**Title**

Information and Communication Technologies and Online Platforms in Developing Economies



**Author**

Nakamura, Shotaro Nishimura

**Publication Date**

2023

Peer reviewed|Thesis/dissertation

# Information and Communication Technologies and Online Platforms in Developing Economies

By

SHOTARO NISHIMURA NAKAMURA
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

ECONOMICS

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Arman Rezaee, Chair

David Rapson

Giovanni Peri

Committee in Charge

2023

To Miho and Jiro Nakamura, whose unwavering support allowed me to dream beyond the horizon.

And to Masako and Masuzo Nishimura, whose love has been a continual source of strength.

# Abstract

Information and search frictions are often cited as sources of market distortions and failures in developing economies. Information communications technology (ICT) and online platforms may help solve these problems, as evidenced by increased adoptions of mobile applications, e-commerce, and gig-economy platforms. Technology-enabled tools and marketplaces may also help address challenges faced by regulatory frameworks in developing economies, such as increased visibility on informal labor markets through digital traces. Furthermore, ICTs may also help lower the costs of providing public goods and services with limited resources and under weak institutions. In this dissertation, I address some of the challenges and opportunities that have arisen in the context of rapidly improving information communication technologies and the emergence of online marketplaces.

In the first chapter, I focus on an online-platform-enabled informal labor market for low-skill workers and address the efficacy of regulatory interventions to improve worker welfare. My co-author and I document the effects of a government-mandated price-floor policy intended to improve workers' earnings in the ridesharing market in Indonesia. We hypothesize that when a market-wide impact is accounted for, a price floor regulation may have muted impact on policy objectives or unintended consequences due to adjustment mechanisms and general equilibrium effects. We measure the causal impact of the policy by taking advantage of an exogenous variation in the policy's rollout.

We find that, on average, the policy increases the average trip price, as expected from a binding price floor. However, we do not find significant effects on the overall transaction volume, driver earnings, or wages. The results can be explained by a significant increase in labor supply, which reduces the number of transactions allocated per driver hour. We fail to find evidence that the price floor policy is redistributive; the excess labor supply comes from lower-earning drivers but does not lead to their increased earnings. We also find

that the policy reduces driver productivity through increased excess supply and is driven by two margins: an increased share of less productive drivers in the workforce and reduced individual productivity due to the crowded supply side. In sum, we find that the price floor policy does not achieve its objectives at an efficiency cost.

In the second chapter, I focus on the mechanics of information and search frictions in online marketplaces via a randomized controlled trial (RCT). My co-authors and I document how an information intervention triggers strategic responses, spillovers, and adjustments to the market in developing economies. We identify a context in which there is limited publicly available price information in the used car market in Pakistan and collaborate with a major online listing platform, PakWheels.com. In our randomized intervention, we provide estimates of transaction prices privately to sellers. We then measure the impact on sellers' pricing and advertising choices, and transaction outcomes. We design the experiment such that we capture both direct and spillover effects.

We find that sellers in online markets respond to an exogenously varied information environment by adjusting their pricing and advertising behaviors and that their choices generate spillovers to other sellers. The empirical estimates show that the information intervention brings directly treated sellers' listing prices closer to our price estimates and reduces advertising usage. These adjustments by treated sellers induce spillovers to sellers who are not directly treated by increasing the page views they receive from potential buyers and improving their transaction probability. The findings point to two mechanisms: 1) effects of price information are mediated by advertising tools that could countervail effects of list-pricing choices, and 2) spillovers could propagate direct effects of information intervention via adjustments by competing sellers.

In the third chapter, I address the role that ICTs can play in improving public service provision. My co-author and I focus on the context of air quality information provision in Lahore, Pakistan, where government services are intermittent and hard to access for the general public. With the availability of low-cost air quality sensors, private citizens' groups have emerged to record and disseminate real-time air quality information via social media. We test how citizens' demand for information and beliefs about air quality depend on the provider they receive information from. We conduct a field experiment in which we provide day-ahead air pollution forecasts via SMS and make salient one of the information sources—the government vs. a private citizens group. We then track respondents' willingness to pay for the SMS service and collect a series of incentivized belief measures on service quality and preference for information sources.

We find that respondents have a high willingness to pay for the forecast service, yet not differentially so between the sources. We also find that they have a significantly higher revealed preference for the assigned source against the other. Respondents' beliefs about air pollution levels are not statistically different across treatment groups, but their belief about the government forecast error is 12% higher than for the private alternative. Our findings suggest that respondents have weak priors and malleable preferences for information sources yet expect lower service quality from the government.

In summary, I provide empirical evidence on economic agents' beliefs and choices, their spillover effects, and market forces in the context of ICTs and online platforms in developing economies. In future work, I hope to extend my insights into the following areas. First, on informal labor markets, as discussed in Chapter 1, I hope to conduct with additional causal analysis on how workers adjust to income shocks and uncertainties via RCTs. Second, on information frictions on online platforms, as discussed in Chapter 2, I hope to explore other relevant types of frictions that may limit the efficacy of online markets in developing economies, such as trust and fraud. Third, on air quality information, as discussed in Chapter 3, I hope to understand how citizens adjust their behaviors in response to air quality information, with a focus on choices in the labor market.

# Acknowledgments

I am indebted for the support I have received from the UC Davis community and beyond that allowed me to complete this dissertation.

First, I would like to thank my main academic advisor, Arman Rezaee, whose support was critical to the success of the two field experiment-based chapters and other projects requiring extensive fieldwork. I am grateful for the time he invested in me as I took on risky projects (and often failed), the advice he shared as I developed collaborations with field partners, and the technical expertise I relied on as I implemented my randomized evaluations. I am also grateful to my second advisor, David Rapson, who helped me explore my interests in energy economics, pushed me to think critically with kindness, and provided valuable feedback that often allowed me to think about the research questions with new perspectives. A huge thanks to my third advisor, Giovanni Peri, who has taken me under his wing despite a late shift in the focus of my dissertation. His advice was critical to shaping my job market paper, and his guidance and kindness were essential to navigating the market.

I have also been fortunate to be mentored by a range of academics at UC Davis and beyond. Thank you to Monica Singhal and Dalia Ghanem for being on my oral exam committee and providing valuable feedback. Thank you, Andres Carvajal, for helping the 2022-23 cohort navigate the job market. I am thankful for the mentoring from the development group, particularly Diana Moreira and Ashish Shenoy. On the environment and energy side, I am particularly indebted to Erich Muehlegger and Jim Bushnell, for whom I also worked as a graduate student researcher and teaching assistant. Beyond UC Davis, I am indebted to many former teachers who helped me find my passion in economics, taught me how to conduct research as an RA, and wrote recommendation letters for graduate programs. I am particularly indebted to Nava Ashraf for my time as a research assistant at HBS and LSE and Sundip Sukhtankar and Doug Staiger for my time as an undergraduate at Dartmouth.

I was also able to dream of and put to life a dissertation that truly spoke to my interests, thanks to my collaborators and colleagues across the world. Thank you, Rizki Siregar, for your insight and contextual know-how as we successfully engaged with a ridesharing platform for research collaborations. Thank you to my co-authors and mentors, Ali Hasanain, Adeel Tariq, and Sanval Nasim, for your hospitality and support as I navigated my way around Lahore in search of research collaborations. I am also grateful for the hospitality and the engaging academic environment at the University of Copenhagen, where I was able to do a year-long visit, thanks to Finn Tarp.

And none of this would have been possible without the support from family and friends. Thanks to my parents, Miho and Jiro Nakamura, whose unwavering belief in me allowed me to study in the U.S. from a young age. Thanks to them, I have been fortunate to pursue my dreams away from home. And thanks, in particular, to my mother, Miho, for instilling in me a passion for knowledge and a sense of curiosity while forging her own academic path. Thanks to my grandparents, Masako and Masuzo Nishimura, for their love and enthusiastic support for my endeavors. Thank you, Lauren Hennelly, for your love and the inspiration you have given me as I watched you grow and excel as a researcher. Thank you to my family members—Shinjiro Nakamura, Masuyo and Isao Shioya, and the Nakamuras of Yokkaichi—for being my sources of comfort and strength.

Finally, I would like to thank the Dean's Office at UC Davis College of Letters and Science and the Russell J. and Dorothy S. Bilinski Educational Foundation for their financial support.

# Contents:

# Chapter 1: Distributional and Productivity Implications of Regulating Casual Labor—Evidence from Ridesharing in Indonesia

Shotaro Nakamura*† Rizki Siregar‡

August 11, 2023

**Abstract**

Regulations intended to improve workers' earnings, such as minimum wage, may have muted impact on policy objectives or unintended consequences due to adjustment mechanisms and general equilibrium effects. Using comprehensive transaction data from one of Indonesia's dominant platforms, we study the market-wide implications of a federal policy on minimum fares for drivers on ridesharing apps. We estimate the causal effects of the policy with difference-in-differences and synthetic control methods, exploiting an exogenous variation in the policy's rollout. We find that, on average, the policy increases the trip price but does not significantly affect the overall transaction

*The views expressed in this paper are solely those of the authors, and do not reflect the views of the firm from which we received data. The paper has gone through a check by the firms' employees to ensure confidentiality of their data and other proprietary information, but not on the empirical findings and views expressed in the paper. The authors report no conflict of interest. We are grateful to Arman Rezaee, Giovanni Peri, and Dave Rapson for their guidance. This paper also benefited from feedback from Michael Carter, Daiji Kawaguchi, John List, Diana Moreira, Joris Mueller, Monica Singhal, Kosuke Uetake, Justin Wiltshire, as well as participants in the UC Davis applied micro brown bag series and development tea meetings, Tokyo Labor Economics Conference, and NEUDC. All errors remain our own.

†Department of Economics, University of California, Davis, CA; snnakamura@ucdavis.edu

‡Gutenberg School of Management and Economics, the University of Mainz; rsiregar@uni-mainz.de

volume nor increase driver earnings or wages. These effects are driven by a higher excess labor supply, reducing the number of transactions per driver. The excess labor supply comes from lower-earning drivers but does not lead to their increased earnings. We also find lower driver productivity, driven by two margins: an increased share of less productive drivers in the workforce and reduced individual productivity due to crowding on the supply side.

**JEL:** J38, O18, R48

**Keywords**–Ridesharing, Minimum Wage, Casual Labor, Distributional Impact, Labor Productivity

# 1 Introduction

A significant amount of labor around the world takes forms other than formal and salaried employment. Informal labor, which we define as types of work not registered and regulated by an existing legal framework, is the dominant form of work in developing economies, accounting for approximately 70 to 80 percent of all employment in lower and lower-middle-income countries (World Bank 2018).[1] Informal employment is characterized by its lower pay, higher uncertainty, and lower productivity in developing economies (Ulyssea 2018; Kochar 1995; Kochar 1999; La Porta and Shleifer 2014). If the characteristics associated with informal and casual labor markets result from market frictions or failures, policymakers may be interested in regulating them to improve workers' earnings. Price floors may be one way to achieve such outcomes through a transfer from employers to workers.

It is unclear, however, if price floors would deliver intended outcomes on workers' earnings in informal and casual labor markets. In a neoclassical setting with no market failures or frictions, a binding price floor would reduce quantities demanded and induce excess supply. However, empirical findings may deviate from this intuition; in the context of formal, salaried employment, the literature on minimum wage has found mixed evidence on employment, suggesting that reality may deviate from a neoclassical intuition (e.g., Card and Krueger 1994; Cengiz et al. 2019; Jardim et al. 2018). Price-floor policies may also have distributional and productivity implications (e.g., Engbom and Moser, forthcoming). However, the effects of labor regulations in informal and casual labor markets likely differ from those in formal and salaried settings for many reasons, including differences in regulatory structures, the relevance of market frictions and failures, and the extent of employers' market power.

[1]For this statistic, informal employment is defined as "not [having] a contract, social security, and health insurance and is not a member of a labor union" World Bank 2018. Although various definitions of informal work exist, we follow the one set by ILO and emphasize that the vast majority of work in developing economies remains outside of the regulatory framework, including wage floors (International Conference of Labour Statisticians 1993).

In this paper, we study the market-wide impact of labor regulation directly imposed on a casual labor market: two-wheel taxi rides on mobile ridesharing apps in Indonesia. This labor-market segment is "informal adjacent" in that the app-based services have largely replaced the traditionally informal ride-hailing in Indonesia and other developing markets.[2] Digital platforms provide visibility into a type of informal-adjacent labor market that was once almost entirely offline. The platforms also provide an opportunity for policymakers to regulate a once-unregulatable labor market segment, as prices and other attributes of transactions can be controlled algorithmically. In this context, we study the effects of a price-floor policy that had a similar intention as minimum wage: to increase drivers' earnings.

We apply insights and empirical tools from the minimum wage literature to understand the market-wide adjustments, distributional impact, and productivity consequences of price-floor regulation in these labor markets. We study the implications of a government regulation on ridesharing platforms in 2019 that introduced a minimum price that drivers received per trip ("*driver fare*" henceforth). By exploiting the city-level variation in the timing of its implementation, we estimate the causal effects of the policy with difference-in-differences and synthetic control methods. The data-rich environment, in which we have access to the universe of transactions and worker-level productivity data from one of Indonesia's two major online platforms for ridesharing, allows us to identify the effect of policy on distributions and productivity in detail.

We find five strands of results from our empirical analysis. First, when we estimate average treatment effects, we find that the policy leads to a statistically significant 4.6% increase in the driver fare per trip and a statistically insignificant 0.2% increase in transaction volume. These results align with previous work on a minimum wage like Cengiz et al. (2019), who find limited effects of binding minimum-wage increases on employment. However, unlike in

[2] It is difficult to get a reliable estimate of the share of ride-hailing transactions that are mediated by mobile apps. Based on publicly available industry reports, the share of app-based services in the taxi industry is around 40 to 60% in Indonesia and the ASEAN region (Statista 2022; Mordor Intelligence 2022).

the minimum wage literature, we find that the policy only leads to a statistically insignificant 1.7% reduction in daily earnings from driver fare and a statistically insignificant 6.7% reduction in wage, i.e., daily earnings divided by supply hour. The lack of statistically significant effects on driver earnings despite increased per-trip driver fare suggests the importance of adjustment mechanisms, which a feature of the ride-hailing market allows us to identify.

Second, we find that excess supply predominantly explains the limited impact on average drivers' daily earnings despite the higher average driver fare per trip. Because we study a casual labor market in which the cost of labor supply is relatively low, and work is allocated with frictions, we observe realized excess labor supply separately from transactions. We find a statistically significant 24.3% increase in excess supply hours, i.e., the sum of all idle hours on the app from all drivers. This effect is driven by noisy adjustments on both extensive—the number of distinct individuals per day—as well as intensive—how much drivers work conditional on being present on a given day—margins of driver supply adjustments. As a result, we find statistically insignificant effects on the effective wage rate, i.e., earnings per hour available on the app. In the ride-hailing market, where drivers are firms of one that makes their own supply decisions—a setup that may generalize to informal and casual labor markets more broadly—the effect of a higher piece rate on driver wages and earnings is crowded out by excess supply.

Third, we find that the lack of a significant effect on average driver daily earnings masks a heterogeneous impact across their pre-policy labor supply and potential exposure to the policy. We classify drivers' pre-policy earnings into deciles and estimate conditional average treatment effects by them. We find that the policy increases the total labor supply of workers in the bottom 3 to 4 deciles of pre-policy earnings by 20 to 40%. We also find, however, that the per-driver earnings do not increase significantly for those lower-decile workers. This is likely because of the increased extensive-margin effect, i.e., more drivers participate on more days from the lower-earning deciles, though the effects are imprecisely estimated.

Fourth, we find evidence of heterogeneous policy incidence on the demand side. We find that the regulation's impact is not correlated with customers' pre-policy transaction volume, resulting in homogeneous effects on daily expenditure and wait time. On the other hand, when we look at the differential effects by customers' exposure to the policy, i.e., consumption of trips in the pre-policy period that would be regulated, we find increased customer fare and daily expenditure for most exposed customers. Most-exposed customers are, however, not differentially compensated by differentially shorter wait times compared to their less exposed counterparts. These results indicate that: a) customers whose transactions are targeted by the policy have relatively inelastic demand, b) they end up with a high financial incidence of the regulation, and c) driver allocation is not optimally allocated to compensate consumers most affected by the regulation.

Fifth, we find that the minimum-fare policy and associated adjustments come at the cost of reduced driver productivity. There is an 8 to 10% reduction in average driver productivity due to the policy, statistically significant at the 10% level. The following two channels likely drive this effect. First, the policy increases supply hours from low-productivity workers, driving down the average productivity of the fleet via a compositional shift. Second, the policy reduces individual driver productivity regardless of their pre-policy productivity levels. These results suggest that increased driver availability crowds out the productivity of inframarginal and more productive drivers by allocating transactions away to less productive, marginal drivers. In summary, our results suggest that a regulation guaranteeing a minimum payment for a job could affect labor productivity by changing who participates in the labor market and how they perform on the job.

Our empirical findings suggest that the transaction-level price-floor policy intended to increase workers' earnings in an informal and casual labor market may not achieve its policy target; the policy induces a large supply response that reduces the driver-to-customer match rate and productivity, canceling the effect of a higher piece rate on earnings and wages. To rationalize these findings, we introduce a static matching model of the ride-hailing market in which workers' endogenous labor-supply decisions to exogenously shift prices have consequences on productivity and wage. We use this framework to rationalize the differences in effect sizes between our estimates and a study in a similar context by Hall et al. (2021). We find that an idiosyncratically large labor-supply response in our context, likely due to a mix of contextual and policy-design reasons, can describe the differences in the magnitude of effects on productivity and wages.

This article makes several contributions to the literature on labor regulations, casual work in developing economies, and applications of standard labor-policy instruments in the gig economy. First, on the literature on labor regulation and minimum wage, we demonstrate the applications of standard labor policy instruments and the analytical tools used to evaluate the policy impact and adjustment mechanisms. Recent work on minimum wage has focused on identifying changes to the distribution of wages to show if substitutions between less and more productive labor are a relevant mechanism behind the limited effect on employment (Cengiz et al. 2019). Other work decomposes effects by extensive and intensive margins to show that while employment headcounts may not be affected, reduced hours worked could lower net earnings (Jardim et al. 2018). We build these discussions by finding evidence of lower productivity due likely to excess supply.

Second, our platform-wide access to transaction-level data allows us to analyze the productivity effects of labor regulation and its mechanisms. The productivity measures are distinct from wages and available at the driver-day level for all drivers on one of Indonesia's two largest ridesharing platforms. Some previous studies find positive effects on productivity and attribute their findings to the efficiency wage channel, i.e., increases in wages are absorbed by increased worker efficiency (Ku 2022; Coviello et al. 2021; Dustmann et al. 2022). We provide an alternative perspective to this question by showing that productivity is *reduced* for two reasons. First, our data capture an entire labor-market segment of a dominant online platform rather than individual employers or chains, allowing us to detect market-wide implications and effects beyond direct treatment effects. Second, with productivity measures at the worker level rather than at the firm level, we observe not only what happens to individual labor productivity but also how workers of a given productivity level are reallocated. In other words, we capture a more comprehensive effect on productivity, including workers' intensive and extensive margins as well as spillovers.

Third, our work provides novel insights into casual labor in developing cities and the scope of regulation to increase workers' earnings. Our study offers unique insights into the implication of a standard labor policy instrument applied in casual labor markets, where many workers with low socioeconomic status in developing cities make a living. Labor supply decisions of informal, casual-wage workers may be subject to liquidity constraints and reference-dependent preferences (e.g., Dupas et al. 2020). In such a context, public policy that improves the terms of informal labor may help increase earnings and productivity. In terms of minimum wage policies in developing economies, findings are less conclusive than in developed ones, and the studies necessarily focus on the formal sector (Neumark and Corella 2021). For a similar policy that regulates how much individuals are entitled to work rather than how much they get paid per work, there is evidence that guaranteed employment through public works schemes increases earnings and leads to a positive general equilibrium effect in the local economy (e.g., Imbert and Papp 2015; Beegle et al. 2017; Muralidharan et al. 2017). Our work contributes to a gap in the literature with the following insight; a minimum wage-like policy on casual work may have limited efficacy in increasing the earnings of low-wage earners, and it comes at the cost of increased expenditure for price-inelastic consumers and reduced worker efficiency. Our evidence may be helpful in that it shows the potential benefits and costs of policies designed to protect lower-income workers in casual labor markets.

Fourth, we contribute to the understanding of labor-regulation instruments in gig-economy platforms. Our paper is, to our knowledge, one of the first to empirically evaluate a government-initiated price-control policy in an app-based ridesharing market. Our work is closely related to Hall et al. (2021), who assess driver responses and marketplace equilibrium after price shocks, and Horton (2017), who studies the productivity effect of a minimum-wage policy in an online labor market. Our empirical context is similar yet distinct as we primarily focus on the impact of a price-floor policy rather than an average shift in fare. The regulatory structure makes it more likely to induce a distributional effect compared to a uniform price increase. Our context also differs in that it was a government policy that applied to all platforms rather than a platform-led pricing policy in markets served by a single provider (i.e., Uber). Therefore, our environment may be more reflective of the current competitive landscape for gig-economy platforms, and our findings may be more relevant from the regulatory perspective.

We proceed with a description of the empirical context of the fare regulation and the environment of the ridesharing market in Indonesia in Section 2. We then provide details on the data we use in Section 3. Section 4 presents our identification strategies. In Section 5, we describe our findings. We discuss our results in Section 6 with a combination of a conceptual framework and comparisons of estimates with Horton (2017) for external validity. We conclude in Section 7.

# 2 Empirical context

## 2.1 Ridesharing and its regulations in Indonesia

Ridesharing platforms have served Indonesian consumers since the mid-2010s. The platforms provide services by automobile cars and motorcycles called *ojeks*, which are cheaper and more popular modes of ride-hailing in Indonesia. Currently, 60 to 70% of all ridesharing trips are conducted by *ojeks*. Ridesharing platforms also offer various other services by motorcycle drivers, such as food and package delivery. The motorcycle-based services, in particular, disrupted offline services that were the norm before the introduction of such platforms.

In 2019, the government of Indonesia enacted regulations on transactions on ridesharing platforms (Indonesian Ministry of Transportation 2019). It was motivated by growing concerns about the welfare of drivers on ridesharing platforms. Pressures from groups and associations of drivers demanded regulations that protect drivers in terms of, among others, safety and wage security. The government of Indonesia then enacted several laws starting in early 2019 for motorcycle taxi services. In this paper, we focus on one of the regulations, in particular, fare regulations of motorcycle taxi services. Hence, from here forward, we refer to services provided by motorcycle drivers in all of our analyses.

We exploit spatial variation in the timing and threshold of the minimum- and maximum-fare policy. From Indonesian-language news sources, we have identified the following timetable of policy announcements and implementation by the Ministry of Transportation, as shown in Table 1. From our reading of the news coverage, it seems that the May-1st rollout was only temporary, and the government suspended the regulation on May 15th. We have also identified that the policy went into effect for most cities where our ridesharing partner operates by the third phase of the regulation's implementation. Therefore, the main source of policy variation is the second phase (i.e., July-1st implementation) versus the third (August-9th).

## 2.2 Structure of the fare regulation

The fare regulation consists of two components. First, drivers are guaranteed a payment per trip, regardless of the distance or duration of a trip. This component of the policy effectively works as a price floor of a transaction, and we refer to it as the minimum total fare. Second, the policy sets a limit on the range of per-kilometer price that drivers have to be paid, which we refer to as the per-kilometer rate. In reality, the pricing algorithm first estimates the fare based on distance, location, time, and other factors conditional on the constraints on the per-kilometer rate. The estimated fare is then compared to the minimum total fare, and the final driver fare is the maximum of those two values.

The per-km minimum and maximum fare, as well as minimum total fare, varies by groupings of cities, to which the Ministry of Transportation refers as "zones." Zone 1 covers major cities in the populous islands of Bali, Sumatra, and Java (except Jakarta Metropolitan area), Zone 2 consists of Jakarta Metropolitan area (often referred to as Jabodetabek), and Zone 3 includes the rest of the country. Table 2 shows the minimum and maximum per-km fares and minimum total fares imposed by the Ministry. Appendix Section A.1 lists the cities included in each rollout phase of the regulation's implementation. In addition, the fare regulation only targets taxi services. Other services catered by motorcycle drivers, such as delivery, are not subject to this policy.

## 2.3 Context specificity and external validity concerns

Our empirical context is uniquely suited to study the effects of price regulation in equilibrium and the mechanisms that lead to it. First, thickness on both sides of the ridesharing market means that individual "employers" (i.e., customers) likely do not have significant market power. Although the platform decides on transaction prices via its algorithm, it is doubtful that either of the two large platforms has the significant pricing power to deviate from their competitors' prices. The market share of the smaller of the two platforms is 40 to 50% as of 2021 (Measurable AI 2022).[3] The market condition helps us abstract away from monopsony power and instead focus on supply-side mechanisms and distributional effects. Second, the short average contract duration allows us to observe markets in a new equilibrium quickly after the introduction of the minimum-fare policy, unlike in standard wage-employment contracts. Third, unlike uniform price increases led by the platform operators like in Horton

[3] Anecdotally, the two largest platforms have been engaged in fierce competition over price à la Bertrand in the last decade.

(2017), the policy in Indonesia was government-led and generated a price floor, which mimics a traditional minimum wage more closely and may be more likely to induce a distributional impact.

Yet, the data and policy contexts may limit the extent to which our findings may be generalized beyond the ridesharing context. First, the government policy regulated driver fares per transaction rather than driver wages. These differences in the regulated units may limit the relevance of our findings to the minimum wage literature. Second, the policy consisted of both minimum total fare and constraints on per-kilometer rates, making the direct comparison with the minimum-wage design difficult. Third, we cannot observe the long-term impact of labor regulation because of the limited duration of plausibly exogenous policy variation.

## 3 Data

We construct our data sets for analysis using the collaborating platform's database, consisting of the universe of transactions and driver-daily measures. We restrict our sample to all completed motorcycle trips that had non-zero payments to associated drivers. Our data set contains trips from 64 Indonesian cities, 55 of which are in the data for our analysis. The data covers the period of January 1st to August 8th, 2019.

The transaction-level data set contains information about the price and service type of the trip. Each trip is associated with one of the service types, the most popular of which are taxis, food delivery, and non-food delivery. We use data from all service types for the analysis unless specified as being restricted to "regulated" service types, in which case we restrict the data to taxis. For each trip, we can identify payments made to the driver and the price charged to the customer. We define these measures as the driver- and customer-fare, respectively, and use them for analysis. We note that the amount customers are charged is not necessarily what they end up paying, as they may receive discounts through credits or bundled offers. Unfortunately, we are not able to identify the actual amount paid by the customers as we are not able to distinguish discounts and credit usage from non-cash payments in the database to which we have access.

The transaction data also contain information about the trips' characteristics and associated driver and customer IDs. We use booking and completion time stamps to identify the date and duration of the trip. The data also contain route distance estimates used by the pricing algorithm, which we use to compute per-km fare rates. We also use driver and customer IDs to construct driver- and customer-daily level measures.

Other than the transaction-level data, we use the driver-daily level database to measure driver availability. This data source captures a daily measure of driver availability by tracking how long a driver has the platform app open. We take this measure as a proxy of daily supply hours, defined as the amount of time a driver was available for hire on a given day. We also use this measure as the denominator of our productivity measures.

## 3.1 Constructing aggregated analysis data sets

We aggregate the transaction- and driver daily-level data for analysis because we are interested in identifying market-wide effects and distributional impact. We construct city-day level panel data to assess the city-wide average effects on average prices, frequencies, and other outcomes. We describe the empirical strategy that uses this city-day level panel data in Section 4.1.1.

We also aggregate the data at the city-day-"bin" level to identify effects on the distribution of outcomes and to estimate conditional average treatment effects by pre-policy characteristics of drivers and customers. For effects on the distribution, whose results are presented in Section 5.1.1, we aggregate the transaction data at the level of city, day, and driver-fare bin of 1,000 Rupiah. The outcome variable for each row of this panel data is the number of transactions for which the driver fare was in a given fare bin for a given city on a given day. For conditional average treatment effects by pre-policy characteristics, such as results presented in Section 5.3, we construct panel data of averages and counts at the city-day-bin level, where the bin is defined as the decile of the driver- or customer-characteristics in the month prior to the start of our policy variation.

# 4 Identification strategies

We consider two methods for our identification strategy: difference-in-differences and synthetic controls. One key aspect is the extent to which the parallel trends hold. As discussed in Section 2, the government rolled out the new minimum and maximum wage policies non-randomly, and larger cities tended to fall under the new regulation earlier. Violating the parallel-trends condition may mean that the treatment effect estimates are biased.

Figure 1 shows trends of average fees by the rollout phases, with policy rollout timing marked by vertical dash lines of the corresponding color. It shows that cities in Phase 1 have somewhat different trends from those in Phases 2 and 3, which may be because cities in Phase 1 included many of the largest cities in Indonesia. On the other hand, Phases 2 and 3 have remarkably similar time trends on the average fare.

Figure 1 also helps visualize the effect of fare regulation on average prices, which increase sharply right at policy introduction. For confidentiality reasons, we standardize to the average driver fare in cities included in Phase 2 from January 1st all the way up to the day before the policy implementation, and we call this value $X$. For cities in Phase 2, the average fare from the implementation date to the end of the year increased to $1.21X$. For cities in Phase 3, the pre-policy average (i.e., up to August 8th) was $0.82X$, but increased to $1.12X$ post-policy implementation. As such, we will deploy difference-in-differences and synthetic control methods to paint a holistic picture of the policy effect. We present results from our checks on parallel trends and other robustness checks in Appendix Section 5.2.2.

## 4.1 Difference-in-differences

Our preferred empirical approach is the difference-in-differences (DiD) method, where we take advantage of variation in the timing of policy rollout. Due to concerns about the violation of parallel trends, we will restrict our sample to cities in rollout phases 2 and 3, where we are confident of exact policy implementation timing and parallel trends hold, as shown in Figure 1. We restrict the data to dates between January 1, 2019, to August 8, 2019, the last day of policy variation between phases 2 and 3. We conduct difference-in-differences analysis on city-day averages, as well as on fare (and other) bins, as follows.

### 4.1.1 DiD over average outcomes

We estimate the following equations on the city-day level outcomes. Subscript $c$ denotes city, while $t$ denotes day as the unit of time.

$$Y_{c,t} = \beta_0 + \beta_1 * I_{c,t}(c \in Treat, t > 0) + \gamma_c + \delta_w + \rho_d + \epsilon_{c,t} \tag{1}$$

We study the impact of the fare regulation on various city-day outcomes, $Y_{c,t}$. The granularity of the data allows us to construct four groups of our primary city-day outcomes. First, we analyze the impact on transactions, such as average fare, average distance, and average per-km rate. The second group of outcomes is quantities, such as total supply hours, total driver earnings, total distance, number of drivers, and number of bookings. Third, we also estimate the impact of the regulation on driver-daily outcomes, such as drivers' daily supply hours, daily earnings, and wages. We differentiate earnings and wage, with wage defined as daily earnings divided by supply hour. Lastly, we study the impact of regulation on various measures of productivity, such as the ratios of the number of rides, distance worked, and time spent on providing rides over the total hours spent on the platform.

We define the first day of treatment, July 1st, 2019, as $t$ equals 1. Then, the indicator for treatment, $I_{c,t}$, is one if city $c$ is included in phase 2 of the fare regulation's rollout for all days starting on the first day of treatment. This indicator variable is zero otherwise. We include city fixed effects, $\gamma_c$, as well as two sets of time fixed effects: calendar-week fixed effects, $\delta_w$, and day-of-the-week fixed effects, $\rho_d$. Lastly, $\epsilon_{c,t}$ is an idiosyncratic error term.

Our coefficient of interest is $\beta_1$ which represents the impact of the fare regulation on city-day outcomes. This coefficient captures the within-city changes in the outcome variables. It provides the average treatment effect of the fare regulation. We also cluster the standard errors at the city level.

#### 4.1.2 DiD over frequency by fare bins

We use the city-level difference-in-differences (DiD) over various bins to uncover the heterogeneity in localized effects of fare regulation. Again, we exploit the timing of the policy implementation in cities included in phases 2 and 3.

We estimate the effect of minimum-fare policies on transactions and distributions of wages. The basic specification is the following:

$$E_{j,c,t} = \sum_{j=0}^{31} \beta_j I_{j,c,t}(c \in Treat, t > 0, j) + \mu_{j,c} + \rho_{j,t} + \epsilon_{j,c,t} \tag{2}$$

$E_{j,c,t}$ is the number of transactions in $j$*1000-Rupiah per-km fee bin in city $c$ on day $t$. $I_{c,j,t}$ is a treatment indicator term, equaling 1 if the minimum fare policy is implemented in city $c$ on day $t$ and the fare falls between $j * 1000$ and $(j+1) * 1000$ Rupiah. These fee bins are at the 1000-Rupiah increment, except for fees bigger than 30,000 Rupiah, denoted as $j = 31$. Table 2 shows that minimum fare varied by zones (and was, in fact, given as a range rather than a particular value), and cities in different zones are spread over phases 2 and 3 of the policy rollout. In practice, however, the new minimum fare in most cities was 8,000 Rupiah. In order not to introduce any other noise by adjusting these fare bins relative to possibly misidentified minimum thresholds of a given city, we use fare bins in levels.

We also include wage-bin-by-city and wage-bin-by-time fixed effects, $\mu_{j,c}$ and $\rho_{j,t}$ respectively. These terms allow us to control for city-specific fee distributions as well as trends in the nationwide fee distributions. $\epsilon_{c,j,t}$ denotes the error term. We cluster our standard errors by the city level.

Using the estimated $\hat{\beta}_j$s, we can calculate changes in transaction volume over sections of the fee distribution in response to the minimum-fare policy. One of our primary questions is whether the policy reduces transactions below the new minimum-fare threshold as well as increases transactions just above. We define transaction losses as the change in the number of transactions below the new minimum fare over the policy period.

As we can capture the estimated $\hat{\beta}_j$s over the whole distributions of transactions, we can also compute and illustrate the cumulative effect: $\sum_{j=0}^{31} \hat{\beta}_j$. This cumulative effect can be compared to the average treatment effects. Nevertheless, our main aim is to uncover the heterogeneity in impacts to understand better the mechanisms driving the average treatment effects.

## 4.2 Synthetic control-based method

To address potential concerns arising from the violation of the parallel-trends condition, we also estimate the treatment effect using a version of the synthetic controls estimator first proposed by Abadie and Gardeazabal (2003) and Abadie et al. (2010). At its core, the synthetic control methods allow us to construct a counterfactual using a weighted average of untreated units that best resembles treated units' pre-treatment trends. We deviate from the classic implementation of the synthetic control method, where there would be one treated unit and a set of donors in the control group. Rather, we have multiple treated units and multiple donors in the control group. We approach estimation and inference as follows.

First, we construct a synthetic control for each city in the treated group. In our case, these are the cities in the rollout Phase 2, which we denote with subscript $i \in I$ Cities in the donor pool are those in the rollout Phase 3, denoted with subscript $j \in J$. We restrict the data to April 11 to May 31, and June 11 to August 8, 2019 for the following reasons. First, we match to the pretrends closer to the point of policy variation (July 1st) by dropping data from earlier dates, i.e., those before April 11. Second, we find, as we discuss in further detail in Section 5.2.2, that there are significant, one-time drops in the number of transactions and drivers during the week of *Eid al Fitr* in early June. We found in our preliminary analysis that matching on data from this period generates significant noise, so they are dropped when constructing synthetic controls. Third, we trim the data beyond August 8, the last day before the Phase 3 implementation, to ensure that the donor pool cities are not under the policy regime throughout our sample period.

For an outcome $Y$ at time $t$ each treated city $i$ in Phase 2, we construct $\hat{Y}_i t^N$, i.e., a potential outcome for $i$ at $t$ if it were $\{N\}ot$ treated:

$$\hat{Y}_{i,t}^{N} = \sum_{j=I+1}^{I+J} \hat{w}_{j,i} Y_{j,t} \tag{3}$$

The weights $\hat{\mathbf{W}} = (\hat{w}_{I+1,i}, ..., \hat{w}_{I+J,i})'$ minimize the distance between $i$ and the donors on the predictors used to determine the relative importance of each donor. Because we do not have relevant covariates that we can gather outside of the data received from the online platform, we use means of the primary outcome variables from the pre-treatment period as predictors with equal weight.[4] They are:

[4]Because our pre-treatment periods are disjoint due to the omission of data from the *Eid* period, we use two means from April 11 to May 31 and June 11 to June 30.

- log(average driver fare)
- log(number of trips)
- log(average daily customer expenditure)
- log(average daily income)
- log(average number of rides per supply hour), i.e., our main measure for productivity

Next, We construct "placebo" synthetic control for each of the donor-pool cities, using the rest of the donor-pool cities as its donor. In other words, for city $j$ in t

$$\hat{Y}^N_{j,t} = \sum_{k=\in J-\{j\}} \hat{w}_{k,j} Y_{k,t} \tag{4}$$

After having constructed synthetic controls for both treated and control cities, we estimate the treatment effect as a weighted average of the difference between actual and synthetic control outcomes over the post-treatment period. Since we have multiple treated cities, we use the log pre-treatment transaction volume as weights for the average treatment effect and denote it as $\gamma_i$. In other words, the treatment estimate is:

$$\hat{\bar{\tau}}_e = \sum_{i=1}^{I} \gamma_i \hat{\tau}_i = \sum_{i=1}^{I} \gamma_i (Y_{i,t} - \hat{Y}^N_{i,t}) \tag{5}$$

For inference, we measure the extent to which $\hat{\bar{\tau}}$ is an outlier relative to the standard error of the estimator we propose. In order to assess this, we observe the distribution of "placebo" estimate, defined exactly as in Equation 5, except the $i$ subscript is swapped by $j$, and $\hat{Y}^N_{i,t}$ with $\hat{Y}^N_{j,t}$. We denote this placebo estimate $\hat{\bar{\tau}}_p$. We construct 999 such estimates via sampling of donor cities with replacement.

For each estimate, $\hat{\hat{\tau}}_e$ or $\hat{\hat{\tau}}_p$, we take the ratio of mean squared root errors (RMSRE) of the synthetic control from the actual outcome between the post- and pre-treatment periods. Here, for the pre-treatment period we use data from April 11 to June 31, including the *Eid* period. We rank the estimates based on the RMSRE. The ranking of how much larger the post-treatment MSRE is relative to the pre-treatment one, divided by 1,000, is the p-value.

We present estimates from difference-in-differences in the main result sections. We find that these synthetic control-based inference procedure estimates are not qualitatively different. The results are included in Appendix Section B.6.

# 5 Results

Through our analysis using difference-in-differences and synthetic-control methods, we find that price regulation increases average fare and has noisy effects on the transaction volume. However, we find insignificant effects with relatively tight standard errors on driver earnings and statistically insignificant yet considerable estimated reduction in the wage. We find evidence of increased supply and reduced demand on both intensive and extensive margins, though the statistical precision is higher for the results on the supply side. The combination of the demand- and supply-side effects leads to fewer trips per driver, canceling the effect of higher fees on driver earnings.

In terms of the distributional impact of the policy, we find evidence of increased labor supply from low-earners but a much smaller and less robust positive impact on their daily earnings. We confirm that the policy affects the driver fares of those most exposed to the policy, as measured by the share of pre-policy earnings from trips that would qualify for the minimum payment. We find evidence consistent with the average treatment effect's mechanism that increased per-trip earning is canceled out by fewer trips per day, despite increased supply hours. On the demand side, we find that the effects on customer fare, expenditure, and wait time are not substantially different across customers' pre-policy transaction volume. Furthermore, we find that the effect on customers' expenditures is larger for those whose pre-policy orders were most exposed, at no greater compensation in their wait time.

On productivity, we find that the policy reduces drivers' distance and the number of trips per supply hour. We find suggestive evidence that the reduction in driver productivity is driven both by extensive and intensive margins, i.e., an increased presence of less productive drivers as measured at the pre-policy baseline and reduced individual productivity across the pre-policy levels.

The rest of this section is organized as follows. First, Section 5.1 presents the overall treatment effects on price and quantity. It is followed by results on adjustment mechanisms in price distribution in Section 5.1.1. We then discuss demand- and supply-side responses in Section 5.2. We proceed with discussions on the distributional impact of this policy on both the demand and supply sides in Section 5.2. We conclude with results showing the implications on labor productivity in Section 5.4. Additional results are included in Appendix Section B, to which we will refer in the remainder of this section when relevant.

## 5.1 Price and quantity

First, we present results on the average treatment effect on prices and quantity. If the price floor on the amount paid to the driver per trip binds, then the policy should increase the average price of a trip, all else equal. On the other hand, if the ceiling on the per-kilometer price ranges binds, then the policy may reduce the average driver fare per trip. Therefore, the "first-stage" empirical question is if we could detect the effect of regulation on the average driver fare and, if so, in which direction. Coupled with this first-stage question is if we observe effects on the quantity margin, i.e., the number of transactions, and in total driver earnings, i.e., $P \times Q$.

Table 3 summarizes the treatment effects on price and quantity. It shows a 4.6% increase in average driver fare, statistically significant at the 5% level. If we restrict our sample to the type of transactions regulated by the policy, i.e., taxi services, we see an approximately 13-percent increase in the average price that is statistically significant at the 1% level. We further identify which aspects of the price regulation seem to have bound in Section 5.1.1. On transaction volume, we find a minimal point estimate at 0.2 percent, although it is statistically insignificant, and the standard error is much larger than that of the average price. In the subset restricted to the regulated segment, we see a noisy yet large reduction of 10% of transactions. When taken together, the combination of a statistically significant impact on price and a noisy effect on the number of transactions lead to a statistically insignificant yet positive impact on total driver revenue in columns 5 and 6.

We also find shifts in transactions between regulated and unregulated labor subsegments. As discussed in Section 2, the government's policy only regulated taxi trips but not other services carried out by the two-wheel drivers, such as food and delivery. We present results on these possible spillovers in Appendix Section B.3.

A higher average driver fare and a small change to the overall transaction volume should, all else equal, increase daily earnings per driver. Yet, we find that the policy does not significantly increase driver-daily earnings or wages. Table 4 shows that the policy insignificantly reduces drivers' average daily earnings by 1.7% and insignificantly increases it by 2% from the regulated segment. We also find a noisy yet large 6.7% reduction in the imputed average wage, defined as total daily earnings from trips divided by supply hours.

Results from Tables 3 and 4 suggest the presence of some mechanisms through which an increase in drivers' average fare and a noisy yet small effect on overall transaction volume somehow leads to a relatively small and insignificant effect on driver daily earnings. Though noisily estimated, a large reduction in average wage seem to indicate the roles of demand- and supply-side adjustment mechanisms in bringing down drivers' daily earnings to the pre-policy level. We explore this mechanism in Section 5.2.

### 5.1.1 Binding aspects of the price regulation

A statistically significant increase in the average driver price indicates that at least some aspect of the price regulation, consisting of minimum fare and a floor and ceiling for the per-kilometer rates, is binding. To narrow in on the binding aspects of the regulation, we estimate changes to the distributions of driver fares and conditional average effects on per-kilometer rates by the fare bin, as described in Equation 2.

First, we find that the policy shifts the distribution of transaction prices toward the new minimum threshold from below. We adopt this method to identify changes in employment over wage bins from Cengiz et al. (2019). Figure 2 shows the result by 1,000-rupiah bins of driver fare, with red dash lines showing the cumulative effect from the lowest bin. The figure shows noisy yet large reductions in transactions below the new threshold of 8,000 rupiah and a significant increase at this new threshold. The cumulative effect remains relatively close to zero throughout the rest of the distribution, though it seems to converge further to zero. The changes in the distribution on the right-hand tail depend somewhat on the choice of fixed effects, but the shift from the left to the new threshold is robust across various specifications. We, therefore, find that the minimum total fare aspect of the regulation is binding.

To identify if the constraints on per-kilometer rates is also binding, we estimate changes to the mean and median of the total driver fare divided by distance driven. We do so because we cannot separate fixed (i.e., non distance-dependent) fare from distance-based fare. We estimate the treatment effects by the trip-distance bins at a 500-meter increment. If the fixed fare increases but the per-kilometer fare does not, then we should see positive effects for short distances and the effects fade for longer distances. If, on the other hand, the increases on per-kilometer rates matter, then we should see positive effects even on long distances.

The results are shown in Figure 3. We find positive effects on the "per-km fare," here defined as the total driver fare divided by distance driven, for shorter trips, but statistically significant effects beyond 3 kilometers. Effects on other quantiles, shown in Appendix Figure

D.6, however, suggest that the 75th and 95th percentile values of the per-kilometer rate is statistically significantly higher for trips longer than 5 kilometers. We, therefore, find suggestive evidence that the per-kilometer rate is also binding for a subset of transactions.

## 5.2 Demand and supply-side adjustments

We conjecture that the overall lack of a positive effect on drivers' earnings is driven by increased driver availability, leading to crowding on the supply side. Such supply responses likely cancel out gains from higher fares on earnings in the new equilibrium. To demonstrate this, we show the policy's effect on driver allocation to customers. First, we identify the effect on the overall supply, as shown in Table 5. We find an 8.7% increase in the total supply hours, significant at the 10% level. On the other hand, the policy insignificantly increases the total transaction hours, i.e., hours of all trips demanded, by 1.7%. We have also found in Table 3 that the number of transactions is not significantly affected. These results indicate that the fare policy increases excess supply.

Next, we demonstrate that the policy adjusted the extensive-margin labor supply relative to demand, i.e., the number of distinct drivers on the market in a given day, such that there are fewer customers per available driver. We then show that this adjustment affects the intensive margin, i.e., the number of transactions per driver; we find that the per-person number of trips for both drivers and customers is reduced.

Table 6 shows the extensive margin effect, i.e. changes to numbers of distinct drivers and customers. We find noisy effects on driver and customer headcounts, although the number of distinct drivers seems to have increased by 6.5% for all services, and the number of customers decreased by 6.3% for the regulated segment. When looking at a ratio of distinct customers to drivers, however, we find a significant reduction of 4.5% for all services. When restricted to the regulated segment, the effect sizes are larger at -7.4% and significant at the 5% level. We also note that these effects are not signs of a large and significant influx or exodus of drivers into or from the platform; in Appendix Section B.3.1 we show that the policy does not significantly increase the rate of new drivers' entry.

Table 7 shows our findings on demand- and supply-side responses on the intensive margin. We find that both drivers and customers reduce the number of trips per person; when looking at all service types, drivers reduce their daily number of trips by 6.3% (significant at 10% level), and customers by 1.7% (insignificant at conventional levels). If we restrict to the taxi service type, we identify statistically significant effects of -10.9% and -3.5% for drivers and customers, respectively. These intensive-margin results indicate that the per-customer demand for rides is negatively affected by the policy, though its magnitude is relatively small. Consequently, because of the increased labor supply, there is a larger negative impact on the number of trips allocated per driver.

In summary, the minimum-fare policy increases excess supply by inducing demand and supply adjustments on a combination of intensive and extensive margins, such that fewer trips are allocated to a given driver. Lower allocation seems to have canceled out the effect of a higher per-trip fare, resulting in a statistically insignificant effect on daily earnings with a relatively small coefficient.

### 5.2.1 Demand-side response and compensation via wait time

Tables 6 and 7 suggest some reduction in the quantity demanded by the customers as a result of policy, though the magnitude is relatively small and most of the effects may be canceled out by the shift into non-regulated service types. In this section, we explore the extent of the policy's incidence on customers and potential compensations in wait time. As such, we look at the average treatment effects on per-driver daily estimated expenditure and a proxy of wait time. There are, however, some limitations to these outcome measures, and results should be taken with caution; first, the customer fare, a daily customer-level sum of which is used as their daily expenditure, is inclusive of any credits and discounts they may receive.

Second, our proxy of the wait time is "wait distance," i.e., the linear distance between where the driver accepted the order and where it is picked up.

Table 8 shows the results on driver-level daily expenditures and their wait time per trip. We find that the policy increased average daily expenditure on trips by customers by 8.0% and 19.6% if restricted to the regulated segment. The increased customer expenditure, however, is coupled with a significant reduction in wait time; we find a significant 14.9% reduction in wait distance for all services and a 7.6% reduction in the regulated segment. These results suggest that customers may have incurred a higher cost of trips, though we do not know if or the extent to which this effect is canceled out by credits and discounts. We also find that customers are, on average, compensated for seemingly higher fares by shorter wait time. However, we find in Section 5.3 that this compensation is not differentially higher for trips whose fares are increased due to the policy.

We also estimate demand and supply elasticities from the policy variation, though results on this should also be taken with caution due to the measurement concern on the customer fare. We use the variation in average fare from the policy. Consumer- and producer-facing prices differ because the platform charges transaction fees and may implement other policies that may dissociate driver payments from customer fees. We show that the increased driver fare per government policy is passed through to the consumer in Appendix Section B.4, and that noisily estimated demand and supply elasticities suggest that the average elasticity is low for demand and high for supply in Appendix Section B.5.

### 5.2.2 Identification: Parallel trends

As discussed in Section 4, the causal interpretation of our coefficient estimate hinges on the assumption that, in the absence of the minimum-fare policy, trends of the outcome variables for the phase-2 rollout cities are identical to those of the phase-3 cities. Although the counterfactual itself is unobservable, we test for evidence of parallel pre-trends in the four main outcomes: average price, number of transactions, driver-average daily earnings, and driver-average number of trips per supply hour (i.e. a productivity measure).

We use a distributed lag model to estimate the presence of pre-trends. We identify weekly treatment effects 6 weeks into the policy variation period, as well as 7 weeks prior. We use the following equation for the city-day level outcomes where subscript $c$ denotes city, and $t$ and $w$ denote day and calendar week as units of time respectively.

$$Y_{c,w,t} = \beta_0 + \Sigma_{k=-6}^{6}\beta_1^k * I_{c,t}(c \in Treat, w = k) + \gamma_c + \rho_w + \epsilon_{c,t} \tag{6}$$

$\hat{\beta_1^k}$ is the DiD estimate for week $k$, where $k = 0$ corresponds to the week right before the start of policy variation. Figure 4 shows the results. Subfigure (a) shows that the weekly pre-policy coefficients ($k \leq 0$) on the log-average driver fare outcome are small statistically indistinguishable from 0. We then observe a significant positive jump at $k = 1$, and remains at the same level into week 6. We observe similar trends, except with a negative impact post-policy, on the productivity outcome in Subfigure (e). On the log-number of transactions and drivers, however, we find evidence of differential pre-treatment patterns, especially in weeks -4 to -2, as shown in Subfigures (b) and (c). This significant pre-trend in the transaction outcome may also contribute to the noise in the average treatment effect. In Subfigure (d), we also find differential pre-trends in the driver-average daily revenue outcome, though the magnitude of it is significantly smaller.

Upon closer inspection of the data, we find that the points of significantly different pre-treatment patterns in Subfigures (b) and (c) coincide with Eid al-Fitr, a major Islamic holiday marking the end of Ramadan. We find that the transaction volume and the number of market participants drop substantially in most cities. We find that this drop-off happened to be significantly larger for the Phase-2 (i.e., “treated”) cities than in Phase-3 ones. We also find that this differential drop-off in transaction is particularly pronounced in lower-earning and productivity deciles, as Appendix Section B.1 shows.

We address the deviations from parallel trends condition in the following ways. First, we estimate the average treatment effects on data excluding the Eid al-Fitr period on outcomes for which the parallel trend is shown not to hold in Figure 4. The results are presented in Appendix Tables C.9 and C.10, showing that the estimates and the statistical significance of the results are qualitatively unchanged. Second, we present the results from analysis using the synthetic control-based inference procedure in Appendix Section B.6, and find qualitatively similar results as ones from difference-in-differences. Third, for analysis using panel data at the city-bin-day level, we include a bin-date fixed effects on top of city-bin ones to control for bin-specific trends to the best extent possible.

### 5.3 Distributional impact

In Section 5.2, we identify average demand- and supply-side responses to the regulation. However, price-floor policies often aim to protect and transfer welfare to the most vulnerable or lowest-earning workers in the labor market. In this subsection, we assess if and how the policy achieves its potential goal of redistribution. We focus on two aspects of driver and customer heterogeneity: earnings/expenditure and potential exposure of their supply/consumption to the policy.

We run the DiD model by pre-policy transaction-volume quantiles as described in Section 4.1.2. Aside from the main specification, which includes city-bin and bin-date fixed effects, we present an alternative version that includes city-bin and date fixed effects. We select four measures from which to construct deciles of individuals' transaction volume and policy exposure, all from the month prior to policy variation (June 1 to 30, 2019). We restrict the sample to existing drivers and customers. The measures are:

- Drivers' total earnings from trips
- Drivers' share of total earnings that would have qualified for the minimum fare
- Customers' total expenditures for trips
- Customers' share of total expenditures that would have qualified for the minimum fare

#### 5.3.1 Distributional impact on the supply side

Figures 5 and D.7, as well as their Appendix analogues, show the conditional average treatment effects by drivers' pre-policy earnings and policy exposure deciles, respectively. The interpretations of subfigures are as follows. In subfigures (a), we show the conditional average treatment effects on driver fare by decile. We then show the effect on aggregate supply by the decile in subfigure (b). To identify mechanisms on the aggregate-supply effect, we break down the effects into extensive and intensive margins, i.e. the number of distinct drivers and supply hours per driver, in subfigures (c) and (d). We then identify the effects on driver earnings and wages on subfigures (e) and (f), respectively.

Figure 5 shows that while low-earning drivers increase their supply, they do not experience increased earnings. Subfigure (a) shows that the policy increases the per-trip driver fare fairly uniformly across earning deciles, though the effect seems to be slightly larger for higher income deciles. The larger effect on the highest earnings deciles does not seem to be robust to the choice of fixed effects and the span of the pre-policy period over which to calculate the pre-policy decile. Subfigure (b) shows that the policy increased the labor supply of workers in the 4 lowest pre-policy earning deciles, by approximately 15 to 40%. We then break down these effects on aggregate supply in subfigures (c) and (d) into the extensive and intensive margins, respectively. Subfigure (c) shows that the policy increases the number of distinct drivers participating from the bottom 3 to 4 deciles by 10 to 20%, depending on the decile and the set of fixed effects. Subfigure (D) shows that, conditional on driver's availability, the policy increases the driver supply hours by 5 to 20%, again depending on the decile and fixed effects. The consequences of the supply responses, and its intensive- and extensive-margin effects, on earnings and wages are shown in subfigures (e) and (f). We find that the estimates on the effects on earnings for the low earners are smaller, at around 5 to 10%, and statistically indistinguishable from zero. In sum, we find evidence consistent with the overall supply-side responses highlighted in Section 5.2; large increases in excess supply comes from lower earners, but their effects on post-policy earnings are canceled out by the supply-side crowding.

Appendix Figure D.7 also provides suggestive evidence that confirms the supply side mechanisms in response to higher fare. The figures show that the policy increases the average fare for drivers most exposed to the policy, but it did not increase their driver earnings. In sum, we find that the price policy induces supply responses from lower earners and those most exposed to the policy, but the impact on earnings and wages is limited.

### 5.3.2 Distributional impact on the demand side

Figures 6 and 7 show the distributional impact by the differential impact on customers' outcomes by their expenditure and policy exposure, respectively. In these figures, we identify the distributional impact on the quantity and price of customers' orders, as well as their daily expenditure and the compensation mechanism in shorter wait time, in subfigures (a), (b), (c), and (d), respectively.

Figure 6 shows a relative lack of differential impact by customers' pre-policy expenditure levels. We find a uniform increase in customer fare and their expenditure across the pre-policy expenditure bins, accompanied by a uniform reduction in the wait time. Figure 7, on the other hand, shows a more distinct pattern by the policy exposure decile. We find that those most exposed to the policy reduce their number of trips slightly and face significantly higher per-trip fares. This results in higher expenditure for the most-exposed customers.

Interestingly, higher fare and expenditure are not coupled with a greater reduction in the wait distance, as shown in subfigure (d).

In summary, we find that the incidence of the policy falls significantly on customers whose preferred types of trips were the target of the minimum fare. They are relatively demand-inelastic and absorb the higher cost of a trip induced by the policy. It is possible that the minimum-fare policy functions as a transfer from customers with high, inelastic demand to workers with low earnings, as may be an intention of a price-floor policy. The evidence for such a mechanism is relatively weak and should not be taken for granted. We also find that the most exposed customers are not compensated for their higher incidence with shorter wait time any more than less frequent ones. This result indicates that shorter wait time is a byproduct of supply-side adjustments in the market at large rather than a mechanism to offset the impact of certain customers' higher fares.

## 5.4 Effects on labor productivity

A potential implication of increased labor supply that competes over a fixed amount of transactions, along with other distortions and reallocations that the pricing policy might have introduced, is that drivers are less effective at finding and executing their trips. In this section, we investigate the impact on driver productivity. We identify productivity effects on two channels; reallocation of work to less productive workers and reduction in individual productivity. We find suggestive evidence that the policy reduces average driver productivity, and this is due to both increased participation of less productive drivers, as well as a reduction in individual productivity regardless of their pre-policy productivity levels.

We define three measures of daily labor productivity as follows: the number of paid rides, distance traveled on those paid rides, and time engaged in them, all over supply hours. The last measure–time engaged in paid rides per supply hour—is also referred to as utilization rate. We note that, given a positive effect on total labor supply in Table 4 and a small and statistically insignificant effect on an overall number of transactions in Table 3, there should be a negative mechanical effect on average productivity measures. In fact, we find in Table 9 that using these three measures, there is approximately an 8 to 10% reduction in labor productivity, some of which is statistically significant at the 10% level.

#### 5.4.1 Do less productive workers participate more?

One possible mechanism behind the reduced average driver productivity is that the policy induces less productive drivers to participate more frequently, thereby increasing the share of inherently less productive drivers in the fleet. This seems plausible, as we find in Section 5.3 that lower-earning drivers increase their labor supply. If they are also less productive according to our measures, then the resulting stock drivers on the market would be less productive.

We find evidence for this mechanism in Figure 8. We identify effects on the log daily number of distinct drivers by the pre-policy productivity decile, with the city-bin fixed effects, as well as either date or bin-date fixed effects. We use distance traveled per supply hour as the pre-policy productivity measure. We find that drivers in the two to three lowest pre-policy productivity deciles increase daily participation significantly, depending on the set of fixed effects. On the other hand, drivers in the highest deciles of pre-policy productivity may reduce daily participation, though the results do not hold with more robust fixed effects. These results suggest that the policy induces a compositional shift in the fleet toward less productive drivers.

#### 5.4.2 Are drivers less productive on average?

Another possible mechanism is that, as a result of the crowded labor market or changes to the on-the-job incentives, drivers are less productive individually than they were before the introduction of the policy. To assess this possibility, we again look at treatment effects by pre-policy productivity deciles. The outcome measure this time, however, is the productivity measure itself. Results in Figure 9, indicate a noisy reduction in individual productivity of the similar magnitude as in Table 9, depending on the decile.The results, therefore, imply that crowding on the supply side, in turn, reduced individual productivity.

# 6 Discussions

Our empirical results show that a price-floor policy in a ridesharing market increases labor supply from the lower-earning and less productive drivers. We also find that excess supply crowds the supply side, canceling out the effects of higher per-trip fares on driver wages. Increased labor supply and inelastic demand also reduce individual productivity by lowering the match rate.

In this section, we introduce a conceptual framework to highlight the adjustment mechanisms of the minimum-fare policy through the supply-side response and its impact on the match rate. We illustrate that an exogenous trip-price increase leads to an increase in labor supply and a reduction in match rate. We find a positive effect on wages, but the relative magnitude to the labor supply response depends on parameter values, such as the convexity of the labor cost function.

We then compare the effect sizes from our analysis with another study with the most similar empirical setting—a study by Hall et al. (2021) on the effect of Uber's base-fare changes on drivers' responses, transaction outcomes, and productivity. Based on the conceptual framework and its prediction, we compare the results with Hall et al. (2021) and hypothesize on the reasons behind the similarities and differences of our results.

## 6.1 Conceptual framework: setup

The conceptual framework focuses on market-wide effects and abstracts away from details on the search and matching processes. Many papers studying pricing policies on ridesharing platforms, such as Frechette et al. (2019) and Castillo (2020), focus on dynamic and spatial aspects of search and match and benefits of centralized allocation system and surge pricing, respectively. Instead, we take an approach similar to that of Hall et al. (2021) by focusing on market-wide effects on labor supply and earnings.

We imagine a marketplace in equilibrium where platform operators derive piece-rate price $p$ for a representative trip in the market via an algorithm.[5] In other words, we treat the price $p$ as being set exogenously by the platform rather than being endogenously determined to clear the market. Driver availability and demand are then endogenously determined in response to the exogenously set price. Trip allocations are made to drivers via an algorithm rather than to the lowest bidders, leading to some number of idle drivers (i.e., excess supply) in equilibrium. The assumptions we make are similar to those of Hall et al. (2021).

Drivers are matched with a trip at a rate $x$, which is endogenous to demand and supply. A driver $i$ decides how many hours $l_i$ to make themselves available on the app. Not all units of $l_i$ translates into earnings, as some are spent idling. Drivers determine $l_i$ based on the exogenously determined price of a ride $p$, match rate $x$, and their cost function in terms of labor supplied, $c()$, where $c'(.) > 0$ and $c''(.) > 0$. A driver's utility function is $U_i^d = pxl_i - c(l_i)$. We include individual subscripts to account for driver heterogeneity in a latter subsection. Market-wide supply is $L = \sum_{i \in \mathbb{I}} l_i$.

On the demand side, a customer's quantity demanded is a function of price and wait time $w$, which itself is a function of match rate $x$. Customer's utility is $U^c = f(p, w(x))$,

[5] In reality, two platforms compete against each other for customers while maintaining an efficient fleet of drivers. We abstract away from aspects of platform competition because, anecdotally, these platforms engage in price competition and offer relatively homogeneous prices and services on a given trip attribute. We focus on market-wide responses to government regulation on pricing, assuming that the two platforms respond similarly.

and the market demand $D(f(p, w(x)))$, where $D'(.) < 0$. Because we do not find significant demand-side responses empirically in Section 5.2, we simplify the demand-side equation to be of homogeneous agents responding deterministically to the price and wait time.

In equilibrium, drivers maximize their utility by selecting $l_i$ subject to the market price $p$ and match rate $x$, demand $D$, and the match rate $x$. In other words:

$$px - c^{'}(l_i^*) = 0 \tag{7}$$

and,

$$x^* = \frac{D}{L} \tag{8}$$

## 6.2 Conceptual framework: Comparative statics

Next, we identify the effect of exogenous price increases on driver supply, match rate, and wages. For now, we treat drivers as homogeneous agents of a unit mass. Implicitly differentiating the driver first-order condition in Equation 7 with respect to $p$, we get:

$$\frac{dl_i}{dp} = \frac{x + p\frac{dx}{dp}}{c''} \tag{9}$$

Differentiating Equation 8 with respect to $p$ gives us:

$$\frac{dx}{dp} = \frac{1}{L}\frac{dD}{dp} - \frac{x}{L}\frac{dL}{dp} = -\frac{x}{L}\frac{dL}{dp} = -\frac{x}{L}\frac{dl_i}{dp} \tag{10}$$

The last equality is true if $\frac{dD}{dp} = 0$, a scenario which seems to fit with empirical evidence and assertion we make to simplify our comparative static results. Plugging 10 into 9 gives:

$$\frac{dl_i}{dp} = \frac{xL}{c''L + xp} > 0. \tag{11}$$

**Proposition 1** *With homogeneous drivers in the market, an exogenous price shock increases driver supply by* $\frac{xL}{c''L+xp} > 0$.

**Corollary 1.1** *With homogeneous drivers in the market, an exogenous price shock reduces match rate* $x$ *by* $-\frac{x^2}{c''L+xp} < 0$.

We define wage as $xp$, i.e., the expected earnings per unit of labor supply $l_i$. Using equations above and totally differentiating $xp$, we derive the response in wage to a change in price $p$:

$$\frac{dxp}{dp} = x + p\frac{dx}{dp} = \frac{xc''L}{c''L + xp} > 0. \tag{12}$$

**Proposition 2** *With homogeneous drivers in the market, an exogenous price shock increases driver wage by* $\frac{xc''L}{c''L+x+xp} > 0$.

Propositions 1 and 2 show that with an exogenous price increase, driver supply and wages increase. We also see that the effect on wages is a combination of those on prices and match rate. The effect on wage coming from the changes in match rate is negative, as shown in Corollary 1.1. The effect of exogenous price increases on wages is, therefore, tampered by a reduction in match rate.

There are some areas in which our empirical results do not necessarily align with the findings from our conceptual framework; we do not find evidence for increased average wage. This could be because of parameter values; for instance, our theoretical framework would predict a small effect on the wage if workers' cost function is not very convex (i.e., low $c''$). Second, the matching rate may decrease exponentially to crowded supply side in reality, which our conceptual framework does not capture. Third, agents may be heterogeneous, and so are their labor-supply responses and effects on their match rate, a key parameter in driver productivity.

## 6.3 Comparisons with other empirical estimates of pricing policy

We compare our empirical estimates to other work that studies the effect of pricing in ridesharing markets on labor supply, match rate, and wages. Our main point of comparison is Hall et al. (2021) (HHK for the remainder of this subsection), which studies the effect of Uber's base-fare changes on drivers' responses, transaction outcomes, and productivity. Following the setup and results of the conceptual framework, we focus on the following outcomes:

- Number of trips
- Aggregate supply hours
- Productivity (kilometers driven on trip/supply hour)
- Wage

The rationale behind these choices are as follows; first we compare the effect on the number of transactions, as we consider it to be determined by the demand-side responses. We confirm that the aggregate demand effect between our context and HHK's are similar, in that we both find limited responses. We then identify the relative effect sizes on the aggregate supply hours, i.e., the choice variable of the conceptual framework. We then analyze how the choices on supply influences productivity (and its analogue, the match rate between drivers and customers) and driver wage.

To compare effect sizes, we standardize our estimates into elasticities to driver fare, i.e., the percentage change in the outcome variable in response to a percentage change in the average driver fare. We use the two-stage least squares (2SLS) regression, where we instrument for the driver fare with the policy variation variable $I_{c,t}(c \in Treat, t > 0)$, the regressor of the main specification in Equation 1. We provide further detail of the 2SLS approach we use in Appendix Section B.5.

We find that our estimates are in line with those of HHK, albeit with differences in magnitude. Table 10 shows the elasticity estimates for some of our key outcome variables. The table also reports corresponding estimates from HHK. First, Column 1 shows that the effect on the number of transactions is small and indistinguishable from 0 in both our estimates and HHK, making it easier to compare the supply-side effects. Second, Column 2 shows that our supply hour elasticity estimate of 1.76 is, though statistically insignificant, more than five times greater than that of HHK (0.34). The drastic difference in the supply response may drive the difference in the magnitude of productivity elasticity, defined as kilometers driven on the trip per supply hour, in Column 3. We find the productivity elasticity of -1.87, nearly three times as much as in HKK (-0.66). Lastly, Column 4 shows that the elasticity on wage is negative in our estimation at -1.37, as opposed to small and negative in HHK of 0.075, though neither are statistically significant.

We speculate that the larger labor-supply response in our context than those in HHK drives the differences in the effects on productivity and wages. We first note the differences in contexts, such as the locale and the structure of the price policy—a price floor versus a uniform increase in the base fare. These contextual differences make it difficult to attribute the cause of a larger labor-supply response in our findings. However, conditional on this effect on labor supply, our results, combined with HHK's, seem to confirm that supply-side crowding reduces productivity and wages. We provide a conceptual framework to elucidate this mechanism in the next section.

# 7 Conclusion

Regulation to improve workers' earnings via a price floor on a unit of labor may not achieve its intended policy targets when considering its effects in market equilibrium. In this article, we study the impact of the introduction of a minimum fare policy for ridesharing app workers in Indonesia. We find that the exogenously shifted price of labor would have a limited impact on workers' average earnings in equilibrium. We identify demand- and supply-side mechanisms through which such muted impact would result. We also find that the average effect on driver earnings masks the heterogeneous impact, where lower earners supply more labor but may not necessarily earn more. We also find suggestive evidence that adjustments from the policy led to reduced worker productivity via changes in the composition of workers and individual reduction in productivity. These findings have several important implications for the economics literature and policy-making.

First, the findings add to our understanding of regulating informal work in developing countries. Previous work such as Muralidharan et al. (2017) studies policies that guaranteed both the quantity (i.e., number of days employed) and price (daily earnings) of work to casual laborers, leading to positive earnings and structural transformation. In our context, however, only the price on the piece rate is guaranteed, and the market-equilibrium effect cancels out higher payments per transaction for an average worker. These differences in the policy design—and likely the scale and the fiscal commitment from the government—may have contributed to different outcomes between our analysis and one by Muralidharan et al. (2017). Our findings suggest that a simple minimum wage-like policy would not trigger increased earnings for all workers, even when enforcement of such policy is made feasible by the increased presence of online platform-mediated marketplaces. Instead, the policy may induce more labor supply from lower earners and induce competition over a given amount of demand. We are, however, unable to offer definitive insight as to whether the policy would result in a meaningful redistribution toward lower earners.

Second, our findings also provide novel insights into market-wide implications of and adjustment mechanisms triggered by the implementation of minimum wage, despite some contextual differences. We find noisy yet small effect sizes on the overall transaction volume and shifts in the price distribution, similar to findings of Cengiz et al. (2019). Our results suggest that a small overall effect on disemployment can occur without monopsony power, although the statistical precision on that statement is limited. Our results are also in contrast with Jardim et al. (2018), who finds lower net earnings via reduced hours worked. We find a relatively tight estimate of around zero on average earnings. Furthermore, we find that through crowding on the supply side, the policy *reduces* productivity in equilibrium. This result stands in contrast to previous work that found that an increased minimum wage would attract more efficient workers when offered by more localized, specific employers rather than across the market (e.g., Ku 2022; Coviello et al. 2021). The fact that we find a reduction in productivity implies that regulation on labor price may have negative allocative consequences; with the policy, it now takes more working hours to deliver the same amount of transactions, while the excess labor supply could have been more productive elsewhere. Hence, our findings suggest that policymakers may want to weigh the allocative costs against any potential benefit of the policy.

Lastly, our findings also provide insights into the efficacy of regulating labor markets on online platforms. Our results are in line with findings from previous work on pricing policy like Horton (2017), who show increased labor supply and crowding to cancel out the effects of higher piece-rate on earnings. Our additional contributions are the distributional impact that may be driven by the policy design; minimum fares rather than uniform increases in fare may affect some drivers positively and others negatively. Our findings, therefore, suggest not only limited efficacy of minimum fare on earnings at the cost of lower labor productivity but also potentially uneven effects on drivers based on their pre-policy productivity and the extent of reliance on ridesharing as a source of income.

# References


Alberto Abadie and Javier Gardeazabal. The economic costs of conflict: A case study of the basque country. *American economic review*, 93(1):113–132, 2003.

Alberto Abadie, Alexis Diamond, and Jens Hainmueller. Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American statistical Association*, 105(490):493–505, 2010.

Kathleen Beegle, Emanuela Galasso, and Jessica Goldberg. Direct and indirect effects of malawi's public works program on food security. *Journal of Development Economics*, 128: 1–23, 2017.

David Card and Alan B Krueger. Minimum wages and employment: A case study of the fast-food industry in new jersey and pennsylvania. *The American Economic Review*, 84 (4):772, 1994.

Juan Camilo Castillo. Who benefits from surge pricing? *Available at SSRN 3245533*, 2020.

Doruk Cengiz, Arindrajit Dube, Attila Lindner, and Ben Zipperer. The effect of minimum wages on low-wage jobs. *The Quarterly Journal of Economics*, 134(3):1405–1454, 2019.

Decio Coviello, Erika Deserranno, and Nicola Persico. Minimum wage and individual worker productivity: Evidence from a large us retailer. 2021.

Pascaline Dupas, Jonathan Robinson, and Santiago Saavedra. The daily grind: Cash needs and labor supply. *Journal of Economic Behavior & Organization*, 177:399–414, 2020.

Christian Dustmann, Attila Lindner, Uta Schönberg, Matthias Umkehrer, and Philipp Vom Berge. Reallocation effects of the minimum wage. *The Quarterly Journal of Economics*, 137(1):267–328, 2022.

Niklas Engbom and Christian Moser. Earnings inequality and the minimum wage: Evidence from brazil. *American Economic Review*, forthcoming.

Guillaume R Frechette, Alessandro Lizzeri, and Tobias Salz. Frictions in a competitive, regulated market: Evidence from taxis. *American Economic Review*, 109(8):2954–92, 2019.

Jonathan V Hall, John J Horton, and Daniel T Knoepfle. Pricing in designed markets: The case of ride-sharing. Technical report, Working paper, Massachusetts Institute of Technology, 2021.

John J Horton. Price floors and employer preferences: Evidence from a minimum wage experiment. *Available at SSRN 2898827*, 2017.

Clement Imbert and John Papp. Labor market effects of social programs: Evidence from india's employment guarantee. *American Economic Journal: Applied Economics*, 7(2): 233–63, 2015.

Indonesian Ministry of Transportation. Decree of the minister of transportation of the republic of indonesia number 348, 2019.

International Conference of Labour Statisticians. Resolution concerning statistics of employment in the informal sector, adopted by the fifteenth international conference of labour statisticians. In *The Fifteenth International Conference of Labour Statisticians*, 1993.

Ekaterina Jardim, Mark C Long, Robert Plotnick, Emma Van Inwegen, Jacob Vigdor, and Hilary Wething. Minimum wage increases and individual employment trajectories. Technical report, National Bureau of Economic Research, 2018.

Anjini Kochar. Explaining household vulnerability to idiosyncratic income shocks. *The American Economic Review*, 85(2):159–164, 1995.

Anjini Kochar. Smoothing consumption by smoothing income: hours-of-work responses to idiosyncratic agricultural shocks in rural india. *Review of Economics and Statistics*, 81(1): 50–61, 1999.

Hyejin Ku. Does minimum wage increase labor productivity? evidence from piece rate workers. *Journal of Labor Economics*, 40(2):325–359, 2022.

Rafael La Porta and Andrei Shleifer. Informality and development. *Journal of economic perspectives*, 28(3):109–26, 2014.

Measurable AI. Ride-hailing race in indonesia: Gojek versus grab. https://blog.measurable.ai/2022/01/18/ride-hailing-marketshare-in-southeastasia-indonesia-gojek-versus-grab/, 2022. Accessed: 2022-10-27.

Mordor Intelligence. Asean taxi market - growth, trends, covid-19 impact, and forecasts (2022 - 2027). https://www.mordorintelligence.com/industry-reports/asean-taxi-market/, 2022. Accessed: 2022-10-27.

Karthik Muralidharan, Paul Niehaus, and Sandip Sukhtankar. General equilibrium effects of (improving) public employment programs: Experimental evidence from india. Technical report, National Bureau of Economic Research, 2017.

David Neumark and Luis Felipe Munguia Corella. Do minimum wages reduce employment in developing countries? a survey and exploration of conflicting evidence. *World Development*, 137:105165, 2021.

Statista. Ride-hailing and taxi - indonesia. https://www.statista.com/outlook/mmo/shared-mobility/shared-rides/ride-hailing-taxi/indonesia/, 2022. Accessed: 2022-10-27.

Gabriel Ulyssea. Firms, informality, and development: Theory and evidence from brazil. *American Economic Review*, 108(8):2015–47, 2018.

World Bank. *World development report 2019: The changing nature of work*. The World Bank, 2018.

# 8 Tables

Table 1: Timeline of fare regulation

| Time | Description |
|---|---|
| 11 March 2019 | Safety regulation released. The Minister of Transportation stated that the Ministry is still working on the minimum fare regulation. He hinted it will be around IDR 2,400/km. |
| 25 March 2019 | The Ministry of Transportation released the fare regulation. |
| 1 May 2019 | Start of the implementation of the regulation. |
| 1 May 2019 | First phase of the fare regulation's implementation: Jakarta, Bandung, Yogyakarta, Surabaya, dan Makassar. |
| 1 July 2019 | Second phase of the fare regulation's implementationn: 41 cities. |
| 9 August 2019 | Third phase of the fare regulation's implementation: 123 cities. |
| 2 September 2019 | Fare regulation is implemented in all cities where ride-share platforms operate. |

*Source*: Minister of Transportation Decree No. 348 Year 2019 and various news sources. List of news sources are available upon request.

Table 2: Structure of the fare regulation

| Zone | per-km min-max fares | minimum total fare |
|---|---|---|
| Zone 1 | Rp 1,850-2,300 | Rp 7,000 - 10,000 |
| Zone 2 | Rp 2,000-2,500 | Rp 8,000 - 10,000 |
| Zone 3 | Rp 2,100-2,600 | Rp 7,000 - 10,000 |

*Source*: Minister of Transportation Decree No. 348 Year 2019.

Table 3: Average treatment effects on driver fare and number of transactions

| | log(Avg driver fare) | | log(N trips) | | log(Sum driver fare) | |
|---|---|---|---|---|---|---|
| | All services | Regulated | All services | Regulated | All services | Regulated |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Treat | 0.0461** | 0.1286*** | 0.0021 | -0.0976 | 0.0483 | 0.0310 |
| | (0.0177) | (0.0322) | (0.0829) | (0.0914) | (0.0751) | (0.0813) |
| Observations | 12,760 | 12,760 | 12,760 | 12,760 | 12,760 | 12,760 |
| $R^2$ | 0.93870 | 0.91247 | 0.98193 | 0.98331 | 0.98381 | 0.98426 |
| Within $R^2$ | 0.03673 | 0.14282 | $3.5 \times 10^{-6}$ | 0.00605 | 0.00189 | 0.00066 |
| Day fixed effects | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| City fixed effects | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

*Notes:* All dependent variables are in log. "Avg driver fare": The city-day average of the price drivers receive for a ride. "N trips": Number of trips per city per day. "Sum driver fare": City-day level aggregate of the driver fare. Point estimates are of average treatment effects, $\hat{\beta}_1$, as in equation 1. Standard errors are reported in parentheses. Regressions are run on a panel of cities over day as unit of time. All regressions include city fixed effects and time fixed effects, and standard errors clustered at the city level. Two-tailed significance: p<0.1*; p<0.05**; p<0.01***.

Table 4: Average treatment effects on driver daily earnings, and wage

| | log(Avg earnings/day) | | log(Avg wage) |
|---|---|---|---|
| | All services | Regulated | |
| | (1) | (2) | (3) |
| Treat | -0.0167 | 0.0199 | -0.0674 |
| | (0.0248) | (0.0347) | (0.0503) |
| Observations | 12,760 | 12,760 | 10,962 |
| $R^2$ | 0.92638 | 0.94523 | 0.82260 |
| Within $R^2$ | 0.00132 | 0.00134 | 0.00902 |
| Day fixed effects | ✓ | ✓ | ✓ |
| City fixed effects | ✓ | ✓ | ✓ |

*Notes:* All dependent variables are in log. "Avg earnings/day": City-day average of drivers' daily earnings from fares. "Avg wage": City-day average of drivers daily earnings, divided by their daily total available hours on the app. Point estimates are of average treatment effects, $\hat{\beta}_1$, as in equation 1. Standard errors are reported in parentheses. Regressions are run on a panel of cities over day as unit of time. All regressions include city fixed effects and time fixed effects, and standard errors clustered at the city level. Two-tailed significance: p<0.1*; p<0.05**; p<0.01***.

Table 5: Average treatment effects on driver supply hours, and trip duration

| | log(Sum supply hrs) (1) | log(Sum transaction hrs) (2) | log(Sum idle hrs) (3) |
|---|---|---|---|
| Treat | 0.0865* | 0.0169 | 0.2430*** |
| | (0.0500) | (0.0738) | (0.0886) |
| Observations | 10,962 | 12,760 | 10,912 |
| $R^2$ | 0.98475 | 0.98198 | 0.92888 |
| Within $R^2$ | 0.00996 | 0.00024 | 0.01580 |
| Day fixed effects | ✓ | ✓ | ✓ |
| City fixed effects | ✓ | ✓ | ✓ |

*Notes:* All dependent variables are in log. "Sum supply": City-day aggregate of drivers' daily total available hours on the app. "Sum transaction hrs": City-day aggregate of the total duration of all trips conducted. Point estimates are of average treatment effects, $\hat{\beta}_1$, as in equation 1. Standard errors are reported in parentheses. Regressions are run on a panel of cities over day as unit of time. All regressions include city fixed effects and time fixed effects, and standard errors clustered at the city level. Two-tailed significance: p<0.1*; p<0.05**; p<0.01***.

Table 6: Average treatment effects on the extensive margin market participation

| | log(N drivers) | | log(N customers) | | log(N customer/driver) | |
|---|---|---|---|---|---|---|
| | All services (1) | Regulated (2) | All services (3) | Regulated (4) | All services (5) | Regulated (6) |
| Treat | 0.0649 | 0.0111 | 0.0195 | -0.0630 | -0.0454* | -0.0741** |
| | (0.0605) | (0.0599) | (0.0749) | (0.0826) | (0.0262) | (0.0327) |
| Observations | 12,760 | 12,760 | 12,760 | 12,760 | 12,760 | 12,760 |
| $R^2$ | 0.98460 | 0.98515 | 0.98345 | 0.98470 | 0.88919 | 0.95239 |
| Within $R^2$ | 0.00494 | 0.00014 | 0.00035 | 0.00295 | 0.01159 | 0.02297 |
| Day fixed effects | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| City fixed effects | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

*Notes:* All dependent variables are in log. "N drivers": Number of distinct drivers at the city-day level. "N customers": Number of distinct customers at the city-day level. "N customer/driver": Number of distinct customers divided by the number of distinct drivers. Point estimates are of average treatment effects, $\hat{\beta}_1$, as in equation 1. Standard errors are reported in parentheses. Regressions are run on a panel of cities over day as unit of time. All regressions include city fixed effects and time fixed effects, and standard errors clustered at the city level. Two-tailed significance: p<0.1*; p<0.05**; p<0.01***.

Table 7: Average treatment effects on the intensive margin market participation

| | log(Avg n. bookings/driver) | | log(Avg n. bookings/customer) | |
|---|---|---|---|---|
| | All services | Regulated | All services | Regulated |
| | (1) | (2) | (3) | (4) |
| Treat | -0.0628* | -0.1087** | -0.0173 | -0.0346*** |
| | (0.0322) | (0.0409) | (0.0115) | (0.0116) |
| Observations | 12,760 | 12,760 | 12,760 | 12,760 |
| $R^2$ | 0.91089 | 0.95232 | 0.84742 | 0.83562 |
| Within $R^2$ | 0.01750 | 0.03698 | 0.00751 | 0.03185 |
| Day fixed effects | ✓ | ✓ | ✓ | ✓ |
| City fixed effects | ✓ | ✓ | ✓ | ✓ |

*Notes:* All dependent variables are in log. "Avg n. bookings/driver": Average number of daily trips per driver. "Avg n. bookings/customer: Average number of daily trips per customer. Point estimates are of average treatment effects, $\hat{\beta}_1$, as in equation 1. Standard errors are reported in parentheses. Regressions are run on a panel of cities over day as unit of time. All regressions include city fixed effects and time fixed effects, and standard errors clustered at the city level. Two-tailed significance: p<0.1*; p<0.05**; p<0.01***.

Table 8: Average treatment effects on customer expenditure and wait time

| | log(Avg expenditure/day) | | log(Avg wait distance) | |
|---|---|---|---|---|
| | All services | Regulated | All services | Regulated |
| | (1) | (2) | (3) | (4) |
| Treat | 0.0801* | 0.1957*** | -0.1488*** | -0.0759* |
| | (0.0408) | (0.0391) | (0.0472) | (0.0387) |
| Observations | 12,760 | 12,760 | 7,620 | 7,620 |
| $R^2$ | 0.82291 | 0.89509 | 0.85736 | 0.90391 |
| Within $R^2$ | 0.02451 | 0.17482 | 0.03352 | 0.01501 |
| Day fixed effects | ✓ | ✓ | ✓ | ✓ |
| City fixed effects | ✓ | ✓ | ✓ | ✓ |

*Notes:* All dependent variables are in log. "Avg expenditure/day": Average of total customer fare per customer per day, at the city-day level. "Avg wait distance": Average of the linear distance between the accepted driver and the pickup point per trip, at the city-day level. Point estimates are of average treatment effects, $\hat{\beta}_1$, as in equation 1. Standard errors are reported in parentheses. Regressions are run on a panel of cities over day as unit of time. All regressions include city fixed effects and time fixed effects, and standard errors clustered at the city level. Two-tailed significance: p<0.1*; p<0.05**; p<0.01***.

Table 9: Average treatment effects on productivity

| | log(Avg util. rate) (1) | log(Avg km/supp hr) (2) | log(Avg rides/supp hr) (3) |
|---|---|---|---|
| Treat | -0.0800 | -0.0919* | -0.1050* |
| | (0.0516) | (0.0521) | (0.0556) |
| Observations | 10,962 | 10,962 | 10,962 |
| $R^2$ | 0.75871 | 0.87826 | 0.83735 |
| Within $R^2$ | 0.00852 | 0.01600 | 0.02264 |
| Day fixed effects | ✓ | ✓ | ✓ |
| City fixed effects | ✓ | ✓ | ✓ |

*Notes:* All dependent variables are in log. Point estimates are of average treatment effects, $\hat{\beta}_1$, as in equation 1. "Avg util. rate": average utilization rate, i.e., the number of hours spent on paying trips divided by active hours spent on app, i.e., "supply hours." "Avg km/supp hr": average cumulative distance traveled on paying trips per supply hour. "Avg rides/supp hr": average daily number of rides per supply hour. Standard errors are reported in parentheses. Regressions are run on a panel of cities over day as unit of time. All regressions include city fixed effects and time fixed effects, and standard errors clustered at the city level. Two-tailed significance: p<0.1*; p<0.05**; p<0.01***.

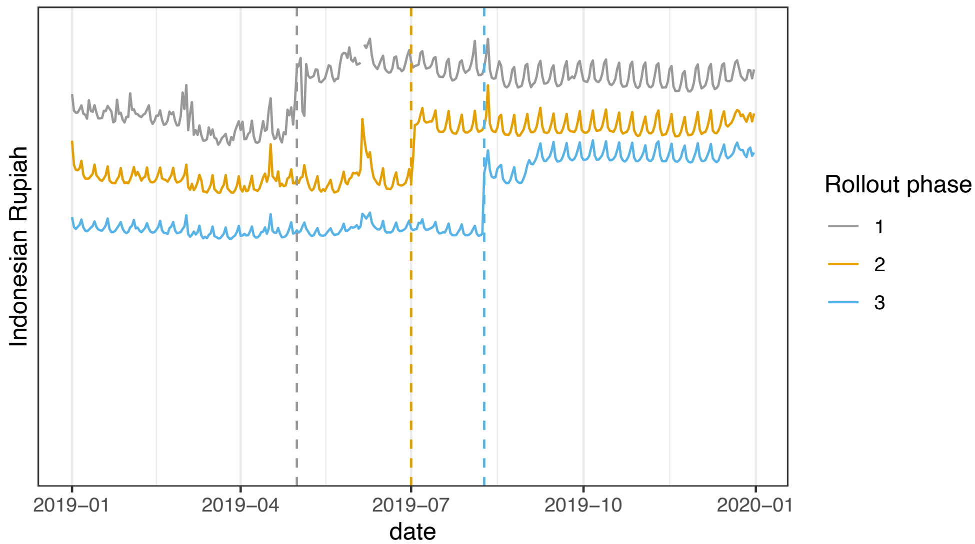
Table 10: Elasticities to driver fare and comparisons from Hall et al. (2021)

| | log(N trips) (1) | log(Sum supply hrs) (2) | log(Avg km/supp hr) (3) | log(Avg wage) (4) |
|---|---|---|---|---|
| log(Avg customer fare) | 0.0219 | | | |
| | (0.8594) | | | |
| log(Avg driver fare) | | 1.756 | -1.867** | -1.370 |
| | | (1.242) | (0.7973) | (0.8560) |
| Observations | 12,760 | 10,962 | 10,962 | 10,962 |
| $R^2$ | 0.98178 | 0.98004 | 0.84658 | 0.80637 |
| Within $R^2$ | -0.00832 | -0.29534 | -0.24009 | -0.08163 |
| Day fixed effects | ✓ | ✓ | ✓ | ✓ |
| City fixed effects | ✓ | ✓ | ✓ | ✓ |
| HHK estimates | -0.099 | 0.342*** | -0.655*** | 0.075 |
| HHK SEs | (0.081) | (0.034) | (0.059) | (0.064) |

*Notes:* All dependent variables are in log. Point estimates are elasticity 2SLS estimates based on Equation 15. Standard errors are reported in parentheses. Regressions are run on a panel of cities over day as unit of time. All regressions include city fixed effects and time fixed effects, and standard errors clustered at the city level. Two-tailed significance: p<0.1*; p<0.05**; p<0.01***.

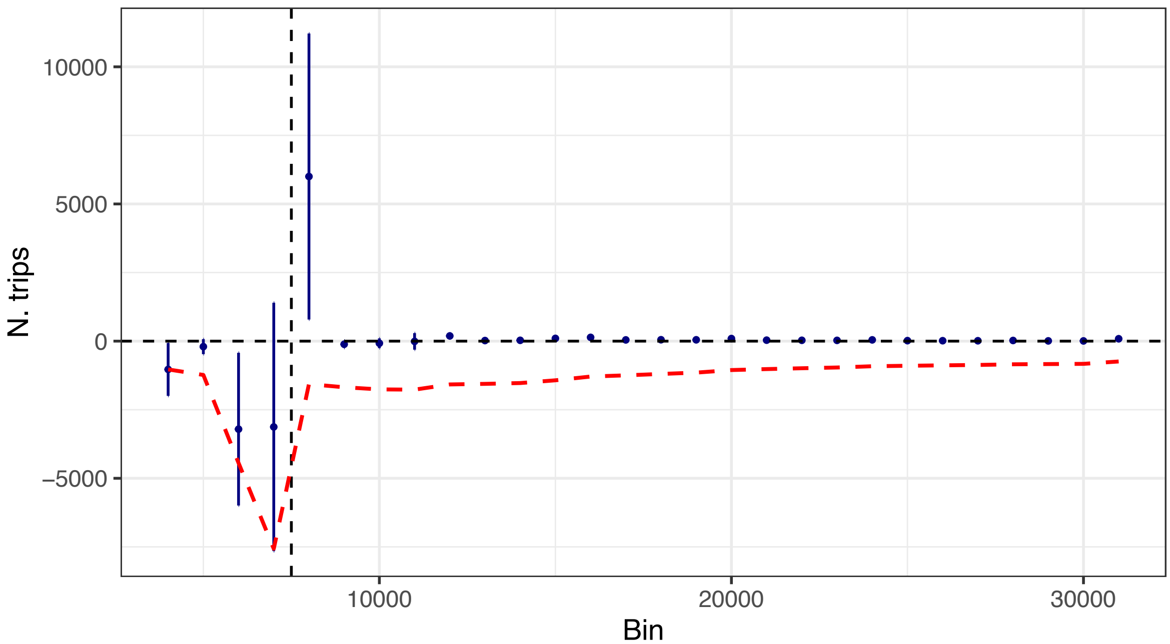
# 9 Figures

Figure 1: Average driver fare, by policy rollout phase




*Notes*: We plot the daily averages of per-trip fares received by drivers for each rollout phases. The vertical scale does not have value labels for confidentiality reasons, but the range of the axis starts from 0. Vertical dashed lines are the timing of policy rollout for the phases in the corresponding color. The first rollout phase was on 1 May 2019. The second rollout phase was on 1 July 2019. The third rollout phase was on 9 August 2019.

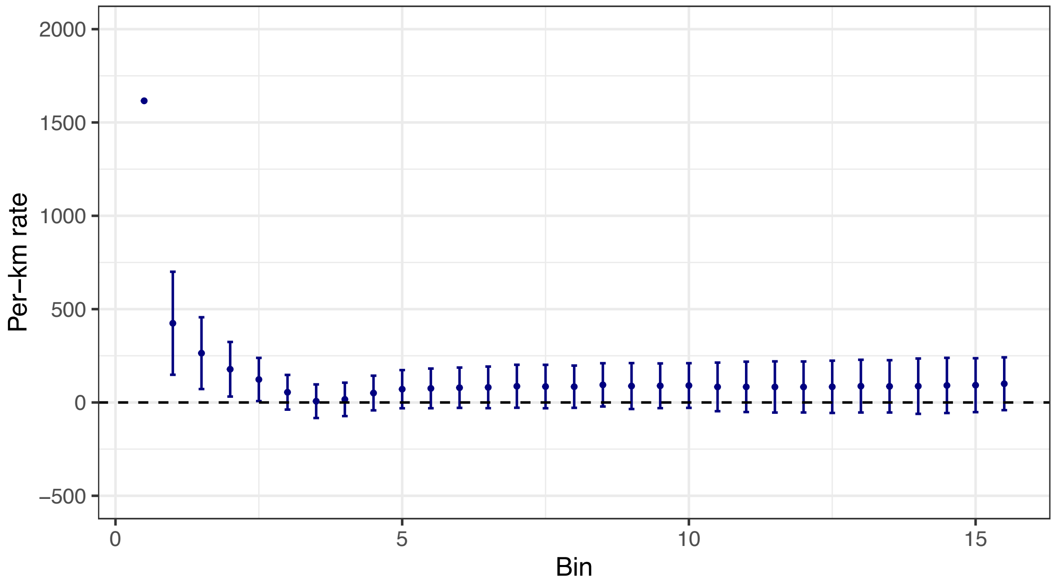
Figure 2: Treatment effect on frequency by 1,000-rupiah bins



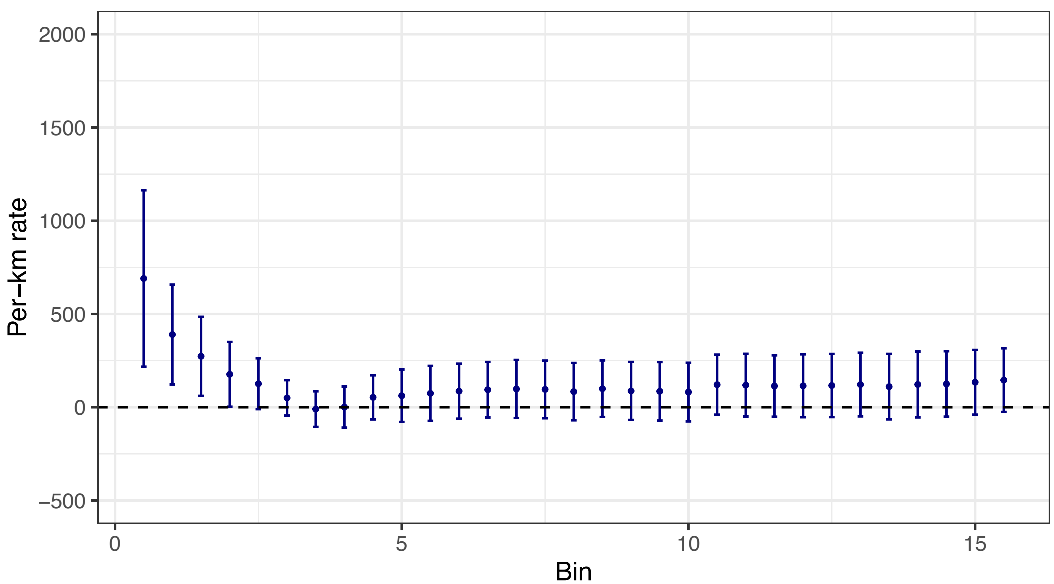

*Notes*: The dependent variable is the number of transactions in 1000-Rupiah bins. Point estimates in blue are the estimated impact on each bin, $\hat{\beta}^j$, as in equation 2. The 95% confidence intervals for coefficients are shown as the whisker bar in blue. The dashed red line shows the cumulative effect, i.e., the sum of the point estimates, from the left to right. Regression is run on a bins-city by day panel data. The model includes fare-bin-by-city fixed effects and fare-bins-by-day fixed effects. Standard errors are clustered at the city-bin level.

Figure 3: Treatment effects on average and median driver fare divided by distance driven, by the 500-meter trip distance bin

(a) Average per-km fare

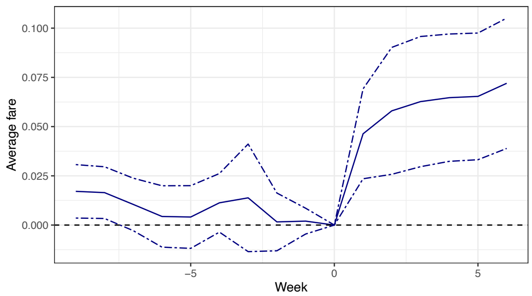



(b) Median per-km fare

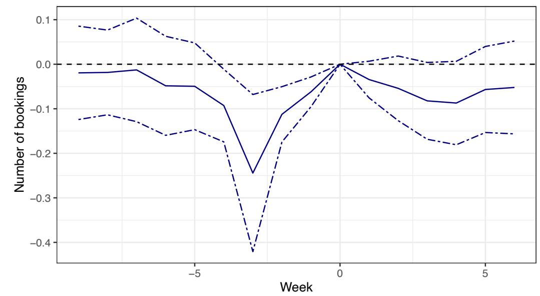



*Notes*: The dependent variables are the conditional average and median measures of total driver fare divided distance driven for subfigures (a) and (b), respectively. Point estimates in blue are the estimated impact for each 500-meter trip distance bin. The 95% confidence intervals for coefficients are shown as the whisker bar in blue. The model includes bin-by-city fixed effects and fare-bins-by-day fixed effects. Standard errors are clustered at the city-bin level..

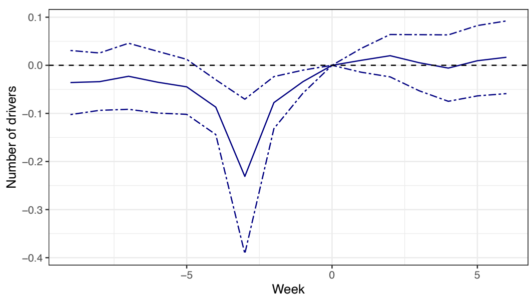
Figure 4: Weekly treatment effects on average driver fare and number of bookings



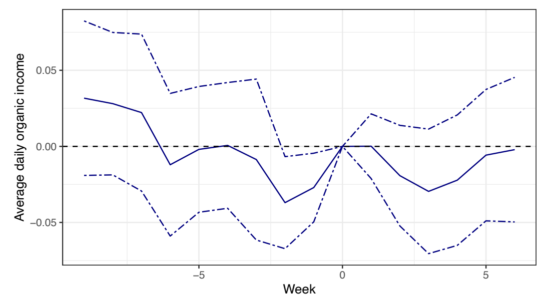

(a) Log(average driver fare)



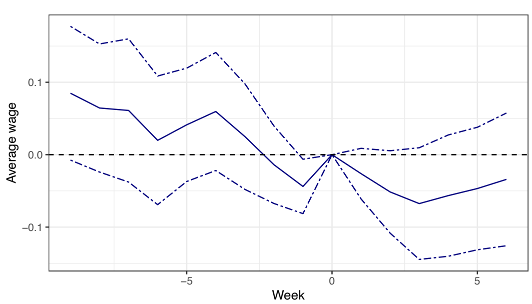

(b) Log(number of transactions)



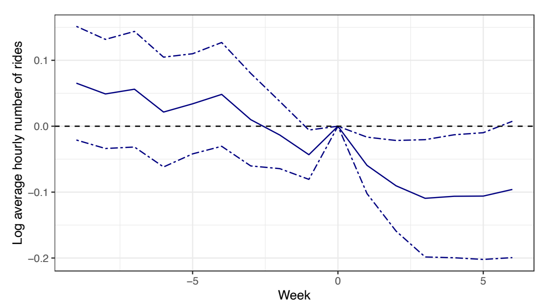

(c) Log(number of drivers)




(d) log(driver average daily revenue)




(e) log(average wage)

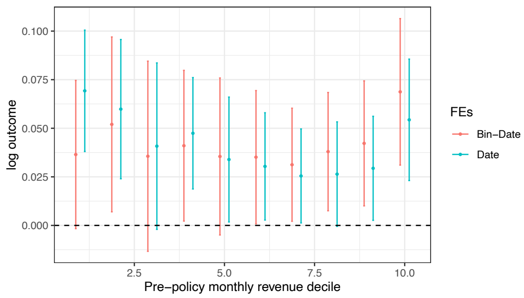



(f) log(average num. trips/supply hour)

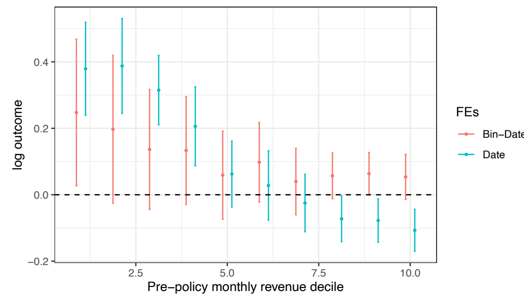
*Notes*: All dependent variables are in log. Point estimates, in solid lines, are the estimated weekly average treatment effects $\hat{\beta_1^k}$ , where $k$ corresponds to the weeks since measure on the horizontal axis, as in equation 6. The 95% confidence intervals for coefficients are shown as range bound by the dashed lines. Regressions are run on a panel of cities over day as unit of time. All regressions include city fixed effects and time fixed effects, and the errors clustered at the city level.

Figure 5: Conditional average treatment effects by drivers' total earning deciles

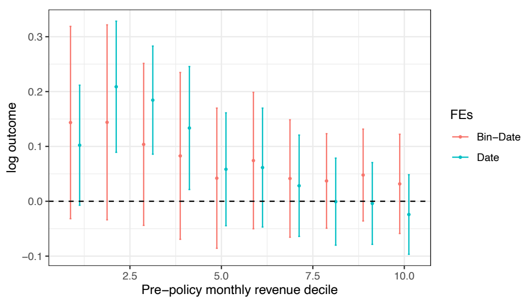
(a) Driver fare/trip



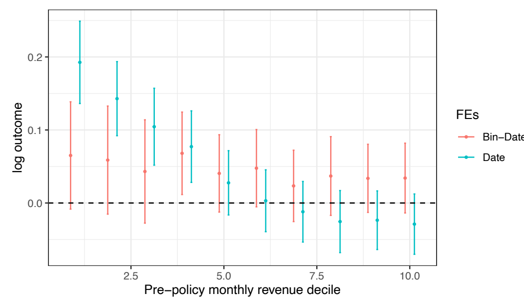

(b) Supply hours/decile



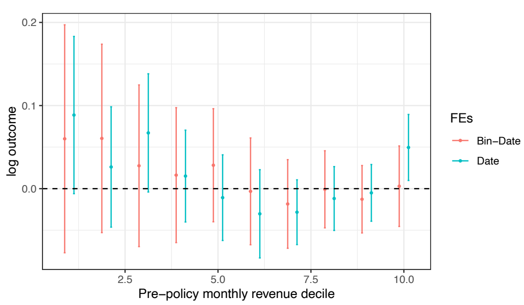

(c) N. distinct drivers/bin



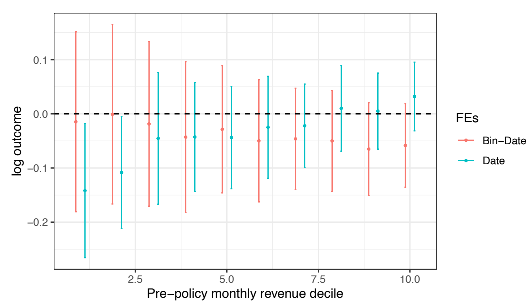

(d) Supply hours/driver




(e) Driver earnings/driver



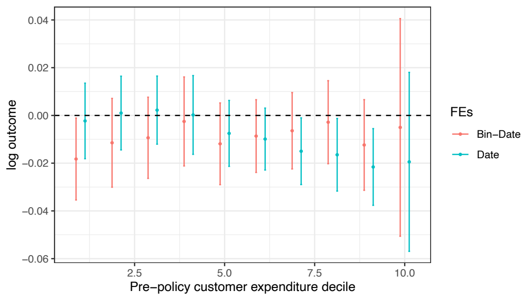

(f) Driver wage



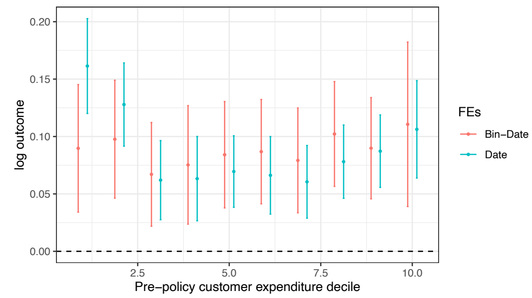

*Notes*: The dependent variables are the log-transformed outcomes listed in subfigure captions. The deciles are defined by the pre-policy transaction volume of drivers, defined as the sum of all driver fares over the pre-policy period of June 1 to June 30, 2019. Data are restricted to transactions from drivers who were present in the data as of June 30, 2019. Point estimates in blue are the estimated impact on each bin of pre-policy average daily transaction. The 95% confidence intervals for coefficients are shown as the whisker bar in blue. The model includes bin-by-city fixed effects and either fare-bins-by-day fixed effects or day fixed effects. Standard errors are clustered at the city-bin level. The sample is restricted to drivers who joined the platform by June 30, 2019.

Figure 6: Treatment effects by customers' pre-policy total expenditure deciles

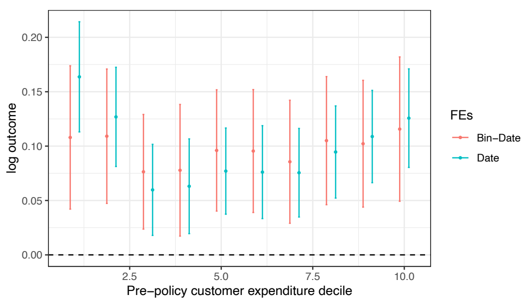
(a) Number of trips/customer



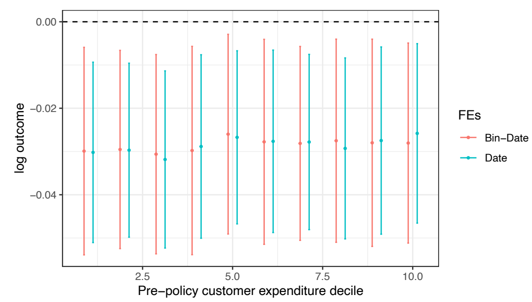

(b) Customer fare/trip




(c) Daily expenditure/customer



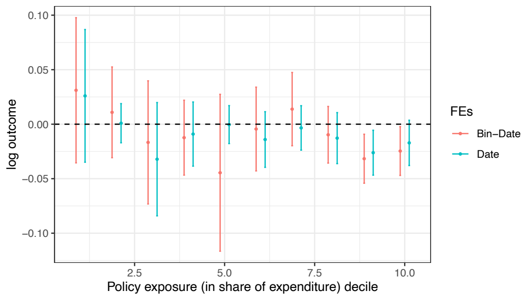

(d) Wait distance/trip



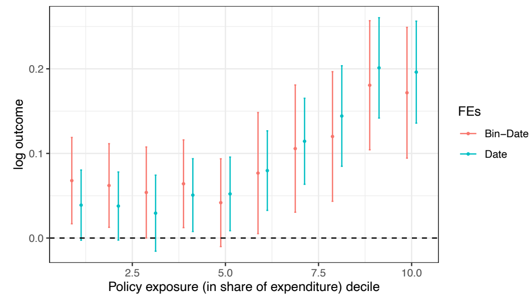

*Notes*: The dependent variables are the log-transformed outcomes listed in subfigure captions. The deciles are defined by the pre-policy customer expenditure, defined as the sum of all customer fare over the pre-policy period of June 1 to June 30, 2019. Data are restricted to transactions from drivers who were present in the data as of June 30, 2019. Point estimates in blue are the estimated impact on each bin of pre-policy average daily transaction. The 95% confidence intervals for coefficients are shown as the whisker bar in blue. The model includes bin-by-city fixed effects and fare-bins-by-day fixed effects. Standard errors are clustered at the city-bin level. The sample is restricted to customers who had made at least one order by June 30, 2019.

Figure 7: Conditional average treatment effects by customers' pre-policy exposure

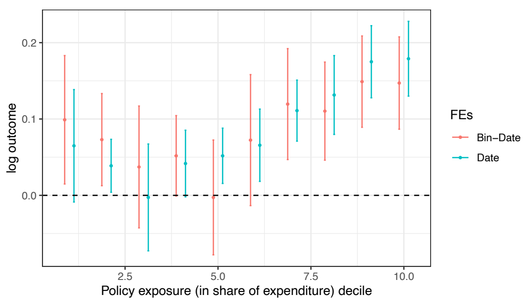
(a) Number of trips/customer



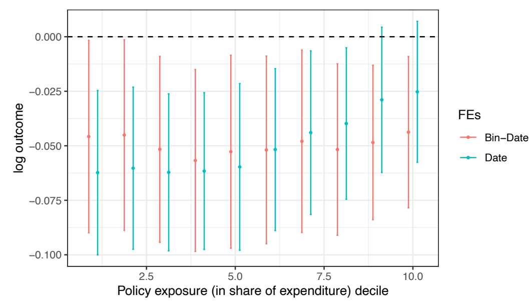

(b) Customer fare/trip




(c) Daily expenditure/customer

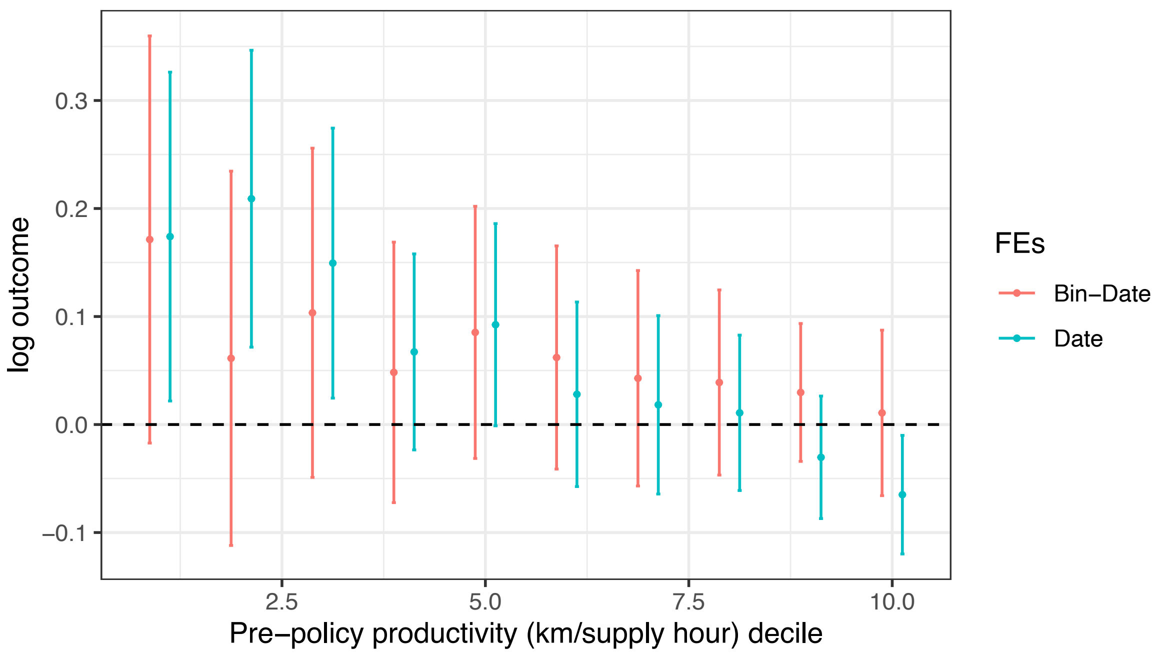



(d) Customer-average wait distance



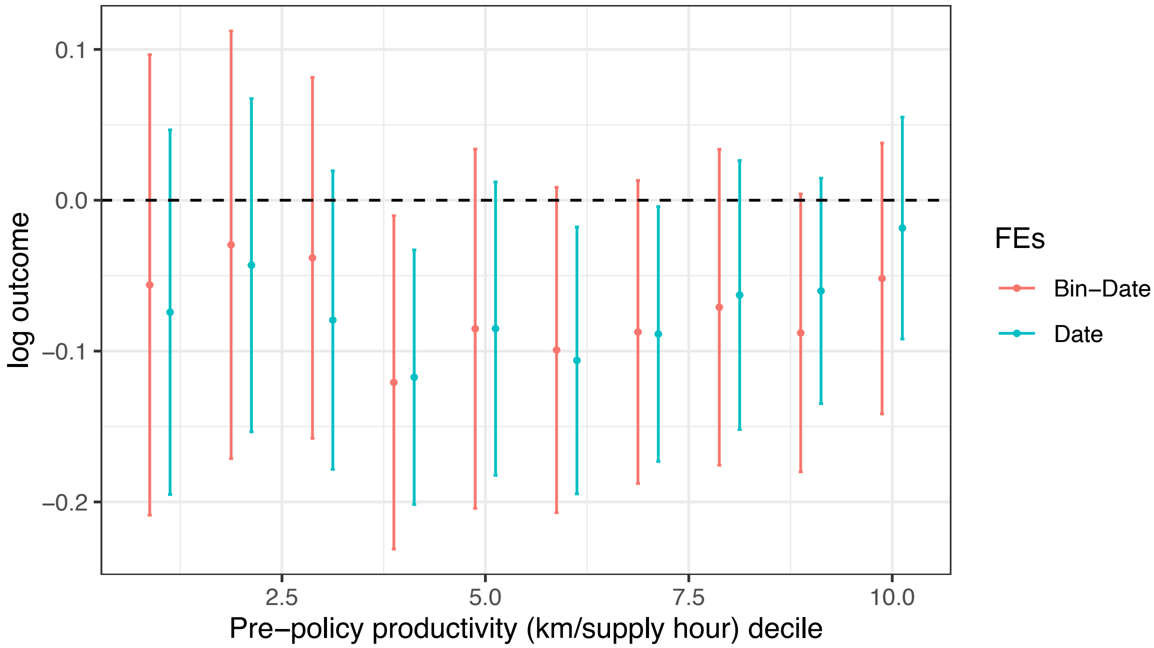

*Notes*: The dependent variables are the log-transformed outcomes listed in subfigure captions. The deciles are defined by the potential policy exposure of customers during the pre-policy period, defined as the share of the customer's monthly expenditure that would qualify for the policy. The pre-policy period is defined as June 1 to June 30, 2019. Point estimates in blue are the estimated impact on each bin of pre-policy average daily transaction. The 95% confidence intervals for coefficients are shown as the whisker bar in blue. The model includes bin-by-city fixed effects and fare-bins-by-day fixed effects. Standard errors are clustered at the city-bin level. The sample is restricted to customers who had made at least one order by June 30, 2019.

Figure 8: Conditional average treatment effects on the number of distinct drivers, by drivers' pre-policy productivity




*Notes*: The dependent variable is the log-transformed number of distinct drivers per city per day. The deciles are defined by the pre-policy productivity level of drivers, defined as distance driven on trips per supply hour. Pre-policy period is June 1 to June 30, 2019. Data are restricted to transactions from drivers who were present in the data as of June 30, 2019. Point estimates and the 95% confidence intervals are shown as dots and whiskers. The models include city-bin fixed effects, as well as either bin-date or date fixed effects. Standard errors are clustered at the city-bin level.

Figure 9: Conditional average treatment effects on productivity, by drivers' pre-policy productivity




*Notes*: The dependent variable is the log-transformed average distance driven on trips per supply hour per city-bin-day. The deciles are defined by the pre-policy productivity level of drivers, defined as distance driven on trips per supply hour. Pre-policy period is June 1 to June 30, 2019. Data are restricted to transactions from drivers who were present in the data as of June 30, 2019. Point estimates and the 95% confidence intervals are shown as dots and whiskers. The models include city-bin fixed effects, as well as either bin-date or date fixed effects. Standard errors are clustered at the city-bin level.

# A Policy context

## A.1 Fare regulation's implementation by city

**Phase 1 (5 cities):** Jakarta, Bandung, Yogyakarta, Surabaya, Makassar

**Phase 2 (41 cities):**

- Zone 1: Banda Aceh, Kota Medan, Kota Batam, Kota Pekanbaru, Kota Palembang, Kota Bandar Lampung, Kota Metro, Kota Belitung, Kota Bandung, Kota Semarang, Kota, Solo, Kota Yogyakarta, Kota Surabaya, Kota Denpasar, Kab. Probolinggo, Kab. Pasuruan, Kab. Kudus, Madura.

- Zone 2: Jakarta, Kab. Bogor, Kota Bogor, Kota Depok, Kab. Tangerang, Kota Tangerang, Kota Tangerang Selatan, Kab. Bekasi, Kota Bekasi.

- Zone 3: Kota Pontianak, Kota Palangkaraya, Kota Samarinda, Kota Balikpapan, Kota Banjarmasin, Kota Mataram, Kota Kupang, Kota Manado, Kota Gorontalo, Kota Palu, Kota Makassar, Kota Kendari, Kota Ambon, Kota Jayapura.

**Phase 3 (88 cities/districts):**

- Zone 1: Kota Sabang, Kota Bukittinggi, Kabupaten Agam, Kabupaten Lima Puluh Kota, Kabupaten Tanah Datar, Kota Padang Panjang, Kota Payakumbuh, Kota Duri, Kabupaten Bengkalis, Kota Tanjung Pinang, Kota Jambi, Kabupaten Muaro Jambi, Kabupaten Kisaran, Kabupaten Asahan, Kabupaten Karo, Kabupaten Toba Samosir, Kota Tanjung Balai, Kota Padangsidempuan, Kota Padang Lawas Utara, Kabupaten Tapanuli Selatan, Kabupaten Serdang Bedagai, Kota Pematangsiantar, Kabupaten Simalungun, Kota Tebing Tinggi, Kota Rantau Prapat, Kabupaten Labuhan Batu, Kabupaten Batang, Kabupaten Cilacap, Kabupaten Kebumen, Kabupaten Banyumas, Kabupaten Brebes, Kabupaten Purworejo, Kota Pekalongan, Kabupaten Pekalongan, Kabupaten Pemalang, Kabupaten Banjarnegara, Kabupaten Purbalingga, Kota Salatiga, Kabupaten Banyuwangi, Kabupaten Bojonegoro, Kabupaten Jember, Kabupaten Bondowoso, Kabupaten Jombang, Kabupaten Kediri, Kota Kediri, Kabupaten Nganjuk, Kota Madiun, Kabupaten Magetan, Kabupaten Ngawi, Kabupaten Ponorogo, Kota Mojokerto, Kabupaten Mojokerto, Kota Serang, Kabupaten Lebak, Kota Cirebon, Kabupaten Cirebon, Kabupaten Garut, Kabupaten Indramayu, Kabupaten Kuningan, Kabupaten Majalengka, Kota Tasikmalaya, Kabupaten Tasikmalaya, Kabupaten Subang, Kota Sukabumi, Kabupaten Sukabumi, Kabupaten Cianjur, Kabupaten Purwakarta, Kabupaten Sumedang, Kabupaten Ciamis, Kabupaten Pangandaran, Kota Banjar, Kota Malang, Kabupaten Malang, Kota Batu, Kota Tegal, Kabupaten Tegal, Kabupaten Demak, Kabupaten Kendal, Kabupaten Pati, Kabupaten Jepara

- Zone 3: Kota Bitung, Kota Tomohon, Kota Palopo, Kota Tarakan, Kota Ternate, Kota Sorong, Kabupaten Merauke, Kota Pare-Pare

**Phase 4:** all cities and districts where ride-share platforms operate.

# B Additional results

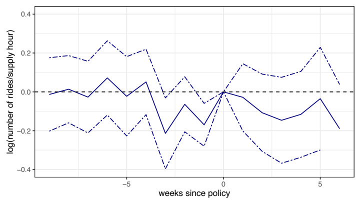
## B.1 Effects over time

One concern is that the exogenous price shocks may have differential short-term vs is the long-term. If it is so and, for instance, the causal effect of the policy diminishes over time, the estimates we present may not be relevant for the effect in equilibrium. Appendix Section 5.2.2 addresses this concern, by showing that the causal effects of policy are significant and persistent for the 6 weeks in which we have policy variation for the price and productivity outcomes. Furthermore, we argue in Section 1 that the estimate we identify *are* the equilibrium effects because of short contract duration and low cost of supplying labor on the platform once registered. Yet we cannot rule out that drivers and customers adjust their responses over time.

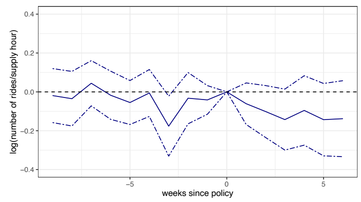
This question may be particularly relevant regarding the policy effect, which we show to be driven by both entry of lower productivity driver and reduction in individual productivity in Section 5.4. It is possible that less productive drivers, who now work more, may learn to improve their productivity as they work more, therefore diminishing the negative impact over time. To identify if such dynamic effects exist, we estimate Equation 6 from Appendix Section 5.2.2 separately for each decile of the pre-policy driver productivity.

Figure B.1 shows the result. We find that the negative productivity effect is noisily estimated and statistically indistinguishable from 0 for the first and second deciles. For higher deciles, however, we observe negative and persistent effects on productivity. If the aforementioned mechanism of productivity adjustment were to take place, then we would expect that estimates for lower deciles be positive in the later weeks post policy introduction. We do not observe such a trend.

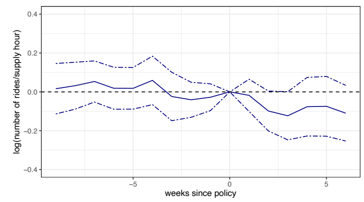
Figure B.1: Weekly treatment effects on average driver productivity, by pre-policy productivity (km driven per supply hour) decile



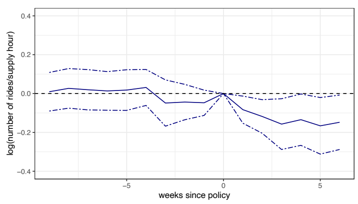

(a) 1st decile



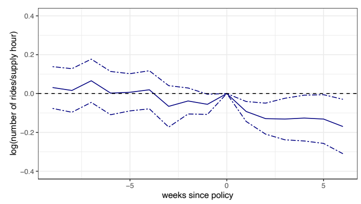

(b) 2nd decile



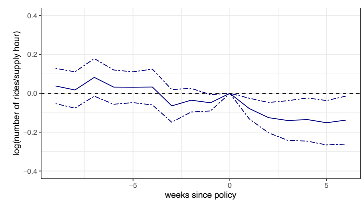

(c) 3rd decile



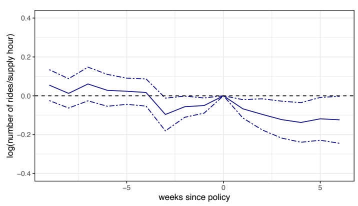

(d) 4th decile



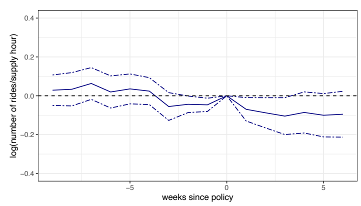

(e) 5th decile



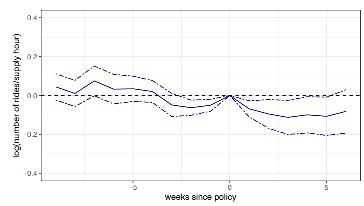

(f) 6th decile



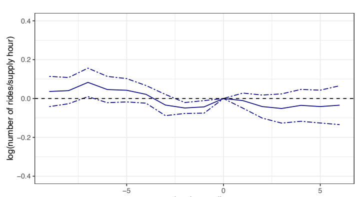

(g) 7th decile




(h) 8th decile



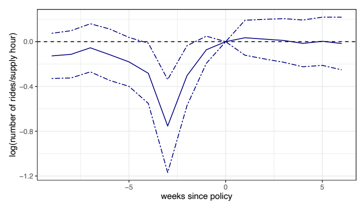

(i) 9th decile



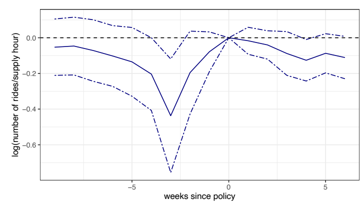

(j) 10th decile

*Notes*: All dependent variables are in log. Point estimates, in solid lines, are the estimated weekly average treatment effects $\hat{\beta_1^k}$ , where $k$ corresponds to the weeks since measure on the horizontal axis, as in equation 6. The 95% confidence intervals for coefficients are shown as range bound by the dashed lines. Regressions are run on a panel of cities over day as unit of time. All regressions include city fixed effects and time fixed effects, and the errors clustered at the city level.

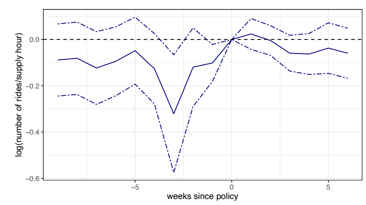
Figure B.2: Weekly treatment effects on the number of distinct drivers, by pre-policy productivity (km driven per supply hour) decile



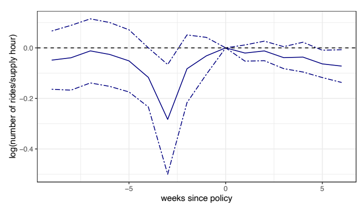

(a) 1st decile



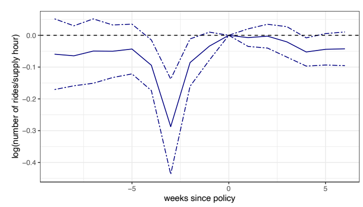

(b) 2nd decile



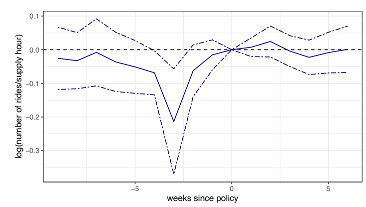

(c) 3rd decile



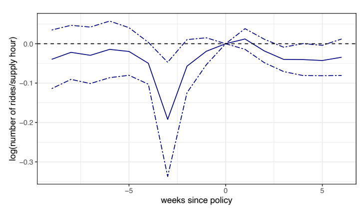

(d) 4th decile



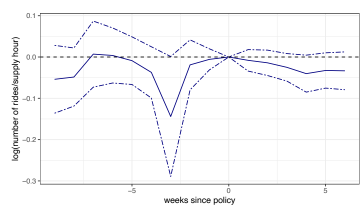

(e) 5th decile



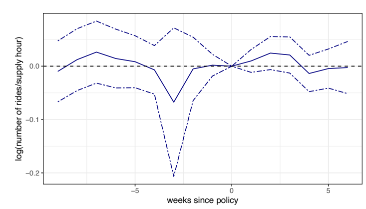

(f) 6th decile



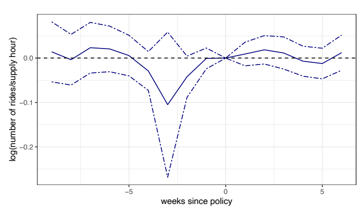

(g) 7th decile




(h) 8th decile




(i) 9th decile




(j) 10th decile

*Notes*: All dependent variables are in log. Point estimates, in solid lines, are the estimated weekly average treatment effects $\hat{\beta_1^k}$ , where $k$ corresponds to the weeks since measure on the horizontal axis, as in equation 6. The 95% confidence intervals for coefficients are shown as range bound by the dashed lines. Regressions are run on a panel of cities over day as unit of time. All regressions include city fixed effects and time fixed effects, and the errors clustered at the city level.

## B.2 Average treatment effects on trip attributes

Table B.1: Average treatment effects on driver fare and number of transactions

| | log(Avg driver fare) (1) | log(Avg distance) (2) | log(Avg driver fare/km) (3) |
|---|---|---|---|
| Treat | 0.0461** | 0.0142 | 0.0393 |
| | (0.0177) | (0.0165) | (0.0298) |
| Observations | 12,760 | 12,760 | 12,760 |
| $R^2$ | 0.93870 | 0.89941 | 0.56430 |
| Within $R^2$ | 0.03673 | 0.00252 | 0.00124 |
| Day fixed effects | ✓ | ✓ | ✓ |
| City fixed effects | ✓ | ✓ | ✓ |

*Notes:* All dependent variables are in log. Point estimates are of average treatment effects, $\hat{\beta}_1$, as in equation 1. Standard errors are reported in parentheses. Regressions are run on a panel of cities over day as unit of time. All regressions include city fixed effects and time fixed effects, and standard errors clustered at the city level. Two-tailed significance: p<0.1*; p<0.05**; p<0.01***.

## B.3 Adjustment mechanism: Policy spillovers to unregulated segment

In Section 5.1.1, we show that the policy shifted the distribution of driver fare from below-threshold levels and higher fare toward the new threshold. This distributional shift underlines the effect on the total number of transactions, on which we estimated small yet noisy treatment effect in Section 5.1.

We find suggestive evidence that some of the null effects are driven by spillovers to unregulated, non-taxi service, including food and other delivery services. First, Tables B.2 and B.3 show that reduction in the number of taxi bookings is coincided with an increase in non-taxi counterparts, which are not regulated by the minimum fare policy. However, the estimates for th enon-taxi segments are noisily estimated and cannot be distinguished from zero.

Second, Figure B.3 shows that the changes in distribution of fares had different patterns for taxi and non-taxi segments. For non-taxi segments, the minimum fare policy does not reduce transaction quantities on lower price ranges, yet only increases it at the policy threshold that should only apply to taxis. Although the effect is attenuated by reduction in transactions at higher fares, the figure shows an increase in transaction for the non-taxi segment without any significant losses in price ranges below the minimum-fare threshold. This pattern is indicative of positive spillover effects in the unregulated segment, perhaps as a shift in transactions from the regulated counterpart. Although effects of the overall city-level effect on the number of transactions by service type are still imprecisely estimated, these point estimates and changes to the distribution of fares by service type suggest that certain types of taxi trips were replaced by comparable non-taxi ones.

The shift in transactions from taxi to non-taxi service types is corroborated by Figure B.4, which shows changes in distributions of transactions across trip distance, separated by service type. The figure shows that a reduction in a mass of short-distance taxi trips is more than made up by an increase in non-taxi trips of similar distance. This would be in line with the idea that taxi trips of certain attributes were switched by non-taxi of the same attributes, such as switching to food delivery from a restaurant instead of visiting it.

Table B.2: Average treatment effects on average price by service type

| | log(Avg driver fare) | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| Service type | Taxi | Non-taxi | All |
| Treat | 0.1286*** | -0.0107 | 0.0461** |
| | (0.0322) | (0.0095) | (0.0177) |
| Observations | 12,760 | 12,760 | 12,760 |
| $R^2$ | 0.91247 | 0.93400 | 0.93870 |
| Within $R^2$ | 0.14282 | 0.00177 | 0.03673 |
| Day fixed effects | ✓ | ✓ | ✓ |
| City fixed effects | ✓ | ✓ | ✓ |

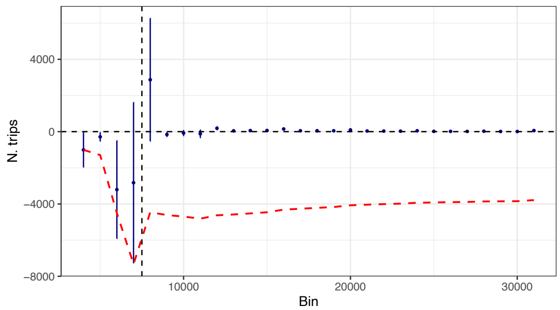
*Notes:* All dependent variables are in log. Point estimates are of average treatment effects, $\hat{\beta}_1$, as in equation 1. Standard errors are reported in parentheses. Regressions are run on a panel of cities over day as unit of time. All regressions include city fixed effects and time fixed effects, and standard errors clustered at the city level. Two-tailed significance: p<0.1*; p<0.05**; p<0.01***.

Table B.3: Average treatment effects on average price by service type

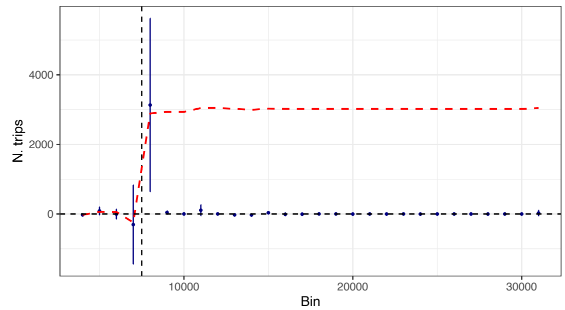
| | log(N trips) | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| Service type | Taxi | Non-taxi | All |
| Treat | -0.0976 | 0.0516 | 0.0021 |
| | (0.0914) | (0.0747) | (0.0829) |
| Observations | 12,760 | 12,760 | 12,760 |
| $R^2$ | 0.98331 | 0.97862 | 0.98193 |
| Within $R^2$ | 0.00605 | 0.00213 | $3.5 \times 10^{-6}$ |
| Day fixed effects | ✓ | ✓ | ✓ |
| City fixed effects | ✓ | ✓ | ✓ |

*Notes:* All dependent variables are in log. Point estimates are of average treatment effects, $\hat{\beta}_1$, as in equation 1. Standard errors are reported in parentheses. Regressions are run on a panel of cities over day as unit of time. All regressions include city fixed effects and time fixed effects, and standard errors clustered at the city level. Two-tailed significance: p<0.1*; p<0.05**; p<0.01***.

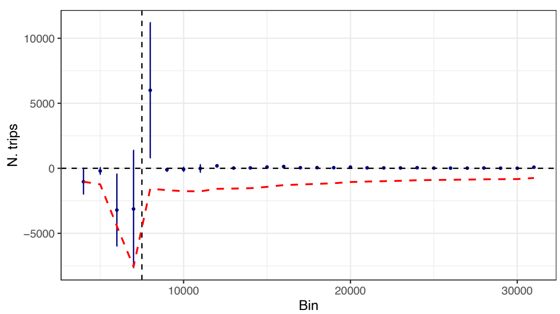
Figure B.3: Treatment effect on frequency by 1,000-rupiah bins, by service type




(a) Taxi



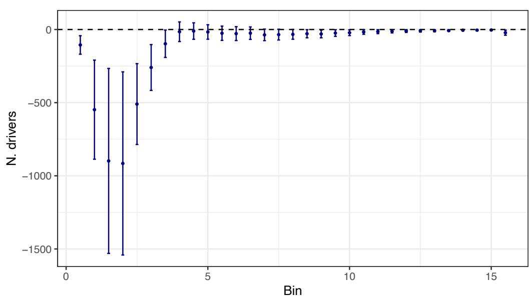

(b) Non-taxi



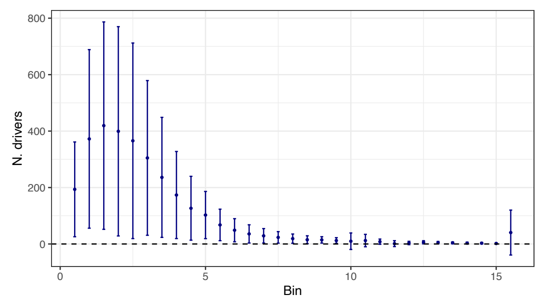

(c) All job types

*Notes*: The data is restricted to taxi trips and non-taxi trips for subpanels a and b, respectively. The result on subpanel C is for all job types, and is identical to Figure 2. The dependent variable is the number of transaction in 1000-Rupiah bins. Point estimates in blue are the estimated impact on each bin, $\hat{\beta}^j$, as in equation 2. The 95% confidence intervals for coefficients are shown by the range plots in blue. The dashed red line illustrates the cumulative effects, i.e., the sum of the point estimates. Regression is run on a panel of bins-city over day. The model includes fare-bin-by-city fixed effects and fare-bins-by-day fixed effects. Standard errors are clustered at the product of zones and policy rollouts.

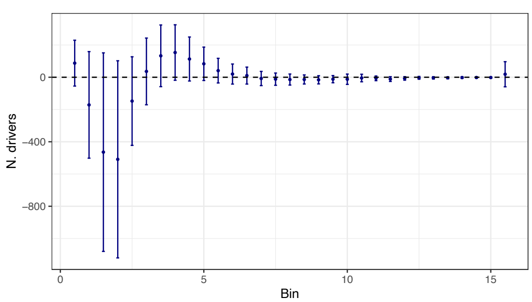
Figure B.4: Treatment effect on frequency by 500-meter distance bins, by service type




(a) Taxi




(b) Non-taxi




(c) All job types

### B.3.1 New driver entries and permanent exits

In Section 5.2, we find a noisy yet positive impact of the policy on the number of workers per day. This effect may be driven by new driver entries to the platform, increased days worked of the existing drivers, or both. The extent of new worker entries also matters for the analysis on distributional impact and productivity in Sections 5.3 and 5.4, where we condition the data to drivers who worked in June, 2019.

We assess the magnitude of new worker entries and permanent exits by constructing a city-month level panel data of driver entries and exits. Driver entry and exit dates are defined as the minimum and maximum dates of the timestamps on their transactions. We define the entry month as the first month in which drivers had their first trip, and the exit month as the month in which they conducted their last trip, plus 1. We construct a data set covering February (the first month in which we can observe entry or exit) to July, 2019. We convert the entry and exit variables to the share of the number of drivers of a given city in June, 2019.

We find statistically insignificant effects on new driver entry or permanent exit of that the minimum-fare policy. Table B.4 shows the results. The effect sizes on new entry are 1.7 percentage points relative to the pre-policy fleet size (i.e., the number of distinct drivers in operation). The effect size on the log outcome indicates that this effect corresponds to a 22.1% increase, i.e., a large percentage increase off a small value. We refrain from interpretating the effect sizes of the exiters, as they have inconsistent signs between the share- and log-share outcomes.

## B.4 Passthrough of driver-cut increases on customer fare

We investigate the extent to which the effects on transactions and spillovers are driven by the demand side response. First, we note that the exogenous increase in drivers' cut of the fare

Table B.4: Average treatment effects on driver entry and exit

| | Share(entrants)<br>(1) | Share(exiters)<br>(2) | log(Share(entrants))<br>(3) | log(Share(exiters))<br>(4) |
|---|---|---|---|---|
| Treat | 0.0172 | 0.0033 | 0.2213 | -0.0871 |
| | (0.0186) | (0.0070) | (0.1924) | (0.1117) |
| Observations | 348 | 295 | 348 | 295 |
| $R^2$ | 0.70367 | 0.67418 | 0.68840 | 0.83072 |
| Within $R^2$ | 0.00463 | 0.00023 | 0.00777 | 0.00248 |
| Month fixed effects | ✓ | ✓ | ✓ | ✓ |
| City fixed effects | ✓ | ✓ | ✓ | ✓ |

*Notes:* Point estimates are of average effects. Standard errors are reported in parentheses. Regressions are run on a panel of cities over month as unit of time. All regressions include city fixed effects and month fixed effects, and standard errors clustered at the city level. Two-tailed significance: p<0.1*; p<0.05**; p<0.01***.

leads to an increase in the amount billed to the consumer. This is not a given in ridesharing apps and other two-sided markets, where the platform operators may opt to absorb price shocks in one side (i.e. drivers) of the market so as not to affect the other (ii.e. customers). Table B.5 shows that the the policy increased both the drivers' cut as well as fare faced by the customers, and that these increases only occur in the taxi segment. In this appendix section we show that the pass-through of driver-cut increases on consumer price is greater than 100%, likely indicating complete pass-through plus administrative costs charged by the platform.

Table B.5: Average treatment effects on customer- and driver-fare

| | log(Avg driver fare) (1) | log(Avg customer fare) (2) |
|---|---|---|
| Treat | 0.0461** | 0.0974** |
| | (0.0176) | (0.0447) |
| Observations | 12,760 | 12,760 |
| $R^2$ | 0.93644 | 0.82979 |
| Within $R^2$ | 0.03547 | 0.03517 |
| Week fixed effects | ✓ | ✓ |
| DoW fixed effects | ✓ | ✓ |
| City fixed effects | ✓ | ✓ |

*Notes:* All dependent variables are in log. Point estimates are of average treatment effects, $\hat{\beta}_1$, as in equation 1. Standard errors are reported in parentheses. Regressions are run on a panel of cities over day as unit of time. All regressions include city fixed effects and time fixed effects, and standard errors clustered at the city level. Two-tailed significance: p<0.1*; p<0.05**; p<0.01***.

We estimate the passthrough rate using the two-stage least squares (2SLS) regressions. Using exogenous variation in the driver fare, we estimates the passthrough rate of changes in driver cut to customer fare, by instrumenting the driver fare ($ln(P^{driver}_{c,t})$) with the exogenous policy variation ($I_{c,t}(c \in Treat, t > 0)$) in the following equation:

$$ln(P^{customer}_{c,t}) = \beta_0 + \beta_1 * ln(P^{driver}_{c,t}) + \gamma_c + \rho_t + \epsilon_{c,t} \tag{13}$$

Table B.6 shows that the for taxi segment, a 10% increase in driver cut amount is associated with 18% increase in customer fare. The greater than 100% passthrough likely accounts for platform fees and/or additional discounts (in credits and free rides) that are given to customers. The passthrough rate is negative but imprecisely estimated for non-taxi segments.

Table B.6: **Elasticity of customer fare to driver cut**

| | log(Avg customer fare) | |
|---|---|---|
| | All services | Regulated |
| | (1) | (2) |
| log(Avg driver fare) | 2.111*** | 1.791*** |
| | (0.7822) | (0.3512) |
| | | |
| Observations | 12,760 | 12,760 |
| $R^2$ | 0.88062 | 0.90666 |
| Within $R^2$ | 0.29631 | 0.28041 |
| | | |
| Day fixed effects | ✓ | ✓ |
| City fixed effects | ✓ | ✓ |

*Notes:*

## B.5 Elasticity estimates

We estimate the elasticities of key outcome variables to price via two-stage least squares (2SLS) regressions. We use the policy variation variable to instrument for the price. We use the relevant price measures depending on the outcome; for outcomes that we consider to be "demand side," we use the prices customers face. The "demand-side" outcome measure for our analysis is the number of trips. For the "supply-side" measures, i.e., driver supply hours, productivity, an average wage, we use the driver fare as price.

For the demand-side elasticity measure, we estimate the relationship between the number of trips and the price consumers face by instrumenting log average customer fare ($ln(P_{c,t}^{customer})$) with the exogenous policy variation ($I_{c,t}(c \in Treat, t > 0)$). The second-stage equation for the two-stage least-squares estimator is as follows, in which we regress the log of transaction volumes, $ln(Q_{c,t})$ on the instrumented average customer fare in log:

$$ln(Q_{c,t}) = \beta_0 + \beta_1 * ln(P_{c,t}^{customer}) + \gamma_c + \rho_t + \epsilon_{c,t} \tag{14}$$

Similarly, we estimate supply-side elasticity measures to the price they face, i.e., driver fare. We instrument log average driver fare ($ln(P_{c,t}^{driver})$) with the exogenous policy variation ($I_{c,t}(c \in Treat, t > 0)$) for the following equation, with the log of average supply hours, $ln(SupplyHr_{c,t})$, as the outcome variable:

$$ln(SupplyHr_{c,t}) = \beta_0 + \beta_1 * ln(P_{c,t}^{driver}) + \gamma_c + \rho_t + \epsilon_{c,t}. \tag{15}$$

## B.6 Results from synthetic control-based inference

We find that the synthetic control-based inference procedure described in Section 4.2 yields similar results as the difference-in-differences counterpart. Table B.7 shows the results on our six main logged outcomes: average driver fare, number of trips, number of drivers, total supply hour, average driver daily earnings, and productivity (average distance driven on trips/supply hour). The dataset is the city-day panel, and the estimates are comparable to Tables 3, 4, 5, 6, and 9. The second and third columns show the pre- and post-policy differences between the weighted treatment units and their synthetic controls. The fourth column shows the p-values based on RSMRE rankings. The differences between the values on the third and second columns can be interpreted as the difference-in-differences estimate between the treatment and placebos, and the fourth column has its p-value.

Table B.7 shows results that are consistent with our difference-in-difference estimates. We find a 7.3% increase in average driver fare but do not find statistically significant effects on the number of trips, drivers, and their driver earnings. We find an 11.3% reduction in driver productivity. One notable difference, however, is that the effects on total driver supply hours is insignificant for the synthetic control-based analysis, with a treatment effect of around 1.0%.

Figure B.5 shows the time-series plots of weighted treatment effects and their placebos from Day 101 (April 11, 2019) to Day 216 (August 8, 2019). We find that, as we see in Table B.7, the differences between weighted treated outcomes and their synthetic controls (i.e., the pre-policy values of the red line) are greater than zero. This may be due to the weekly cyclicality or other daily variations in the outcomes data, where the optimally weighted synthetic control was biased. As such, it seems from the figure that the causal estimate should be the differences in deviations between weighted treated units and the placebos post- and pre-policy.

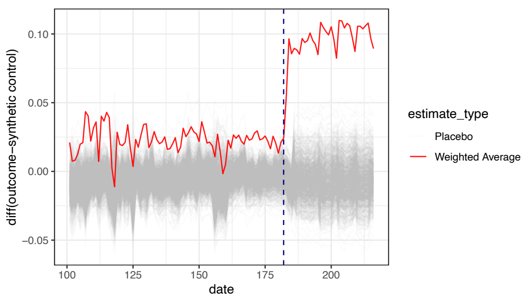
Figure B.5 also confirms the instantaneous and persistent effects on average driver fare, as shown in the raw data in Figure 1 and in distributed lag DiD model in Figure 4. We also find that for outcomes for which the treatment estimates are less statistically significant, the differences between pre- and post-treatment levels are less stark than for statistically significant estimates.

Table B.7: **Average treatment effects on driver fare and number of transaction**

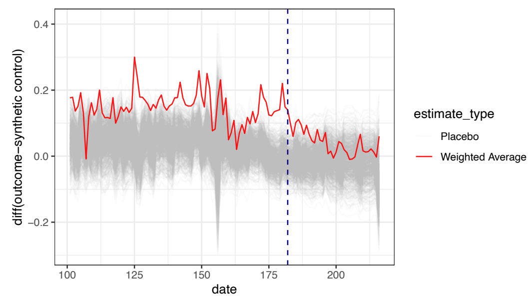
| | Pre-policy difference | Post-policy difference | P-value (RMSRE-based) |
|---:|---:|---:|---:|
| log(Avg driver fare) | 0.022 | 0.095 | 0.016 |
| log(N trips) | 0.149 | 0.043 | 0.820 |
| log(N drivers) | 0.118 | 0.079 | 0.559 |
| log(Total supply hours) | 0.132 | 0.142 | 0.245 |
| log(Avg daily income) | 0.047 | 0.019 | 0.918 |
| log(Avg km/supply hr) | 0.018 | -0.095 | 0.040 |

*Notes:* All dependent variables are in log. Outcome variables are listed on the first column, with corresponding estimates and p-values on the second to fourth columns. The second and third columns report the pre- and post-policy differences between the weighted treatment effects and their synthetic controls. The fourth column shows the p-values based on RSMRE rankings of the treatment effect, relative to their "placebo" counterparts.

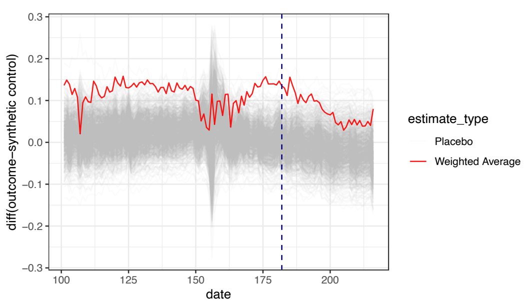
Figure B.5: **Synthetic control-based estimates and placebo simulations**



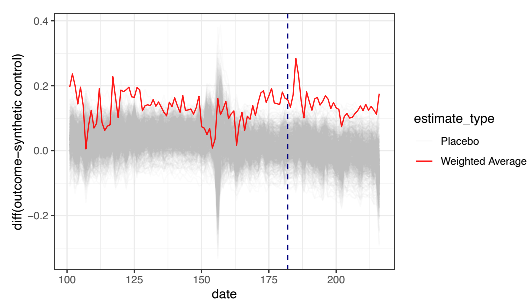

(a) log(Avg. driver fare)



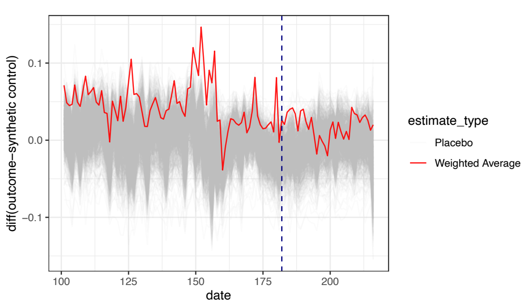

(b) log(N. trips)



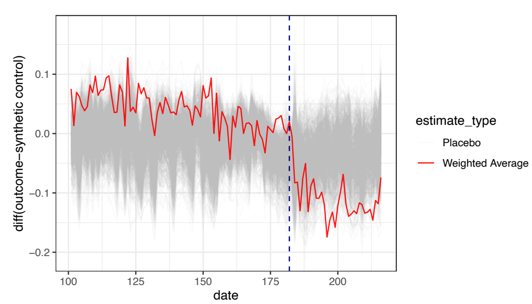

(c) log(N. drivers)




(d) log(Total supply hours)




(e) log(Avg. daily income)




(f) log(Avg. n rides/supply hr)

# C Appendix Tables

Table C.8: Average treatment effects on customer fare and total spending

| | log(Avg customer fare) | | log(Sum customer fare) | |
|---|---|---|---|---|
| | All services | Regulated | All services | Regulated |
| | (1) | (2) | (3) | (4) |
| Treat | 0.0974** | 0.2303*** | 0.0995* | 0.1327** |
| | (0.0450) | (0.0476) | (0.0544) | (0.0585) |
| Observations | 12,760 | 12,760 | 12,760 | 12,760 |
| $R^2$ | 0.83656 | 0.89504 | 0.98499 | 0.98572 |
| Within $R^2$ | 0.03658 | 0.19079 | 0.00926 | 0.01437 |
| Day fixed effects | ✓ | ✓ | ✓ | ✓ |
| City fixed effects | ✓ | ✓ | ✓ | ✓ |

*Notes:* All dependent variables are in log. Point estimates are of average treatment effects, $\hat{\beta}_1$, as in equation 1. Standard errors are reported in parentheses. Regressions are run on a panel of cities over day as unit of time. All regressions include city fixed effects and time fixed effects, and standard errors clustered at the city level. Two-tailed significance: p<0.1*; p<0.05**; p<0.01***.

Table C.9: Robustness check: Average treatment effects on the number of transactions, excluding the Eid al-Fitr period

| | log(N trips) | | |
|---|---|---|---|
| | Full sample | 1 week | 2 weeks |
| | (1) | (2) | (3) |
| Treat | 0.0021 | -0.0044 | -0.0079 |
| | (0.0829) | (0.0847) | (0.0867) |
| Observations | 12,760 | 12,354 | 11,948 |
| $R^2$ | 0.98193 | 0.98332 | 0.98361 |
| Within $R^2$ | $3.5 \times 10^{-6}$ | $1.66 \times 10^{-5}$ | $5.53 \times 10^{-5}$ |
| Day fixed effects | ✓ | ✓ | ✓ |
| City fixed effects | ✓ | ✓ | ✓ |

*Notes:* All dependent variables are in log. Point estimates are of average treatment effects, $\hat{\beta}_1$, as in equation 1. Labels above the column number refer to the sample restriction. "Full sample" uses all data from January 1 to August 8, 2019. Column with the "1 week" label excludes data from June 2 (one day before the start of Eid al-Ftr) to June 8, 2019. Column with the "2 weeks" label excludes data from June 2 to June 15, 2019. Standard errors are reported in parentheses. Regressions are run on a panel of cities over day as unit of time. All regressions include city fixed effects and time fixed effects, and standard errors clustered at the city level. Two-tailed significance: p<0.1*; p<0.05**; p<0.01***.

Table C.10: Robustness check: Average treatment effects on the number of drivers, excluding the Eid al-Fitr period

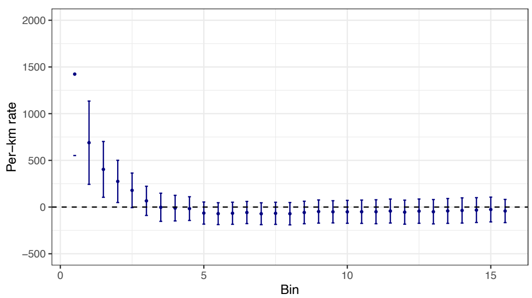
| | log(N drivers) | | |
|---|---|---|---|
| | Full sample | 1 week | 2 weeks |
| | (1) | (2) | (3) |
| Treat | 0.0649 | 0.0582 | 0.0562 |
| | (0.0605) | (0.0615) | (0.0628) |
| Observations | 12,760 | 12,354 | 11,948 |
| $R^2$ | 0.98460 | 0.98603 | 0.98627 |
| Within $R^2$ | 0.00494 | 0.00454 | 0.00441 |
| Day fixed effects | ✓ | ✓ | ✓ |
| City fixed effects | ✓ | ✓ | ✓ |

*Notes:* All dependent variables are in log. Point estimates are of average treatment effects, $\hat{\beta}_1$, as in equation 1. Labels above the column number refer to the sample restriction. "Full sample" uses all data from January 1 to August 8, 2019. Column with the "1 week" label excludes data from June 2 (one day before the start of Eid al-Ftr) to June 8, 2019. Column with the "2 weeks" label excludes data from June 2 to June 15, 2019. Standard errors are reported in parentheses. Regressions are run on a panel of cities over day as unit of time. All regressions include city fixed effects and time fixed effects, and standard errors clustered at the city level. Two-tailed significance: p<0.1*; p<0.05**; p<0.01***.

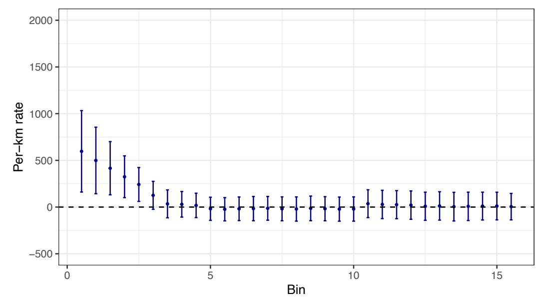
# D Appendix Figures

Figure D.6: Treatment effects on percentiles of the driver fare/trip distance by trip distance bin

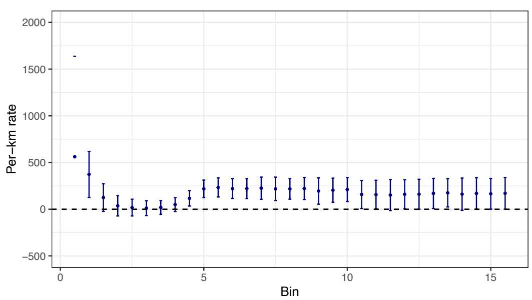
(a) 5th percentile



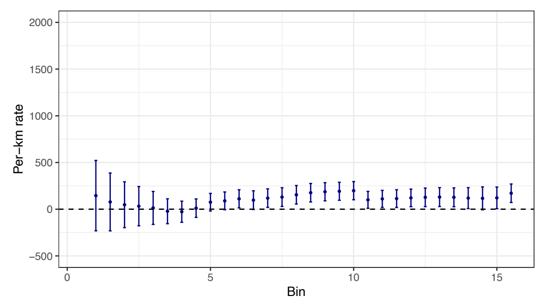

(b) 25th percentile




(c) 75th percentile



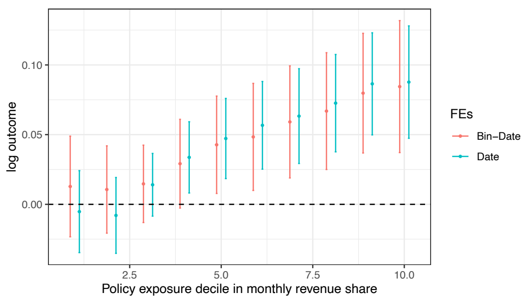

(d) 95th percentile



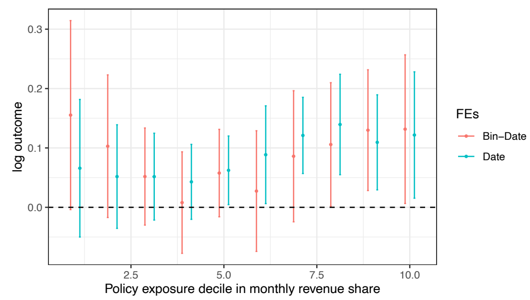

*Notes*: The dependent variables are the percentile measures of total driver fare divided by the distance driven in kilometers. Point estimates in blue are the estimated impact for each 500-meter trip distance bin. The 95% confidence intervals for coefficients are shown as the whisker bar in blue. The model includes bin-by-city fixed effects and fare-bins-by-day fixed effects. Standard errors are clustered at the city-bin level..

Figure D.7: Conditional average treatment effects by drivers' pre-policy exposure

(a) Driver fare/trip



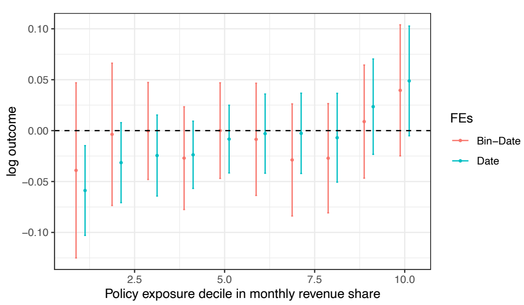

(b) Supply hours/decile




(c) N. distinct drivers/decile


(d) Supply hours/driver


(e) Driver earnings/driver

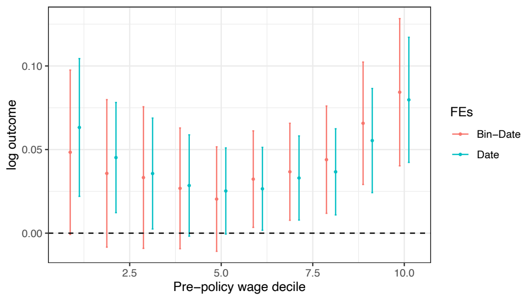



(f) Driver wage

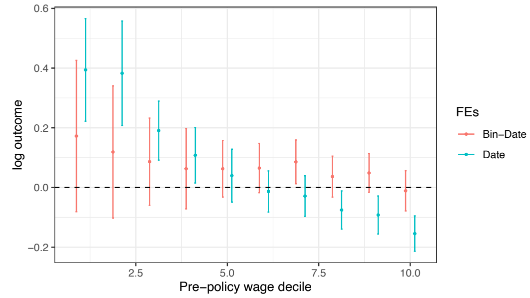

*Notes*: The dependent variables are the log-transformed outcomes listed in subfigure captions. The deciles are defined by the potential policy exposure of drivers during the pre-policy period, defined as the share of the driver's monthly earnings that would qualify for the policy. The pre-policy period is defined as June 1 to June 30, 2019. Data are restricted to transactions from drivers who were present in the data as of June 30, 2019. Point estimates in blue are the estimated impact on each bin of pre-policy average daily transaction. The 95% confidence intervals for coefficients are shown as the whisker bar in blue. The model includes bin-by-city fixed effects and either fare-bins-by-day fixed effects or day fixed effects. Standard errors are clustered at the city-bin level. The sample is restricted to drivers who joined the platform by June 30, 2019.

Figure D.8: Conditional average treatment effects by drivers' pre-policy wage

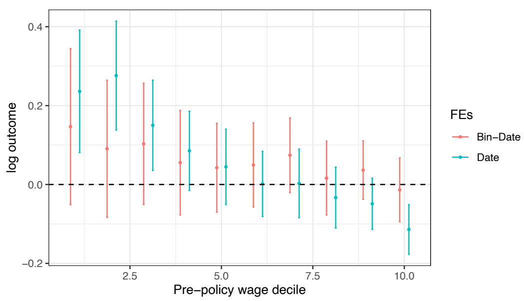
(a) Driver fare/trip



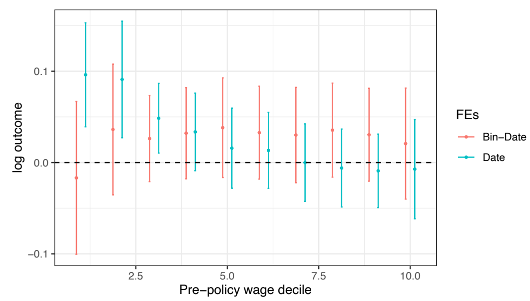

(b) Supply hours/decile



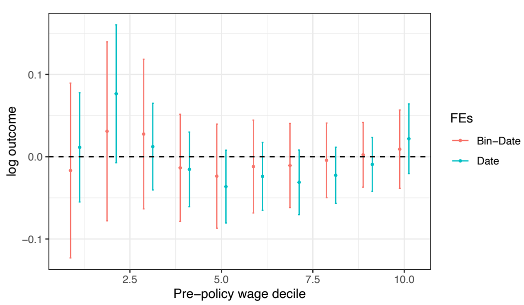

(c) N. distinct drivers/decile



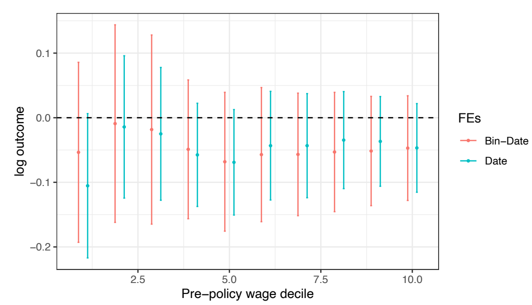

(d) Supply hours/driver




(e) Driver earnings/driver




(f) Driver wage




*Notes*: The dependent variables are the log-transformed outcomes listed in subfigure captions. The deciles are defined by the potential policy exposure of drivers during the pre-policy period, defined as the share of the driver's monthly earnings that would qualify for the policy. The pre-policy period is defined as June 1 to June 30, 2019. Data are restricted to transactions from drivers who were present in the data as of June 30, 2019. Point estimates in blue are the estimated impact on each bin of pre-policy average daily transaction. The 95% confidence intervals for coefficients are shown as the whisker bar in blue. The model includes bin-by-city fixed effects and either fare-bins-by-day fixed effects or day fixed effects. Standard errors are clustered at the city-bin level. The sample is restricted to drivers who joined the platform by June 30, 2019.

# Chapter 2: Spillovers under Information and Search Frictions—Experimental Evidence from an Online Platform in Pakistan*

Shotaro Nakamura Syed Ali Hasanain Adeel Tariq

August 11, 2023

**Abstract**

Information communications technology is shown to reduce search and information frictions in developing markets. Yet the mechanisms through which such interventions trigger strategic responses, spillovers, and adjustments to the market in developing economies remain under-explored. We causally identify spillover effects and their mechanisms via a randomized control trial on a major online listing platform for used vehicles in Pakistan, where there is limited publicly available price information. We

*Nakamura: University of California, Davis, CA; snnakamura@ucdavis.edu. Hasanain: Lahore University of Management Sciences, Lahore, Pakistan. Tariq: Lahore University of Management Sciences, Lahore, Pakistan. The views expressed in this paper are solely those of the authors, and do not reflect the views of PakWheels.com. The paper has gone through a check by PakWheels' employees to ensure confidentiality of their data and other proprietary information, but not on the empirical findings and views expressed in the paper. The authors report no conflict of interest. We thank Arman Rezaee, Monica Singhal, Diana Moreira, Ashish Shenoy for helpful comments. We also thank Ayesha Rao and Turyal Neeshat for excellent research assistance, and Mohammad Malick and Ahsan Tariq at the Institute of Development and Economics Alternatives (IDEAS) for excellent execution of the endline phone survey. We also thank Suneel Munj, Muhammad Raza Saeed, Raahim Rasheed, Hamza Madni, and Waris Ali at PakWheels.com for fascilitating the research collaboration. The endline survey for this study was funded by the Faculty Initiative Fund at LUMS. This study has been pre-registered in the AEA RCT Registry (ID: AEARCTR-0007537) and approved by IRBs in Pakistan (LUMS-IRB/06042021/AT-FWA-00019408) and the US (UC Davis IRB-1647279-1). All errors remain our own.

provide estimates of transaction prices privately to sellers who create new posts and capture their pricing, advertising, and transaction outcomes. We also identify direct and spillover effects via a saturation design at the vehicle-model level. We find that the information intervention brings sellers' listing prices closer to our price estimates, but increases transaction probability only for the spillover group. The findings point to two mechanisms: 1) effects of price information are mediated by advertising tools that could countervail effects of list-pricing choices, and 2) spillovers could propagate direct effects of information intervention via adjustments by competing sellers.

**JEL:** D82, L11, O12

**Keywords:** Price Dispersion, Search Friction, Online Platforms, Advertising, RCT

# 1 Introduction

Information and search frictions are often cited as causes of high price levels and dispersion and a significant share of trade costs in developing markets (Allen 2014; Startz 2016). Information communications technology (ICT), such as mobile phones and apps, has been shown to reduce by making price information more accessible (e.g., Aker 2010; Jensen 2007). Reducing search and information frictions via ICT could also lead to productivity gains and other knock-on effects through supply chains, suggesting the benefit of information interventions beyond the direct effects of reducing information friction (e.g., Jensen and Miller 2018, Hasanain et al. 2019). Large online marketplaces and platforms, which have increased their prominence in developing economies, may further reduce search costs by making it easier to acquire information about competing products (Fu et al. 2021). Yet, evidence of persistent price dispersion and search frictions in developed and developing economies' online markets suggests that platforms would not entirely eliminate information and search friction (Einav et al. 2015; Horton 2019; Fradkin 2015; Bai et al. 2020). Such persistent frictions point to the importance of understanding what types of search and information frictions agents in developing economies face and how they internalize such frictions with available tools.

The mechanisms through which ICT- or platform-based information interventions trigger spillovers, strategic responses, and adjustments in developing economies remain underexplored. First, price information interventions may generate spillover effects through improved market-wide access to information. Yet, we know little about the mechanisms through which such spillovers happen, due in part to the difficulty in scaling interventions and monitoring the mechanisms market-wide. Second, information and search frictions may also affect agents' choices on a wider range of decisions beyond pricing, causing further frictions. There is evidence that search friction and congestion in emerging online markets negatively affect the growth of high-quality traders (Bai et al. 2020). Yet, further empirical evidence is needed on how traders internalize and overcome such market frictions, such as pricing and advertising, particularly in a context where individuals' actions may cause spillovers on other market participants and have implications on market efficiency.

To provide insights into the links between information friction, individual choices, and spillovers in a developing market, we conducted a randomized controlled trial (RCT) in the used car market in Pakistan. In the intervention, we provide transaction-price estimates—called the Price Calculator—privately to sellers on a leading and nationally recognized online classified listing platform for used vehicles in Pakistan, PakWheels.com. With the experimental variation, we address the following research questions; first, we identify agents' pricing and advertising choices under search and information frictions and their internal mechanisms (i.e., changes in beliefs) behind those choices. Second, we identify if and how the information intervention induces spillover effects in the presence of information and search frictions. Specifically, we are able to capture a range of outcomes, such as prices, advertising, and transaction, as well as the direction of the impact relative to the direct treatment effects.

We overcome several traditional challenges in estimating the market-wide impact of an intervention. First, we relax the logistical constraint in conducting large-scale interventions in markets in developing economies through an extensive partnership with a popular and dominant online listing platform, an increasingly popular form of transaction in developing economies. Second, the collaboration with the platform allows us to conduct a natural field experiment, where the intervention is nested within the user interface of a popular platform, and the study sample consists of the vast majority of new listings (List 2007). Third, we measure changes in detailed, individual-level outcomes such as sellers' strategic choices and buyer-side responses using unique data captured by the platform. Forth, we generate variation in treatment saturation at the market sub-section level via a two-stage randomization saturation design. This partial identification strategy allows us to estimate the direct treatment effect, spillover, and saturation effects. Fifth, by privately providing the Price Calculator estimate, we rule out direct information spillovers and instead show knock-on effects of *choices* that treated sellers make.

We pre-specify and measure direct treatment and spillover effects on a) changes to the listing price, b) occurrence of the transaction, c) transaction price, d) usage of advertising tools, and e) index of buyer attention. We find that the intervention brings listing prices closer to our price suggestions for directly treated sellers. We find, however, that the intervention improves transaction outcomes for the spillover group by 1 percentage point, from the base of about 33%, but not for the directly treated group. We also find that the intervention increases the potential buyers' attention to spillover posts and reduces advertisement usage by the directly treated sellers. The findings point to two mechanisms: 1) effects of price information are mediated by advertising tools that could countervail effects of list-pricing choices, and 2) spillovers could propagate direct effects of information intervention via adjustments by competing sellers.

To further clarify the potential mechanisms that drive our main pre-specified results, we exploit our pre-specified model of static search with information friction. We find some alignment between the theoretical predictions and the empirical findings. Where there is misalignment, we speculate that the intervention induces shifts in sellers' beliefs that we need to account for in the theoretical framework. As such, we investigate shifts in sellers' beliefs as a potential mechanism using non-pre-specified outcomes from the endline survey of 2,311 sellers in our experimental sample.

From the survey, we find evidence that suggests that beliefs and their adjustments drive the set of results we find in the pre-specified outcomes. First, we find that price information intervention adjusts beliefs about the demand for the treated sellers. We also find that the intervention affects their beliefs about search frictions and market conditions, suggesting that sellers believe list pricing and advertising are substitutes. Importantly, these effects on beliefs are not detected for the spillover sellers, suggesting that they are responding to publicly visible choices of competitors but not adjusting their beliefs.

Our findings offer insights into the direct and spillover effects of information interventions on online markets' information environment in a developing economy, as well as mechanisms behind pricing and other choice parameters agents in the market have. We contribute to a strand of literature on search and information frictions in developing markets, motivated by a body of evidence that suggests the high transaction costs in trade (Allen 2014; Atkin and Donaldson 2015; Startz 2016; Aggarwal et al. 2022). We follow a body of work focused on ICT-based information intervention on price convergence and extend the knowledge into both detailed individual mechanisms and spillovers at the market level (Aker 2010; Aker and Mbiti 2010;Andrabi et al. 2017; Jensen 2007). Our work is also related to an emerging body of work on the effect of information interventions on spillovers up the supply chain and the roles of market structure in determining such effects (Jensen and Miller 2018, Hasanain et al. 2019; Mitra et al. 2018). Our work also generally addresses the externalities generated from information intervention but focuses on spillovers *within markets*, sellers' strategic choices, and the implications on market structure.

Our work is also motivated by a body of work documenting persistent price dispersion in online market platforms and sellers' and platform operators' incentives in those marketplaces (Dinerstein et al. 2018; Einav et al. 2015; Horton 2019; Fradkin 2015). The question is why search and information frictions persist in a world with plausibly low search and information costs. One view is that price dispersion and friction on online platforms are, in part, endogenous choices that platform operators make relative to other objectives, such as the extent of competitive pressure they want to induce. Dinerstein et al. (2018), for instance, shows using a redesign on eBay that balance between low information friction and competitive pressure is key to efficient online markets. This trade-off may be even more salient in developing economies with higher existing frictions and other market failures.

Lastly, our work contributes to an emerging body of work on the roles of online platforms in emerging economies, with a focus on reducing information and search frictions (Bai et al. 2020; Falcao Bergquist and McIntosh 2021 Couture et al. 2018; Fernando et al. 2020; Jeong 2020). On the extensive-margin access to platforms, Couture et al. (2018), find that while the benefits of access to e-commerce for rural markets in China are sizable, most of the gains accrue to the consumption side and to a minority of younger and richer users. The findings suggest that simply increasing access does not induce investments required to drive adaptation to e-commerce. On the intensive margin, Bai et al. (2020) suggest that search and information frictions still plays a major source of inefficiency on online platforms in developing countries, as they show that positive shocks to demand and information improve firms' performance in the long run, independent of productivity or quality. This suggests that market dynamics may generate inefficient firms and low-quality goods to persist in markets with information and search frictions based on the luck of having received positive initial demand shocks.

The remainder of this paper is organized as follows; Section 2 describes the context in which we conduct our intervention, and Section 3 the research design. Section 4 describes the pre-specified outcomes. Section 4.1 describes the identification strategy for our pre-specified analysis and our approach to multiple-hypothesis testing. Section 5 presents the pre-specified results, followed by an overview of the pre-specified theoretical framework to rationalize the results in Section 6. We then present results on non-pre-specified outcomes, primarily in the endline survey, in Section 7. Section 8 concludes.

## 2 Used car markets in Pakistan

Trading of used vehicles is a capital-intensive and frictional market in Pakistan, where vehicle ownership is low at around 6% (Pakistan Bureau of Statistics 2020). Anecdotally, trade has traditionally remained within existing social circles or through used car dealerships, with limited peer-to-peer transactions outside of their networks. There have been forums in which people exchange information on online bulletin boards and other social media. The most notable of such platforms, PakWheels.com, has evolved into a listing platform in which sellers and buyers can find each other and can find other information such as insurance and taxes. The platform receives approximately 100,000 new valid listings per month and has a similar level of active posts in a given time as its main competitor, olx.com.

Pricing high-value heterogeneous goods is challenging in a context without publicly available transaction price information. There is no publicly accessible and reliable information on transaction prices for used vehicles in Pakistan, where there is no equivalent to services like kbb.com. At the moment, the most comprehensive and publicly accessible price signals are the listing prices from online listing platforms like PakWheels. The lack of information may generate variation in market participants' beliefs about market prices, which, in turn, they emit into publicly accessible information in the form of listing prices.

There are three data points that corroborate this problem. First, our baseline data shows that only 33% of all listings are reported to have been sold, highlighting the underlying search and matching frictions. Second, we show in Figure 1 that even in the subset of listings that reported to have sold their vehicles, there are significant deviations and variations between the listing price and the transaction price. Third, the management at PakWheels has anecdotally indicated that their listing sellers are not pricing their vehicles "right," which led to the collaboration in which we provide price information to the sellers.

# 3 Experimental design

We conduct a field experiment in which we privately provide Price Calculator estimates to a randomly chosen subset of sellers. The Price Calculator estimates are based on a machine learning model using data on self-reported transaction prices from previous listings collected by PakWheels. The experiment is conducted within PakWheels' web and mobile platforms, where sellers create new posts. We assign treatment via a blocked, two-step randomization procedure with two saturation levels. The intervention is conducted over the course of 8 weeks to a flow of new posts. The sample selection and randomization procedure are described in the following sections 3.1 and 3.2. Figure 2 also shows the breakdown of posts into our sample and into treatment groups.

## 3.1 Sample selection

The platform receives up to 100,000 valid listings per month. Our experimental sample is new posts on the platform during the intervention period, except those for which PakWheels do not have sufficient data points to provide a Price Calculator estimate. The exact criteria for inclusion into the sample are masked for confidentiality reasons, but we include approximately 88% of all new posts into the study sample, consisting of approximately 70 distinct make-models. We arrive at the sample restrictions via the following steps.

First, we restrict our sample to the listings for which PakWheels would be able to provide Price Calculator estimates, i.e., vehicle types with large enough transaction volume with reported transaction prices. For instance, we do not include certain rare models for which PakWheels deemed not suitable to provide price estimates, such as luxury models or commercial vehicles, or large trucks. We cannot disclose further details on PakWheels' inclusion criteria into the Price Calculator estimation sample, but the resulting sample constitutes the vast majority of all listings.

Second, we impose restrictions based on when listings are created. For the primary analysis, we restrict the sample to listings created during the 8-week experimental period. For the secondary analysis of spillover effects, on the other hand, we also include listings created eight weeks prior to the start of the experimental period. This allows us to include model- (and model-version) fixed effects and run two-way fixed effect models, allowing for higher power of detecting treatment effects under an assumption on time trends. We discuss the benefit of these approaches in Section **??** and implications for power in Appendix Section E.1.

## 3.2 Two-step treatment assignment procedure

Our two-stage randomization process is as follows. In step 1, we block-randomize market clusters, defined as the make-model (e.g., Toyota Corolla), into two treatment (high vs. medium saturation) and control groups. In step 2, we randomize posts into treatment based on the last digit of the user ID on PakWheels. The assignment probability is 50 percent for the medium saturation group and 90 percent for the high group. In order to ensure that treatment and control groups are comparable in the primary outcome variables, we test for balance using listings data from a pre-treatment period with the same sample inclusion criteria and randomization procedure as the experiment. We bootstrap-sample and iterate this randomization procedure over 500 times and identify seeds for which we fail to reject differences in all primary outcome variables (described in Section 4), adjusted for false discover rate at 5%. We then randomly select one of those qualified seeds.

In step 1 of the two-step process, we run block-randomize make-model clusters into high-treatment, mid-treatment, and control groups. We use standardized cluster-level means of the five primary outcomes, as described in Section 4, and the cluster size.[1] Based on the blocks, we assign 50 percent of the clusters to control and 25 percent each to high- and low-treatment groups. Our choice of shares of clusters to treatment arms is informed by the literature on the optimal design of saturation design and our own Monte Carlo simulations using real data from the platform.[2]

In step 2, we assign treatment to posts based on the last digit of sellers' user-ID on PakWheels.[3] Treatment digits are chosen by a random number generator in $R$. The choice of digits for treatment is fixed across clusters and time in order to limit the extent of potential interference and for logistical simplicity. In other words, if a seller with user-ID $i$ is in a treatment group for model $m$, then all other posts by $i$ in $m$ are treated, as well as any other model $m'$ that is treated at the same saturation intensity as $m$. Treatment intensity of 50% or 90% stays constant for the cluster over the course of the experimental period.

### 3.2.1 Interference between clusters

One potential empirical challenge is interference between assignment clusters at the first stage of the randomization procedure. One may be concerned that if clusters are defined too narrowly, and pricing or advertising choices in one cluster could affect those in another, we would violate the Stable Unit Treatment Value Assumption (SUTVA). We allay this concern by using a relatively broad definition of clusters—the make-model—based on aggregated search logs data. We also address possible ways in which interference across clusters could still occur and their potential magnitude.

[1] Blocking is done with R's *blockTools* package (Moore 2012), which uses the optimal-greedy algorithm over the Mahalambois distance. We weigh the five main outcome variables twice as heavily as the cluster size variable. Our choice of weights is admittedly arbitrary, but the rationale is that the primary objective is to balance over main outcome variables and then with cluster size.

[2] We provide further detail on this process in Appendix Section E.

[3] The reason for this randomization procedure, as opposed to some others that does not rely on the user-ID, is partly for its simplicity in implementation, but also because we are assigning treatment to a *flow* of new listings (and some new users), meaning that we cannot pre-assign treatment to posts.

The aggregated search engine logs tell us which combinations of terms are used most frequently by viewers on PakWheels. For the randomization, our objective is to minimize concerns about inter-cluster interference but also retain as many randomization clusters for the step as possible. The aggregate search logs data are taken from the month of August 2020, containing approximately 68 million searches. The data contain numbers of searches per any combination of search terms (e.g., make, model, model-years in range, city, range of listing prices). We capture 35,000 most common search combinations, which account for 93% of all searches. We do not have information on the remaining 7% percent of less frequent combinations of specified search terms due to the capacity constraint of the partnering firm to address our data requests.

First, we observe that a majority (58%) of specified searches for posts on PakWheels included the make-model and the majority of those 58% also had additional terms (e.g., model year, city, price ranges). On the other hand, 32 percent of specified searches did not include make-models but instead included other fairly broad terms such as city name only (e.g., "Lahore"), vehicle make only (e.g., "Honda"), or all posts with pictures. We infer that these broad searches are mostly speculative and unlikely to lead to meaningful price comparisons between posts. We would have been worried about interference if, for instance, a significant portion of users searched for vehicles of specific characteristics (e.g., "mid-size sedans") that contain multiple make-models (e.g., Honda Civic and Toyota Corolla). Overall, the breakdown of specified searches indicates that the make-model is likely a reasonable and perhaps conservative level of clustering and that we are unlikely to have meaningful interference between clusters.

Second, interference across make-model clusters is likely minor because we provide private information that is specific to treated posts' characteristics, making it unlikely that there would be large direct information spillovers from one make-model to another. We confirm our intuition from our pilot telephone endline survey, in which almost none of the sellers reported having looked at listing prices of other models besides one of their own vehicles.

Yet, the following are some of the ways in which interference *across* make-models could occur, violating SUTVA across treatment clusters:

- Large enough shifts in the distribution of listing prices could eventually induce information spillovers. Such large shifts in list-price distribution could also lead to changes in transaction probability, transaction price, and congestion, which in turn may affect price distributions and market outcomes of similar models.

- Changes to the listing prices or advertising in treated clusters may shift buyers' attention to/from untreated make-models. Changes in buyer attention in untreated make-models may affect sellers' pricing and advertising choices.

## 3.3 Treatment assignment and take-up

The intervention is designed to minimize non-compliance; those randomly assigned treatments are automatically shown the Price Calculator estimate on the interface while they create a post. One exception is if the seller uses an older version of PakWheels' mobile app that does not yet contain the intervention tools. This may generate selection into treatment based on a) users' preference for PakWheels' mobile app as opposed to the web platform, which does not suffer from this issue, and b) their propensity to update the app. In order to mitigate this issue, PakWheels launched a new version of the app with the disabled intervention tool weeks in advance of the experimental period. The timing gave users plenty of time to update the new app before the intervention tool was enabled. Yet, it is possible, but unfortunately unverifiable, that a small fraction of users assigned to treatment did not receive it. As such, we identify both intend-to-treat and treatment-on-treated effects, as highlighted in Sections 4.2 and 4.3.

## 3.4 Intervention instrument: The Price Calculator

We provide estimates of the transaction price for used vehicles on PakWheels while sellers are creating their posts. The price information, which PakWheels calls "the Price Calculator", is based on a machine learning model trained to predict self-reported transaction prices using the firm's database of historical listings. The model estimate is conditional on the self-reported occurrence of the transaction, and we use observable attributes of the vehicle, but not of sellers' characteristics, as explanatory variables. Our hypothesis is that this information would help sellers identify realistic transaction prices and set listing prices accordingly.

To identify an error-minimizing forecast model, we take a gradient-boosting approach primarily for two reasons. First, gradient boosting—a method of ensemble predictions based on tree-based models—allows us to construct a predictive model that does not require estimating each of the make-model-modelyear fixed effects. We are, therefore, able to predict transaction prices for vehicles that had a relatively small number of observations within their own make-model-modelyear, but for which we had sufficient information to provide predictions. Second is that the gradient boosting approach performed best in most measures of error against other approaches in our initial design process, in line with the success of gradient boosting models in recent prediction competitions.

### 3.4.1 Display of the Price Calculator estimate

On PakWheels' web platform and mobile apps, sellers can create a new post by clicking on "Post an Ad." Sellers are first asked to log in so that PakWheels' platform can identify the user ID associated with each post. Users would not know their own user ID, as it is internal to PakWheels, or for which last digits we provide the Price Calculator estimates. Once logged in, sellers are asked to provide information about the vehicle they intend to sell, as shown in Figure 3. They then set the listing price in a box shown in Figure 4. If the seller is assigned to treatment, they are then shown a Price Calculator estimate, i.e., the machine-learning-based transaction price forecast, as well as the 10th and 90th percentiles of reported transaction prices for the make-model-model year (or make-model-modelyear-version for frequently traded models). These percentile measures would be labeled as "Lower end" and "Upper end" of transaction prices. Figure 5 shows how the Price Calculator estimate is displayed along with a brief description. Treated sellers are then given a chance to update their listing price, but not the untreated sellers. All sellers are then guided through the rest of the posting process.

## 3.5 User-experience after selecting the listing price

After providing information on the vehicle and selecting a listing price, sellers put their posts "live" on the platform and can be contacted by potential buyers. Sellers can adjust their listing price at any time as they gather more information about market conditions and be contacted by interested buyers. While list pricing is one of the primary choices that sellers make during initial posting and over the duration of the post's life, advertising is another way in which sellers can try to affect the outcome on PakWheels.com. The following are the three principal advertising strategies on which we create an indexed outcome variable for our analysis.

First, sellers can purchase "bump" credits and use them on their posts. The credits allow sellers to bring their post to the top of the result page in the default, reverse-chronological listing order. This effectively increases the post's visibility as more people look at the first pages of listings. Second, a "feature" credit would put their post in a few reserved spots at the top of the result page and label the listing as a "featured ad," much like promoted ads on Google searches. Posts are otherwise listed in the reversed chronological order within the class of featured ads. Third, sellers can provide signals of vehicle quality by requesting in-person inspections by PakWheels' mechanics. Based on a pre-specified rubric, the inspection would result in scores (out of 100) on eight dimensions: engine, brakes, suspension, interior, AC, electrical, exterior, and tires. The vehicles will pass the inspection if the unweighted average of scores over these eight dimensions is above a threshold. They can then be marked as "PakWheels certified" on the platform for an additional fee.

Sellers and potential buyers connect via the contact information listed on the post, negotiate, and transact outside the platform. As PakWheels.com is only a listing platform, it does not directly observe if a transaction occurs and, if so, to whom and at what price. Instead, the platform contacts the users regularly to request that the sellers self-report the transaction outcomes when they make a sale or want to take down the post. The posts expire after 90 days from the initial posting when sellers are again asked to report the transaction outcome.

## 3.6 Data sources

We leverage access to PakWheels' database, which contains all historical and live posts, to estimate the market-wide impact of our natural field experiment. At the post level, the database contains information on the posted vehicles' attributes, such as the make, model, model year, version, engine capacity, transmission, fuel type (e.g., petrol or CNG), color, and if it was assembled domestically. These fields are required for the sellers to post their vehicle and thus are consistently available for most listings. The database also contains additional information about the vehicles' attributes, such as stereo, air conditioning, and other amenities, that sellers have the option of reporting. For both the construction of our

Price Calculator model and for analysis, we use required fields as inputs.

The database also has information on list pricing and advertising choices, and page views over the course of time per post. For list pricing and advertising choices, the database tracks when prices are adjusted, and advertising tools are activated with timestamps. The page-views variable, on the other hand, is a live count and is updated every 12 hours in the database. Because of the time-varying nature of these variables, we measure them after all posts are expected to go offline, i.e., we collect data 90 days after the last day of the experiment. Similarly, seller-reported transaction outcomes, which we expect to be logged when the posts go offline, at the same time as we collect the time-varying data.

There are two primary potential challenges with relying entirely on PakWheel's database. First, transaction outcomes and prices are self-reported for a subset of sellers and may suffer from selection bias or cannot be verified. Second, PakWheels' database does not contain information about the sellers' beliefs about market conditions, expectations on transaction outcomes, and other measures of experience in the marketplace on a representative sample. As such, we also collect data from the telephone endline survey of 3,000 representative sellers on their transaction outcomes, unincentivized beliefs, and other measures of experience.

Other data items we use for design and secondary analysis include i) aggregated search engine results in terms of keywords and their combinations, ii) daily search listing orders from PakWheels, and iii) a usage log of a previous iteration of the Price Calculator, which preceded the experiment. Further details about each of the data sources can be found in Appendix Section A.

# 4 Prespecified outcomes

The primary objective of this paper is to address how a price information intervention induces direct and spillover effects on sellers' pricing and advertising choices, transaction outcomes, and mechanisms. As such, we pre-specify five primary outcome variables, for which we address the issues of multiple hypothesis testing. The five outcomes are a) changes to the listing price, b) occurrence of the transaction, c) transaction price, d) usage of advertising tools, and e) index of buyer attention, which are defined as follows.

### 4.0.1 log-absolute difference in prices

We consider changes to listing prices as the "first-stage" effect of our intervention, in that impact on other primary outcomes hinges on the changes to listing prices and their distributions. We expect that sellers would adjust their listing price toward the Price Calculator estimate, plus some margin for expected bargaining. In order to capture this type of convergence, we define our primary price outcome to be the natural-log transformation of the absolute difference between the final listing price and the Price Calculator estimate that the seller received or would have received. PakWheels calculates and provides the Price Calculator estimate only to treated posts, so we estimate the prices that control posts *would have received* using the identical model as the one PakWheels uses for this experiment.

As discussed in Section 3.5, sellers can update prices and other features as long as their posts are active on the platform. Direct effects of the Price Calculator estimate may happen when the post is created, while indirect effects may occur even after the post is created through feedback from buyers and competition with other posts. We use the listing price at the end of posts' active status for our primary outcome so that all changes to the listing prices are factored in.

### 4.0.2 Transaction outcome and price

Sellers on PakWheels can take down their posts once they no longer wish to receive inquiries or the post expires after 90 days since the initial posting. When the post is taken down, sellers are asked if they have sold their vehicles. They are required to respond in order to have their ads taken off. They are given options on the form (e.g., sold via PakWheels' website, sold via others, chose not to sell, etc.), and most sellers choose one of them. However, some respond as "Other" yet report in the comment section that they have sold the vehicle. Our transaction outcome variable accounts for this to the best extent possible by string cleaning responses classified as "Other." The transaction variable is 1 if the seller reported a sale and 0 otherwise.

Sellers are also prompted to report the transaction price on the online form if they report having sold their vehicle. The value is missing for those who do not report their transaction outcome. We also remove inputs outside of the reasonable price range for their given make-model. We use the natural log of the transaction price as the outcome variable.

These self-reported outcome data are likely the best source of information on transactions and prices across a wide range of vehicle characteristics and locations in Pakistan. However, they may be vulnerable to biases and are checked against values collected via a telephone survey described in Section A.4. We plan on using responses from this survey to construct analogous outcome variables for robustness checks.

### 4.0.3 Advertisement index

One of our main hypotheses is that, when faced with novel price information, sellers adjust their strategic choices along two margins; list pricing and advertising. We capture sellers' choices on advertising with data on paid services on PakWheels. As discussed in Section

3.5, sellers can increase the visibility of their posts and/or signal quality by "bumping", "featuring," and requesting an inspection for their vehicle. In order to capture both intensive and extensive usage of advertising tools, we construct an index measure consisting of the following variables:

- number of "bumps" the seller applies to the post
- number of weeks the seller "features" the post
- 1 if the seller requests PakWheels to have the vehicle inspected.[4]

### 4.0.4 Buyer-attention index

We also hypothesize that the price information intervention, and causal effects on pricing and advertising, affect treated posts' visibility on the platform. In order to capture this effect on the post's visibility and buyer attention, we construct an indexed measure from data discussed in Section A.3. The index consists of the following variables:

- page views (i.e., clicks on the post)
- clicks on the "Show Phone Number" button within the post to contact the seller.

[4]Given that certification is endogenous to vehicle quality, we use the data on whether or not the vehicle was ever inspected, as opposed to certified.

## 4.1 Empirical strategy

## 4.2 Intend-to-treat effects

We estimate the intent-to-treatment effect of being provided the Price Calculator estimate using Equation 1, where the coefficients of interest are $\beta_1$, $\beta_2$, and $\beta_3$:

$$\begin{aligned} Y_{i,p,m,w} = \beta_0 + \beta_1 * Assign_{i,m} + \beta_2 * Cluster_m + \beta_3 * ClusterHigh_m \\ + \bar{Y}_{m,w\in[-15,-8]} + \psi_w + X'_{i,p}\rho + \epsilon_{i,p} \end{aligned} \tag{1}$$

The subscripts used in the equation above indicate the following:

- $i$: individual user identifier (defined by PakWheel's user ID)
- $p$: post (multiple posts could belong to a given $i$)
- $m$: vehicle make-model cluster
- $w$: posting week. $w = 1$ is the first week of the experimental phase.

This estimating equation is fitted to data of listings that were created during the 8-week experimental period and for which Price Calculator estimates could be generated, as discussed in Section 3.1.

$\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$ capture the ITT effects. $Assign$ is the binary direct treatment variable, $Cluster$ is a dummy variable that equals 1 if the model is selected for first-stage assignment (of either saturation level) and zero otherwise. $ClusterHigh$ is a dummy variable for high-saturation cluster-level treatment. Since we cannot have model fixed effects, we include the pre-experimental, model-level means of the outcome variable from weeks -15 to -8. We select this time period as it would be sufficiently far from the experimental time frame, and the vast majority of posts created in weeks -15 to -8 would already be taken down week 1. $\psi_w$ denotes the week fixed effects, and $X_{i,p}^{\prime}$ is a vector of controls for vehicle and seller characteristics, as follows:

- Vehicle characteristics:
  - vehicle's age (by model year)
  - log(mileage)
  - engine capacity
  - transmission
  - fuel type (e.g., petrol, CNG)
  - color
  - assembly (domestic or imported)
- Seller's characteristics:
  - Seller's city
  - 1 if professional dealer, as observed through PakWheels' account information
  - log(number of listings ever made on PakWheels)
  - log(months since first listing on PakWheels)

For all dependent variables other than the binary transaction outcome, we use linear regressions. For the binary outcome variable, we use the logit model. We cluster the error at the make-model level, as the first stage of the randomization is conducted at this level. We also estimate these models using heteroskedasticity-robust standard errors as a robustness check.

## 4.3 Treatment-on-the-treated

As discussed in Section 3.3, we may encounter some treatment non-compliance by sellers with old versions of the PakWheels app that does not include the intervention tools. This type of non-compliance is rare but likely non-random, so we instrument for the treatment take-up using the assignment variable.

The treatment-on-the-treated (TOT) effect is estimated via 2SLS, with $Assign$ instrumenting for $Treat$, and $Cluster$ and $ClusterHigh$ included as controls.

$$\begin{aligned} Y_{i,p,m,w} = \theta_0 + \theta_1 * \widehat{Treat}_{i,p} + \theta_2 * Cluster_m + \theta_3 * ClusterHigh_m \\ + \bar{Y}_{m,w\in[-15,-8]} + \psi_w + X_{i,p}^{'}\rho + \epsilon_{i,p} \end{aligned} \tag{2}$$

The first-stage specification for $\widehat{Treat}$ is as follows:

$$Treat_{i,p} = \phi_0 + \phi_1 * Assign_{i,m} + \bar{Y}_{m,w\in[-15,-8]} + \psi_w + X_{i,p}^{'}\tau + \xi_{i,p} \tag{3}$$

$\hat{\theta}_1$ represents the estimated TOT effects. The specifications include controls $\psi_w$, $\gamma_m$ , and $X_{i,p}^{'}$ in the first and second stages, as we did for the ITT effect. $\xi_{i,p}$ is error term in the first stage.

## 4.4 p-value adjustments

In order to address the issue of multiple hypothesis testing, we follow Romano and Wolf 2005 and correct for the false discovery related to tests on the five primary outcomes: log(absolute difference between listing price and Price Calculator estimate), binary transaction outcome, log(transaction price), indexed measure of advertisement usage, and buyer-attention index.

Given that we consider direct treatment and spillovers as separate hypotheses, we adjust their critical values separately. We, therefore, report adjusted critical values at five percent, based on the Romano-Wolf procedure. There are three groups of five null hypotheses related to the primary outcomes; that the coefficients on three main exogenous variables ($Assign$, $Cluster$, and $ClusterHigh$ for ITT) in the regressions of five pre-specified primary outcomes are not statistically different from zero. We report both the unadjusted p-values and adjusted q-values for these primary hypotheses and only unadjusted p-values for tests on secondary outcomes.

# 5 Results from pre-specified analysis

## 5.1 Balancing checks

We begin by examining the balance of outcomes in the pre-treatment period. Our test differs from a standard approach in which one measures outcome variables and covariates at baseline and tests for statistically significant differences between treatment and control groups. In our context, we do not observe baseline measures of the experimental sample because we treat a subset of the flow of new listings, and outcomes are observed only after treatment, i.e., the presentation of the Price Calculator estimates.

As such, we conduct a "placebo" test of balance by using listings data between 8th November 2021 and 9th January 2022, the 8-week pre-treatment period from which the data for randomization comes. The null hypothesis is that the coefficients on the placebo treatment variables assigned to the pre-treatment listings would not be statistically different from zero, based on the intent-to-treat estimators discussed in Section 4.2. We conduct the tests on five pre-specified primary outcome variables and adjust for multiple hypothesis testing as described in Section 4.4.

Table 1 shows the means and standard deviations by placebo treatment groups of the pre-treatment period listings, and Table 2 presents the tests of balance via placebo regressions. We present coefficients "Assignment," "Spillover," and "Spillover (high)" corresponding to the direct treatment effect ("Assign"), as high- ("Spillover (high)") and low- ("Spillover") saturation spillover effects based on Equation 1, respectively. We report unadjusted statistical significance with stars next to the coefficients and adjusted q-values at the bottom three rows of the table.

The balance test shows that we fail to reject the null on almost all outcome variables at conventional levels after controlling for the false-discovery rate. First, unadjusted p-values are below 0.05 for one of the fifteen tests and between 0.10 and 0.05 for two tests. When we adjust for false discovery rates via the Romano-Wolf procedure, however, we find that only one q-value ("Spillover" for Column 4, "Page-view index") is between 0.10 and 0.05, and none below 0.05. Although not a direct test for balance on the experimental sample, this placebo test during the period preceding the intervention is the closest conceptual approximation to the standard baseline test for balance.

## 5.2 Treatment effects on primary outcomes

Through the analysis of pre-specified primary outcomes, we find that our price information intervention caused changes to both sellers' choices and their transaction outcomes. Table 3 shows the ITT estimates, as specified in Equation 1 on pre-specified outcomes. Table 4 shows the TOT effects that are qualitatively indistinguishable from the ITT counterparts, so we focus our interpretation based on the ITT estimates. Table 5 also shows ITT results by treatment groups (i.e., assigned and spillover groups for high and low saturation models), following an alternative specification to the pre-registered empirical model. First, we find evidence that the intervention brings the listing price closer to the Price Calculator estimates as a direct treatment effect and suggestively as spillovers. We find, however, that it increases the transaction probability for the spillover group but not for the directly treated group. We find effects on two potential mechanisms: the use of advertising tools and the resulting shift in buyer attention. We find that the direct treatment reduces advertisement usage for the directly treated group. This effect leads to fewer page views for the directly treated and more for the spillover group.

Column 1 in Table 3 indicates that the intervention reduces the absolute difference between the listing price and the Price Calculator estimate by 3.3% as a result of direct assignment and a further 7.8% as a result of spillovers. The direct assignment effect survives the critical value adjustment, but the spillover effect does not. There is also no statistically significant *additional* effect of a higher saturation, though the coefficient estimates of the higher saturation are less precisely estimated across the rest of the outcomes. The set of results from Column 1 shows the direct effect of price signals on reducing price variations away from the Price Calculator estimate and some suggestive evidence of a spillover. Appendix Table F.1 also shows results on non-primary price outcomes. We do not find evidence of adjustments in listing price **levels**, but evidence of increased listing price adjustments for the high saturation group.

Column 2 in Table 3 shows that the direct treatment effect on the reported transaction is negative and statistically significant, yet positive and statistically significant in the equal magnitude for the spillover. Because these effects are additive, the direct effect and the spillover coefficients cancel out for the direct treatment group. On the other hand, the net effect of treatment is positive for the spillover group. The magnitude of the direct and spillover effects is 1 percentage point, as shown with a linear probability model in Appendix Table F.2. Column 3 Table 3 also shows that the intervention does not significantly affect the log transaction price either directly or through spillovers, although there is an endogenous selection of the sample by those who report their transactions and prices. Overall, we find a set of counter-intuitive results that the Price Calculator intervention improves the transaction probability of spillover groups but not the directly treated ones, the mechanisms behind which we explore in the remainder of our analysis.

Column 4 in Table 3 shows that the direct treatment effect on the buyer attention index is negative, but the spillover effect is positive. The effects are relatively small, at -0.02 SD for the direct effect and 0.03 for the spillover effect. Appendix Table F.3 breaks down the effects on the index by its components of the effect. The table shows that direct assignment reduces the number of page views by 60, and the spillover effect increases it by 57. Similarly, the direct assignment reduces phone number views by 0.41, and the spillover effect increases it by 1.15. Table 5 also indicates that, when assessed by treatment groups, the effects on increased page views are found for spillover groups. We, therefore, find robust evidence indicating that buyer attention increases as a result of spillovers, but the effects are somehow muted for direct treatment groups. We hypothesize that the effects on the page-view index are a result of differential choices by directly treated and spillover groups.

Lastly, Column 5 in Table 3 shows that the direct treatment effect on advertising usage is negative, at 0.01 SD, but the spillover effect is statistically and economically insignificant. Appendix Table F.4 shows the results on each component of the advertising index, as well as an unwinsorized index outcome for reference. We find that direct assignment reduces advertising usage in all components: bumps, features, and certifications. We also find positive and significant effects for high-saturation spillover effects on bumps, certifications, and the un-winsorized index but do not find it in the pre-specified winsorized outcome.

## 5.3 What mechanisms explain the set of pre-specified results?

We find in our analysis of pre-specified outcomes that the intervention reduced the list price's deviation from the Price Calculator estimate and induced positive transaction outcomes through increased page views for the spillover group but not for the directly treated. We also find reduced use of advertising by the direct treatment group. Yet, questions remain as to how to make sense of the set of results in combination. What are the overarching mechanisms that relate pricing, advertising, buyer attention, and transaction outcomes in the context of search and information frictions? And how do we account for spillovers into the mechanisms, particularly when the signs of the effects differ from those on the direct treatment counterpart?

To address these challenges, we use a pre-specified conceptual framework of static search and a set of comparative statics, with which we derive a set of predictions. The conceptual framework is specified in the pre-analysis plan and summarized in Section 6. We then evaluate how closely the pre-specified empirical results align with theoretical predictions. We then reconcile any deviations of empirical results from the model predictions or confirm alignment by providing further evidence on survey-based measures. In this analysis in Section 7, we focus on sellers' beliefs, which we highlight in the model as underlying mechanisms.

# 6 Conceptual framework

We present a simple search framework that addresses various mechanisms of search and information frictions incurred by agents in a developing market. The objective of this exercise is to identify mechanisms through which lack of access to information could generate losses in unrealized transactions or may induce externalities in terms of search and information frictions. We combine our theoretical predictions with empirical results to demonstrate how sellers facing such frictions set listing prices, promote their posts through advertising, and respond to information about market conditions.

There are several channels through which search and information frictions may affect prices and transactions in highly frictional markets. First, the sheer lack of access to, or high cost of accessing, price information could result in variations in otherwise optimal choices. Second, even with access to price information and signals, individuals' beliefs about market conditions and the signal quality may vary, leading to variations in their otherwise rational chocies. Third, sellers may generate spillovers through, for instance, spillovers of information itself, changes in their pricing and advertising decisions, and their choices affecting potential buyers responses that then back to affect sellers' choices.

We address a wide range of possible channels listed above with the conceptual framework while staying with a simple and tractable model to generate clear predictions. We use a static search model that derives inspiration from canonical frameworks such as Stigler (1961) and Diamond (1982). Most contemporary models that focus on the effect of access to price information assume full knowledge of parameters on market friction and demand distributions (Baye et al. 2007). We introduce the following deviations from a standard search framework:

- We allow for supply-side heterogeneity of access to information and resulting beliefs about the demand-side distribution. In effect, sellers have biased or noisy beliefs about the distribution of buyers' willingness-to-pay (WTP).
- This, along with possibly noisy beliefs about the match rate and efficacy of advertising, would lead to biased or noisy beliefs about the probability of sale and to suboptimal list pricing.
- We allow the match rate with potential buyers to be endogenous with respect to ad-

vertising choices sellers make. They can influence the match rate by engaging in costly actions, i.e., advertising.

Our approach is similar to that of Bai et al. (2020), who model and empirically estimate the search and information frictions buyers experience and resulting firm and market dynamics. Unlike Bai et al. (2020), who focus on the demand side, we address the role of information friction on the supply-side and search friction that sellers experience. Our focus on mechanisms is founded on previous work such as Bergquist and McIntosh (2021) and Bai et al. (2020), who show that the existence of, or mere access to, online platforms does not resolve issues of search and information frictions and that frictions that persist on such platforms deserve attention.

We set up a model in which a seller $i$ is endowed with an asset and certain unobservable characteristics $s_i$, as well as information set $I_i$. The search process is composed of the following steps:

1. Seller $i$ forms beliefs about the distribution of buyers' WTP based on information $I_i$.
2. Seller $i$ chooses a listing price $p_i^l$ and amount of advertisements $a$ to optimize expected returns from participating in the marketplace.
3. Seller $i$ matches with a potential buyer via a Poisson process.
4. Once matched, seller $i$ makes a take-it-or-leave-it (TIOLI) offer $p_i^t$ below $p_i^l$ to the potential buyer.
5. Transaction occurs if the matched buyer's WTP is higher than $p_i^t$.

We provide further detail on the set-up and derive the model in Appendix C. Section C.1 lays out the setup of our model and provides definitions of terms and parameters. Section C.2

defines the objective function and the maximization problem in terms of the listing price and advertising choices. Section C.3 gives optimality conditions in the case of no information friction. Section C.4 shows how individual choices may be altered when there is noise in beliefs about the demand and how price information signals would induce updates in beliefs and alter input decisions. Section C.5 concludes by providing predictions on the role of information friction and noisy beliefs on demand in terms of sellers' choice variables and transaction outcomes.

## 6.1 Theoretical predictions

We derive the following predictions from the theoretical framework:

1. The price information intervention brings the listing price $p_i^l$ closer to what it would be under no noise in beliefs about demand.
2. The information intervention increases expected returns from the search process.
3. The information intervention increases the consumption of advertising $a$ if sellers' beliefs about expected returns from search are adjusted upward.
4. Spillover effects could occur through lower noise in publicly available price signals, which could increase returns from the platform and from advertising.
5. Spillover effects could occur if the intervention affects transaction outcomes and, consequently, the Poisson match rate in a treated market segment.

## 6.2 Do the theoretical predictions align with empirical results?

The first prediction is consistent with the empirical results that show that the listing price move close to the Price Calculator estimates. One assumption that underlies this conclusion is that the optimal listing price $p_i^{l*}$ under no noise in beliefs about demand and the optimal transaction price $p_i^{t*}$, which the Price Calculator estimates, is close enough that a seller moving their listing price closer to $p_i^{l*}$ would also reduce the distance to $p_i^{t*}$.

The second prediction that the information intervention increases expected returns is not explicitly supported by the empirical results of the pre-specified outcomes. If the effect of direct treatment is increased expected returns, then one might expect positive effects on transaction probability, transaction price, or the revenue from the search process. One possibility is that the intervention affects other mechanisms, such as advertising and buyer attention, that may, in turn, negatively affect the transaction outcomes. The other possibility is that directly treated sellers increased their expectations of transaction price and held more optimistic view on the market conditions, but such effects are not detected by reported transaction outcomes. To explore these possibilities, we analyze survey-based outcomes in Section 7.

The third prediction is that the information intervention would increase advertising if the seller's belief about expected returns from search is adjusted upward. Empirically, we find that the intervention *reduces* advertising. This suggests that either sellers' beliefs about expected returns are negatively altered or their beliefs about other factors, such as the extent of information and search frictions they face, are affected. The latter is a possibility if directly treated sellers believe that they face lower information friction. We confirm the effects on beliefs about frictions and market conditions in Section 7.

The fourth prediction is that spillovers could occur through lower noise in publicly available price signals, which could increase returns from the platform and from advertising. Empirically, we find positive spillovers in terms of the pricing choices but not on the use of advertising. Fifth, the model predicts that the spillover effects could occur if the intervention affects transaction outcomes and, consequently, the Poisson match rate in a treated market segment. Our effects on increased page views for the spillover group seem to confirm this view. In Section 7, we also test if the beliefs of sellers in the spillover group about search and information are affected to narrow in on specific mechanisms behind the spillover effects.

Overall, we find that some of the theoretical predictions line up with the empirical findings, but not perfectly. We speculate that the changes in sellers' beliefs and how the intervention induces it may help explain the deviations between theoretical predictions and empirical findings. As such, we analyze the results of survey outcome measures on sellers' beliefs in the next section.

# 7 Results from non-pre-specified analysis

## 7.1 The endline survey measures on beliefs

We provide insights into mechanisms put forth by the conceptual framework and the empirical mechanisms through data from a telephone endline survey. We conducted the survey on a representative subsample of 3,000 sellers, balanced across make-model clusters and within-cluster treatment assignments. The survey is conducted on sellers 4 to 6 weeks after their posting in order for the timing to be early enough to minimize recall bias and keep a high response rate, but late enough that sellers have engaged with potential buyers and transaction outcomes mostly determined. We received responses from 2,311 of them (77% response rate) of the sampled respondents. The original intention of the survey was to confirm the self-reported outcomes, but additional questions are included to capture potential mechanisms of the treatment and spillover effects.

We test for both direct treatment and spillover effects on these belief measures, using the specifications listed in Section 4.1. We ask a series of questions pertaining to sellers' beliefs about the market conditions, perceptions of market frictions, and perceptions about the Price Calculator instrument. These questions are meant to capture changes in sellers' beliefs about the demand distribution, i.e., the possibility that tailored price information leads sellers to have less noisy beliefs about the eventual transaction price. The belief outcome measures that we test can be categorized as follows:

- sellers' expectations about transaction prices and their willingness to negotiate
- sellers' beliefs about search and information friction
- sellers' demand for the Price Calculator tools and beliefs about the effectiveness of advertising tools.

## 7.2 Treatment effects on survey outcomes

First, we find that the intervention moves sellers' price expectations toward the Price Calculator estimates and makes them more willing to bargain. Table 6 shows that treatment assignment moves sellers' expected transaction prices closer to the Price Calculator estimates by 21.3%, as shown in column 1. The treatment also increases sellers' willingness to bargain by PKR 5,600. Importantly, statistically significant effects are only detected for direct treatment effects and for spillovers. The set of results confirms the idea that sellers' beliefs about prices are adjusted based on exogenously shifted price signals, i.e., the estimates they receive from the Price Calculator. The result of willingness to bargain may also indicate that sellers are more flexible on pricing to market conditions. The set of results conforms with Theoretical Prediction 1.

Second, we find that directly treated sellers' beliefs about search and information frictions are more optimistic. Table 7 shows that directly treated sellers' perceptions about the difficulty of getting inquiries and good prices improve by 0.05 to 0.07 on the Likert scale.

Again, we do not see similar impact as spillover effects. The results on sellers' beliefs about search and information frictions may help explain the reduction in advertising by directly treated sellers, especially given adjustments in their listing price. If sellers adjust their listing prices according to new information, and they believe they reduce friction from search and information asymmetry, then they may opt to spend less on advertising. This may help explain discrepancies between the pre-specified empirical results and Theoretical Predictions 2. and 3.. We also note, however, that we do not find higher demand for the Price Calculator tools or for advertising as a result of direct treatment, as shown in Table 8.

Third, we note that almost all effects on beliefs are direct treatment effects, not spillovers. All tables mentioned above from survey outcomes do not show statistically significant effects on the spillover coefficients, with the exception of Column 3 in Table 7.[5] Table 9 also shows that sellers believe they are exposed to the Price Calculator as a result of direct treatment and not through spillovers. The set of results seems to indicate that the spillovers to sellers who are not directly treated occur through changes to the market conditions caused by choices of the directly treated.

The findings from the endline survey measures suggest that changes in sellers' beliefs about demand and search frictions may play a role in the effects of information interventions. However, changes in beliefs seem to only occur when the agents are directly confronted with new information; spillover sellers' beliefs are not altered but are merely responding to publicly visible choices of competitors. Combined with the effects we see on advertising in Table 3, the evidence seems to point out that pricing and advertising are substitutes, as directly treated sellers adjust their listing price but also lower their advertising usage.

[5] We note that we do not correct for multiple hypothesis testing in the non-pre-specified outcomes and that this is the one statistically significant result out of 26 coefficients in question on Tables 6 to 8.

# 8 Conclusion

In this study, we address challenges sellers and consumers face in developing markets in the form of information and search frictions. These challenges are mitigated yet persistent in a world with access to information and communication technologies and online marketplaces (Aker 2010; Jensen 2007; Dinerstein et al. 2018; Bai et al. 2020). We conduct a natural field experiment with a theoretical framework to understand how sellers on online markets under persistent frictions internalize such frictions with available tools and how spillovers occur when an information environment is exogenously altered. We conduct a natural field experiment with a dominant online listing platform for used vehicles, PakWheels.com, in which we privately provide Price Calculator estimates to sellers when they create a new post. We track the sellers' list pricing choices, advertising usage, number of page views received, and transaction outcomes.

We find that the information intervention affects observed choices on list prices and advertising. The intervention brings the listing price closer to the Price Calculator estimates yet increases transaction probability only for the spillover group. We also find that the intervention increases the potential buyers' attention to spillover posts and reduces advertisement usage by the directly treated sellers. We hypothesize that these effects may be driven by changes in beliefs about demand distributions and search frictions. We apply a pre-specified conceptual framework to evaluate the set of pre-specified empirical results and assess the validity of the mechanisms we propose via survey outcome measures on sellers' beliefs.

We find evidence consistent with the idea that sellers' beliefs about demand and search frictions play a role in the effects of information interventions. First, the information intervention only affects the beliefs about the price levels, information and search frictions, and other market conditions of those who are directly treated. Second, we find evidence that is consistent with the view that sellers use advertising as a substitute for list-pricing. As such, adjustments via changes to the consumption of advertising tools could countervail the effects of information intervention. Third, the information intervention can generate significant spillover effects, not only through the direct spillover of information itself but also through changes in choices like pricing and advertising usage made by the directly treated. This set of results suggests that spillovers and general-equilibrium effects could propagate direct effects of information intervention through sellers' choices of list pricing and advertising.

Our findings show how information interventions via an online platform in developing economies affect price dispersion and market outcomes and offer novel mechanisms through which spillovers can occur. Our findings are in line with previous ICT-based information interventions, such as Aker (2010), Aker and Mbiti (2010), and Jensen (2007), in that we find significant reductions in price dispersion even in the context of heterogeneous goods. Our work also shows that information interventions can cause spillovers and knock-on effects not only through the supply chains as in Jensen and Miller (2018) and Hasanain et al. (2019), but also *horizontally* through choices that competing sellers make on pricing and advertising. Our work shows the importance of accounting for externalities generated from information intervention on sellers within markets for policymakers considering scaling information interventions.

Our work also highlights the potentials and limitations of improving market access and reducing frictions via online marketplaces in developing economies (Bai et al. 2020; Falcao Bergquist and McIntosh 2021 Couture et al. 2018; Fernando et al. 2020; Jeong 2020). Our work highlights the importance of complementary and substitute tools to pricing that sellers use to counter market frictions in developing economies, even in the context in which information and search frictions are reduced through technology. More work is needed to understand how information and marketing strategies interact with other channels on which entrepreneurs in developing economies rely, such as social networks and relational contracting. Lastly, our findings may also be relevant to the literature on small-to-medium enterprises in developing economies and business training interventions, which have had relatively low cost-effectiveness due to high cost (McKenzie and Woodruff 2014; Blattman and Ralston 2015). Our work shows the potential of improving information access to small-scale traders at scale and also points out the importance of advertising as a jointly determined tool.

# References


Shilpa Aggarwal, Brian Giera, Dahyeon Jeong, Jonathan Robinson, and Alan Spearot. Market Access, Trade Costs, and Technology Adoption: Evidence from Northern Tanzania. *The Review of Economics and Statistics*, pages 1–45, November 2022. ISSN 0034-6535. doi: 10.1162/rest_a_01263. URL https://doi.org/10.1162/rest_a_01263.

Jenny Aker. Information from Markets Near and Far : Mobile Phones and Agricultural Markets in Niger. *American Economic Journal: Applied Economics*, 2(3):46–59, 2010.

Jenny Aker and Isaac Mbiti. Mobile Phones and Economic Development in Africa. *Journal of Economic Perspectives*, 24(3):207–232, 2010. ISSN 00220388. doi: 10.1080/00220388.2012.709615.

Treb Allen. Information Frictions in Trade. *Econometrica*, 82(6):2041–2083, 2014. doi: 10.3982/ecta10984.

Tahir Andrabi, Jishnu Das, and Asim Ijaz Khwaja. Report cards: The impact of providing school and child test scores on educational markets. *American Economic Review*, 107(6): 1535–1563, 2017. ISSN 00028282. doi: 10.1257/aer.20140774.

David Atkin and Dave Donaldson. Who's getting globalized? The size and implications of intra-national trade costs. 2015.

Jie Bai, Maggie Xiaoyang Chen, Jin Liu, and Daniel Yi Xu. Search and Information Frictions on Global E-Commerce Platforms: Evidence from AliExpress. 2020.

Sarah Baird, J Aislinn Bohren, Craig McIntosh, and Berk Özler. Optimal design of experiments in the presence of interference. *Review of Economics and Statistics*, 100(5):844–860, 2018.

Michael R. Baye, John Morgan, and Patrick Scholten. Information, Search, and Price Dispersion. In T. Hendershott, editor, *Handbook on Economics and Information Systems*, pages 323–375. Elsevier, 2007. doi: 10.1016/s1574-0145(06)01006-3.

Lauren Falcao Bergquist and Craig McIntosh. Search Cost, Intermediation, and Trade, Experimental evidence from Ugandan agricultural markets. 2021.

Christopher Blattman and Laura Ralston. Generating Employment in Poor and Fragile States: Evidence from Labor Market and Entrepreneurship Programs. 2015.

Victor Couture, Benjamin Faber, Yizhen Gu, and Lizhi Liu. E-Commerce Integration and Economic Development: Evidence from China. 2018.

Peter A Diamond. Aggregate Demand Management in Search Equilibrium. *Journal of Political Economy*, 90(5):881–894, 1982.

Michael Dinerstein, Liran Einav, Jonathan Levin, and Neel Sundaresan. Consumer price search and platform design in internet commerce. *American Economic Review*, 108(7): 1820–1859, 2018. ISSN 00028282. doi: 10.1257/aer.20171218.

Liran Einav, Theresa Kuchler, Jonathan Levin, and Neel Sundaresan. Assessing sale strategies in online markets using matched listings. *American Economic Journal: Microeconomics*, 7(2):215–247, 2015. ISSN 19457685. doi: 10.1257/mic.20130046.

Lauren Falcao Bergquist and Craig McIntosh. Search Cost, Intermediation, and Trade: Experimental Evidence from Ugandan Agricultural Markets. August 2021. doi: 10.26085/C3759K. URL https://escholarship.org/uc/item/6fp1r637.

A. Nilesh Fernando, Gabriel Tourek, and Niharika Singh. Hiring Frictions in Small Firms: Evidence from an Internet Platform-based Experiment. 2020.

Andrey Fradkin. Search Frictions and the Design of Online Marketplaces. 2015. doi: 10.4108/eai.8-8-2015.2260850.

Xiaolan Fu, Elvis Avenyo, and Pervez Ghauri. Digital platforms and development: a survey of the literature. *Innovation and Development*, 11(2-3):303–321, September 2021. ISSN 2157-930X. doi: 10.1080/2157930X.2021.1975361. URL https://doi.org/10.1080/2157930X.2021.1975361. Publisher: Routledge _eprint: https://doi.org/10.1080/2157930X.2021.1975361.

Ali Hasanain, Muhammad Yasir, and Khan Arman. No bulls : Experimental evidence on the impact of veterinarian ratings in Pakistan. 2019.

John J. Horton. Buyer uncertainty about seller capacity: Causes, consequences, and a partial solution. *Management Science*, 65(8):3518–3540, 2019. ISSN 15265501. doi: 10.1287/mnsc.2018.3116.

Robert Jensen. The Digital Provide: Information (Technology), Market Performance, and Welfare in the South Indian Fisheries Sector. *The Quarterly Journal of Economics*, 122 (3):811–846, 2007. doi: 10.1080/02724980343000242.

Robert Jensen and Nolan H. Miller. Market integration, demand, and the growth of firms: Evidence from a natural experiment in India. *American Economic Review*, 108(12):3583–3625, 2018. ISSN 19447981. doi: 10.1257/aer.20161965.

Dahyeon D J Jeong. Creating ( Digital ) Labor Markets in Rural Tanzania. 2020.

John A. List. Field Experiments: A Bridge Between Lab and Naturally-Occurring Data, March 2007. URL https://www.nber.org/papers/w12992.

David McKenzie and Christopher Woodruff. What are we learning from business training and entrepreneurship evaluations around the developing world? *World Bank Research Observer*, 29(1):48–82, 2014. ISSN 02573032. doi: 10.1093/wbro/lkt007.

Sandip Mitra, Dilip Mookherjee, Maximo Torero, and Sujata Visaria. Asymmetric information and middleman margins: An experiment with Indian potato farmers. *Review of Economics and Statistics*, 100(1):1–13, 2018. ISSN 15309142. doi: 10.1162/REST_a_00699.

Ryan T. Moore. Multivariate continuous blocking to improve political science experiments. *Political Analysis*, 20(4):460–479, 2012. ISSN 10471987. doi: 10.1093/pan/mps025.

Pakistan Bureau of Statistics. Pakistan social and living standards measurement survey 2019-2020. Technical report, 2020.

Joseph P. Romano and Michael Wolf. Stepwise multiple testing as formalized data snooping. *Econometrica*, 73(4):1237–1282, 2005. ISSN 00129682. doi: 10.1111/j.1468-0262.2005.00615.x.

Meredith Startz. The Value of Face-to-Face: Search and Contracting Problems in Nigerian Trade, November 2016. URL https://papers.ssrn.com/abstract=3096685.

George J. Stigler. The Economics of Information. *Journal of Political Economy*, 69(9): 213–225, 1961. ISSN 1098-6596.

# 9 Tables

Table 1: **Balance table: mean by treatment group**

| | Pure control (N=63242) | | Assigned (N=50619) | | Spillover (high) (N=2185) | | Spillover (low) (N=30514) | |
|---|---|---|---|---|---|---|---|---|
| | Mean | Std. Dev. | Mean | Std. Dev. | Mean | Std. Dev. | Mean | Std. Dev. |
| log(Price difference) | 11.422 | 1.814 | 11.197 | 1.910 | 11.168 | 1.806 | 11.262 | 1.849 |
| 1 if sold | 0.339 | 0.474 | 0.351 | 0.477 | 0.358 | 0.480 | 0.356 | 0.479 |
| log(Trnsaction price) | 14.342 | 0.650 | 14.091 | 0.713 | 13.947 | 0.712 | 14.179 | 0.712 |
| Page view index | -0.003 | 0.580 | -0.004 | 0.574 | 0.053 | 0.608 | -0.015 | 0.572 |
| Advertising index | -0.052 | 0.523 | -0.108 | 0.442 | -0.107 | 0.444 | -0.111 | 0.437 |

*Notes:* Mean outcomes by the mutually exclusive treatment group. "log(Price difference)": log of the absolute difference between listing price and Price Calculator estimate. "1 if sold": a binary outcome that is 1 if the seller reports the car as sold. "log(Transaction price)": log of the self-reported transaction price. "Page-view index": a standardized index of page-view measures. "Advertising index": a standardized index of advertising usage by the seller.

Table 2: **Balance table: placebo regressions (ITT) on pre-specified main outcomes**

| | log(Price difference) (1) OLS | 1 if sold (2) Logit | log(Transaction price) (3) OLS | Page-view index (4) OLS | Advertising index (5) OLS |
|---|---|---|---|---|---|
| Assignment | -0.0209 | -0.0317* | -0.0015 | -0.0067 | 0.0061* |
| | (0.0195) | (0.0190) | (0.0025) | (0.0065) | (0.0035) |
| Spillover | 0.0074 | 0.0318 | -0.0063 | 0.0512** | -0.0094 |
| | (0.0543) | (0.0420) | (0.0556) | (0.0197) | (0.0075) |
| Spillover (high) | -0.0120 | 0.0304 | -0.0751 | -0.0136 | 0.0056 |
| | (0.0540) | (0.0431) | (0.0662) | (0.0212) | (0.0063) |
| Observations | 104,485 | 116,314 | 19,222 | 117,715 | 117,715 |
| Squared Correlation | 0.05454 | 0.01119 | 0.89064 | 0.09523 | 0.25887 |
| Pseudo $R^2$ | 0.01385 | 0.00882 | 1.0964 | 0.05739 | 0.21179 |
| BIC | 419,391.9 | 151,714.6 | -1,993.4 | 195,543.5 | 133,310.6 |
| Q-values: Assignment | 0.383 | 0.237 | 0.565 | 0.383 | 0.237 |
| Q-values: Spillover | 0.91 | 0.749 | 0.91 | 0.053 | 0.535 |
| Q-values: Spillover (high) | 0.826 | 0.655 | 0.655 | 0.655 | 0.655 |

*Notes:* The outcomes are as defined in Section 4. "log(Price difference)": log of the absolute difference between listing price and Price Calculator estimate. "1 if sold": a binary outcome that is 1 if the seller reports the car as sold. "log(Transaction price)": log of the self-reported transaction price. "Page-view index": a standardized index of page-view measures. "Advertising index": a standardized index of advertising usage by the seller. The specification is as shown in Equation 1. Standard errors are reported in parentheses and clustered at the make-model level. The stars show the two-tailed significance in p-values: p<0.1*; p<0.05**; p<0.01***. Registered tests of statistical significance are q-values, which are reported in the bottom three rows.

Table 3: **Experimental results (ITT) on main pre-specified outcomes**

| | log(Price difference) | 1 if sold | log(Transaction price) | Page-view index | Advertising index |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| | OLS | Logit | OLS | OLS | OLS |
| Assignment | -0.0327** | -0.0499*** | -0.0008 | -0.0172*** | -0.0095*** |
| | (0.0135) | (0.0172) | (0.0037) | (0.0039) | (0.0025) |
| Spillover | -0.0779* | 0.0488*** | -0.0401 | 0.0332*** | -0.0005 |
| | (0.0443) | (0.0140) | (0.0335) | (0.0106) | (0.0042) |
| Spillover (high) | 0.0736 | -0.0138 | -0.0031 | -0.0092 | 0.0062 |
| | (0.0537) | (0.0333) | (0.0457) | (0.0169) | (0.0059) |
| Observations | 101,750 | 111,309 | 14,084 | 117,891 | 117,891 |
| Squared Correlation | 0.10797 | 0.01471 | 0.92874 | 0.12322 | 0.29329 |
| Pseudo $R^2$ | 0.02959 | 0.01197 | 1.2997 | 0.08546 | 0.24067 |
| BIC | 383,275.7 | 141,831.7 | -6,886.7 | 167,975.7 | 131,198.5 |
| Q-values: Assignment | 0.023 | 0.006 | 0.835 | 0.000 | 0.001 |
| Q-values: Spillover | 0.14 | 0.002 | 0.293 | 0.006 | 0.912 |
| Q-values: Spillover (high) | 0.741 | 0.848 | 0.945 | 0.848 | 0.741 |

*Notes:* The table presents the intent-to-treat estimates on the main, pre-specified outcomes. The specification is as shown in Equation 1. The outcomes are as defined in Section 4. "log(Price difference)": log of the absolute difference between listing price and Price Calculator estimate. "1 if sold": a binary outcome that is 1 if the seller reports the car as sold. "log(Transaction price)": log of the self-reported transaction price. "Page-view index": a standardized index of page-view measures. "Advertising index": a standardized index of advertising usage by the seller. Standard errors are reported in parentheses and clustered at the make-model level. The stars show the two-tailed significance in p-values: p<0.1*; p<0.05**; p<0.01***. Registered tests of statistical significance are q-values, which are reported in the bottom three rows.

Table 4: **Experimental results (ToT) on main pre-specified outcomes**

| | log(Price difference) | 1 if sold | log(Transaction price) | Page-view index | Advertising index |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Treatment | -0.0440** | -0.0160*** | -0.0011 | -0.0252*** | -0.0140*** |
| | (0.0180) | (0.0060) | (0.0055) | (0.0067) | (0.0041) |
| Spillover | -0.0775* | 0.0103*** | -0.0401 | 0.0329*** | -0.0007 |
| | (0.0442) | (0.0030) | (0.0335) | (0.0106) | (0.0042) |
| Spillover (high) | 0.0743 | -0.0043 | -0.0032 | -0.0110 | 0.0053 |
| | (0.0541) | (0.0075) | (0.0456) | (0.0174) | (0.0061) |
| Observations | 101,750 | 111,312 | 14,084 | 117,891 | 117,891 |
| $R^2$ | 0.10787 | 0.01464 | 0.92873 | 0.12284 | 0.29303 |
| Within $R^2$ | 0.02809 | 0.00451 | 0.75561 | 0.01610 | 0.00609 |
| Q-values: Assignment | 0.022 | 0.015 | 0.835 | 0.001 | 0.003 |
| Q-values: Spillover | 0.14 | 0.004 | 0.293 | 0.006 | 0.869 |
| Q-values: Spillover (high) | 0.705 | 0.705 | 0.945 | 0.705 | 0.705 |

*Notes:* The table presents the treatment-on-the-treated estimates on the main, pre-specified outcomes. The 2SLS specification is as shown in Equation 2. The outcomes are as defined in Section 4. "log(Price difference)": log of the absolute difference between listing price and Price Calculator estimate. "1 if sold": a binary outcome that is 1 if the seller reports the car as sold. "log(Transaction price)": log of the self-reported transaction price. "Page-view index": a standardized index of page-view measures. "Advertising index": a standardized index of advertising usage by the seller. Standard errors are reported in parentheses and clustered at the make-model level. The stars show the two-tailed significance in p-values: p<0.1*; p<0.05**; p<0.01***. Registered tests of statistical significance are q-values, which are reported in the bottom three rows.

Table 5: **Experimental results (ITT) on main pre-specified outcomes by treatment group**

| | log(Price difference) | 1 if sold | log(Transaction price) | Page-view index | Advertising index |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| | OLS | Logit | OLS | OLS | OLS |
| GroupSatAssigned(high) | -0.0316 | -0.0238 | -0.0441 | 0.0060 | -0.0018 |
| | (0.0372) | (0.0207) | (0.0370) | (0.0113) | (0.0053) |
| GroupSatAssigned(low) | -0.1135*** | 0.0046 | -0.0407 | 0.0166 | -0.0090** |
| | (0.0408) | (0.0212) | (0.0353) | (0.0122) | (0.0043) |
| GroupSatSpillover(high) | -0.0511 | 0.1117** | -0.0421 | 0.0312** | 0.0162** |
| | (0.0552) | (0.0552) | (0.0427) | (0.0119) | (0.0066) |
| GroupSatSpillover(low) | -0.0749* | 0.0431*** | -0.0401 | 0.0327*** | -0.0007 |
| | (0.0435) | (0.0145) | (0.0335) | (0.0106) | (0.0042) |
| Observations | 101,750 | 111,309 | 14,084 | 117,891 | 134,781 |
| Squared Correlation | 0.10799 | 0.01474 | 0.92874 | 0.12322 | 0.29973 |
| Pseudo $R^2$ | 0.02960 | 0.01199 | 1.2997 | 0.08546 | 0.25504 |
| BIC | 383,285.7 | 141,840.4 | -6,877.1 | 167,986.9 | 142,368.9 |
| Q-values: Assignment group (high) | 0.663 | 0.626 | 0.626 | 0.729 | 0.729 |
| Q-values: Assignment group (low) | 0.036 | 0.83 | 0.315 | 0.293 | 0.094 |
| Q-values: High spillover group | 0.359 | 0.072 | 0.359 | 0.038 | 0.038 |
| Q-values: Low spillover group | 0.15 | 0.007 | 0.292 | 0.007 | 0.872 |

*Notes:* The table presents the intent-to-treat estimates on the main, pre-specified outcomes. The specification is *not* identical to Equation 1, but instead, we regress outcomes on the following dummies: assigned to treatment in high-saturation model ("GroupSatAssigned(high)"), assigned to treatment in low-saturation model ("GroupSatAssigned(low)"), not directly assigned to treatment in high-saturation model ("GroupSatSpillover(high)"), and not directly assigned to treatment in low-saturation model ("GroupSatSpillover(low)"). The outcomes are as defined in Section 4. "log(Price difference)": log of the absolute difference between listing price and Price Calculator estimate. "1 if sold": a binary outcome that is 1 if the seller reports the car as sold. "log(Transaction price)": log of the self-reported transaction price. "Page-view index": a standardized index of page-view measures. "Advertising index": a standardized index of advertising usage by the seller. Standard errors are reported in parentheses and clustered at the make-model level. The stars show the two-tailed significance in p-values: p<0.1*; p<0.05**; p<0.01***. Registered tests of statistical significance are q-values, which are reported in the bottom three rows.

Table 6: **Regressions on survey measures: Sellers' expectations on prices**

| | log(absdff(Expectation)) | | | Amt. bargain | Searched listings |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| | OLS | OLS | OLS | OLS | Logit |
| Assignment | -0.2128*** | -0.0531 | -0.1945 | 5,561.1*** | 0.1338 |
| | (0.0765) | (0.0743) | (0.1692) | (1,670.5) | (0.1845) |
| Spillover | 0.0233 | -0.1126 | 0.0777 | -642.9 | -0.1345 |
| | (0.1025) | (0.1005) | (0.1069) | (2,697.6) | (0.1537) |
| Spillover (high) | 0.1712 | 0.1324 | 0.1398 | 3.690 | -0.0276 |
| | (0.1239) | (0.1246) | (0.1510) | (3,644.3) | (0.1391) |
| Observations | 2,046 | 2,045 | 2,045 | 2,321 | 2,185 |
| Squared Correlation | 0.09972 | 0.16112 | 0.10842 | 0.08766 | 0.05688 |
| Pseudo $R^2$ | 0.02475 | 0.04524 | 0.02552 | 0.00370 | 0.05098 |
| BIC | 9,618.9 | 8,733.0 | 10,113.1 | 58,576.7 | 3,010.0 |

*Notes:* The table presents the treatment-on-the-treated estimates on the survey outcomes. "log(diff(Expectation))": log absolute difference between the Price Calculator estimate and their expected transaction price, as measured through the survey. We asked price expectations in three ways: their initial expectation (column 1), the highest price they could have received (column 2), and the lowest price they could have received (column 3). "Amt. bargain": The amount in PKR that the seller is willing to bargain. "Searched listings": a binary outcome that is 1 if the seller reported having searched for or looked at other similar ads to their vehicles. The ITT specification is as shown in Equation 1. Standard errors are reported in parentheses and clustered at the make-model level. The stars show the two-tailed significance in p-values: p<0.1*; p<0.05**; p<0.01***.

Table 7: **Regressions on survey measures: Sellers beliefs about search and information frictions**

| | Difficult to get inquiry<br>(1) | Difficult to get good price<br>(2) | Buyers have good info<br>(3) | Sellers have good info<br>(4) |
|---|---|---|---|---|
| Assignment | -0.0658** | -0.0532** | 0.0275 | 0.0239 |
| | (0.0324) | (0.0232) | (0.0477) | (0.0392) |
| Spillover | $-8.42 \times 10^{-5}$ | -0.0163 | 0.0457 | 0.0221 |
| | (0.0286) | (0.0221) | (0.0316) | (0.0368) |
| Spillover (high) | 0.0197 | 0.0469 | -0.0868** | -0.0176 |
| | (0.0282) | (0.0306) | (0.0394) | (0.0292) |
| Observations | 2,311 | 2,311 | 2,310 | 2,310 |
| $R^2$ | 0.11718 | 0.13054 | 0.10224 | 0.10057 |
| Within $R^2$ | 0.00257 | 0.00212 | 0.00299 | 0.00100 |

*Notes:* The table presents the treatment-on-the-treated estimates on the survey outcomes. The outcomes on this table are in the Likert scale, where 1 = "strongly disagree" and 5 = 'strongly agree." "Difficult to get inquiry": It is difficult to get inquiries from potential buyers. "Difficult to get good price": It is difficult to get good price offers from potential buyers. "Buyers have good info": Buyers have good information about what fair prices for used vehicles are. "Sellers have good info": Sellers have good information about what fair prices for used vehicles are. The ITT specification is as shown in Equation 1. Standard errors are reported in parentheses and clustered at the make-model level. The stars show the two-tailed significance in p-values: p<0.1*; p<0.05**; p<0.01***.

Table 8: **Regressions on survey measures: Valuation of Price Calculator and advertising tools**

| | 1 if WTP at Rs100<br>(1)<br>Logit | WTP<br>(2)<br>OLS | Ad useful-high price<br>(3)<br>OLS | Ad useful-sell faster<br>(4)<br>OLS |
|---|---|---|---|---|
| Assignment | -0.2485 | -5.674 | 0.0239 | 0.0251 |
| | (0.1624) | (3.584) | (0.0392) | (0.0400) |
| Spillover | 0.0314 | 3.710 | 0.0221 | 0.0252 |
| | (0.1094) | (3.503) | (0.0368) | (0.0323) |
| Spillover (high) | 0.1901 | 3.429 | -0.0176 | -0.0152 |
| | (0.1313) | (2.107) | (0.0292) | (0.0286) |
| Observations | 2,247 | 2,261 | 2,310 | 2,301 |
| Squared Correlation | 0.05775 | 0.08011 | 0.10057 | 0.10072 |
| Pseudo $R^2$ | 0.04635 | 0.00781 | 0.06402 | 0.06679 |
| BIC | 3,578.5 | 25,149.6 | 4,764.8 | 4,597.6 |

*Notes:* The table presents the treatment-on-the-treated estimates on the survey outcomes. "1 if WTP at Rs100": a binary outcome that is 1 if the seller reports to be willing to pay PKR 100 for a Price Calculator estimate per post. "WTP": seller's willingness to pay for a Price Calculator estimate per post. "Ad useful-high price": features and bumps are useful for selling the vehicle faster, on a Likert scale (5 = 'strongly agree."). "Ad useful-sell faster": features and bumps are useful for getting a higher price for the vehicle, on a Likert scale (5 = 'strongly agree."). The ITT specification is as shown in Equation 1. Standard errors are reported in parentheses and clustered at the make-model level. The stars show the two-tailed significance in p-values: p<0.1*; p<0.05**; p<0.01***.

Table 9: **Regressions on survey measures: Price Calculator and transaction outcomes**

| | Seen PC (1) Logit | Others seen PC (2) Logit | 1 if sold (3) Logit | log(Transaction price) (4) OLS |
|---|---|---|---|---|
| Assignment | 0.7749*** | 0.4057** | 0.0207 | -0.0130 |
| | (0.1287) | (0.1622) | (0.1224) | (0.0179) |
| Spillover | 0.0013 | 0.1678 | 0.1606 | 0.0219 |
| | (0.1596) | (0.1060) | (0.1430) | (0.0372) |
| Spillover (high) | -0.0925 | -0.0550 | 0.0078 | -0.0694 |
| | (0.1238) | (0.1712) | (0.1727) | (0.0463) |
| | | | | |
| Observations | 2,202 | 2,195 | 2,280 | 1,397 |
| Squared Correlation | 0.09259 | 0.06356 | 0.09186 | 0.76180 |
| Pseudo $R^2$ | 0.08516 | 0.06132 | 0.06989 | 0.65933 |
| BIC | 3,150.7 | 2,823.5 | 3,820.6 | 2,034.9 |

*Notes:* The table presents the treatment-on-the-treated estimates on the survey outcomes. "Seen PC": a binary outcome that is 1 if the seller reports to have seen the Price Calculator estimate. 'Seen PC": a binary outcome that is 1 if the seller reports that other sellers have received the Price Calculator estimates. '1 if sold" and "log(Transaction price)": analogous to main outcomes but collected through the survey instead of the platform's database. The ITT specification is as shown in Equation 1. Standard errors are reported in parentheses and clustered at the make-model level. The stars show the two-tailed significance in p-values: p<0.1*; p<0.05**; p<0.01***.

# 10 Figures

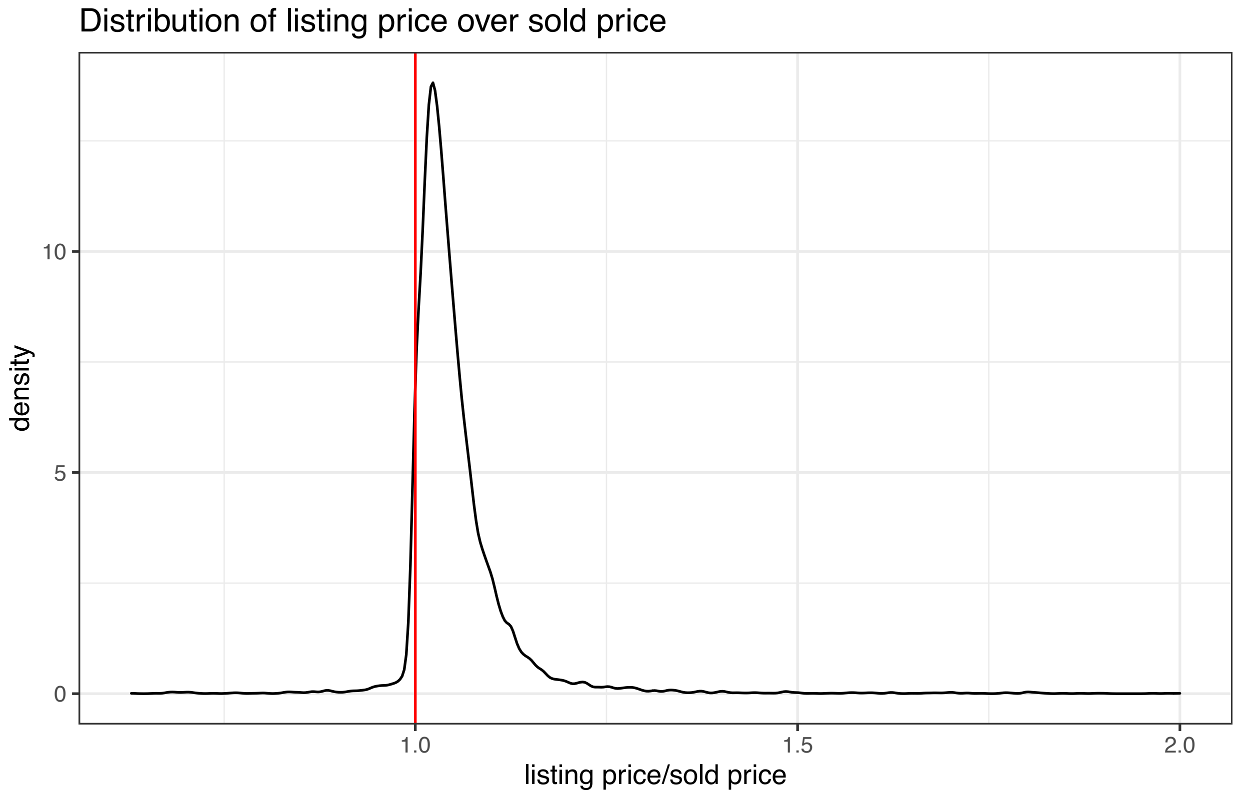



Figure 1: Distribution of the share of listing price to transaction price

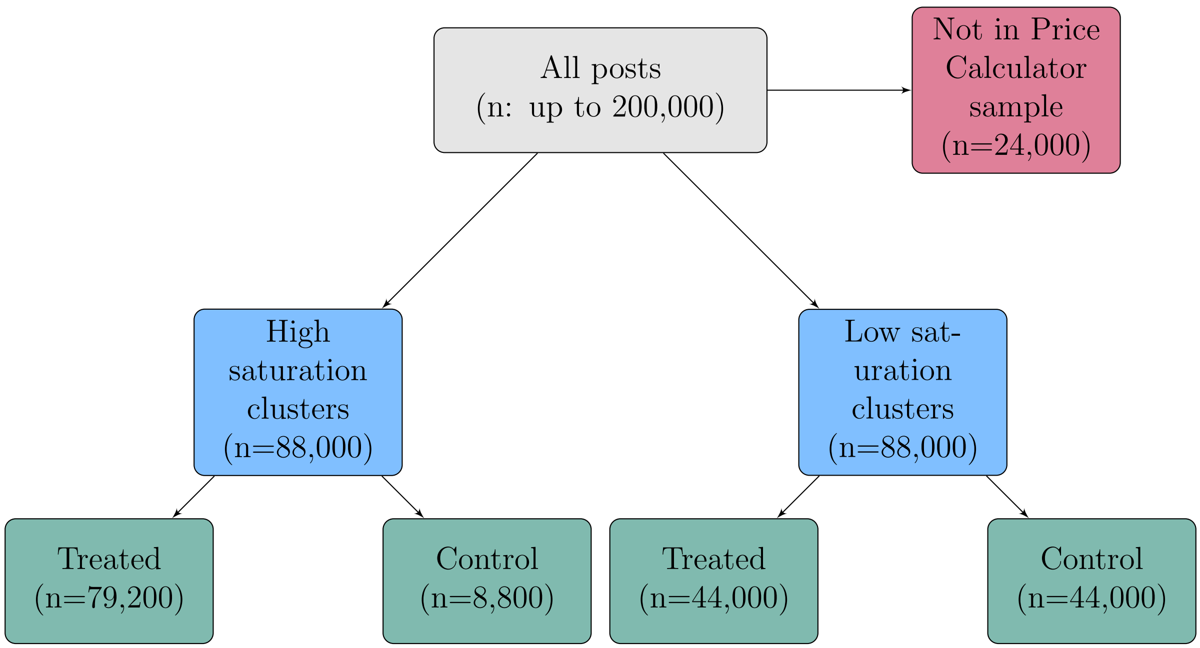



Figure 2: Treatment Groups

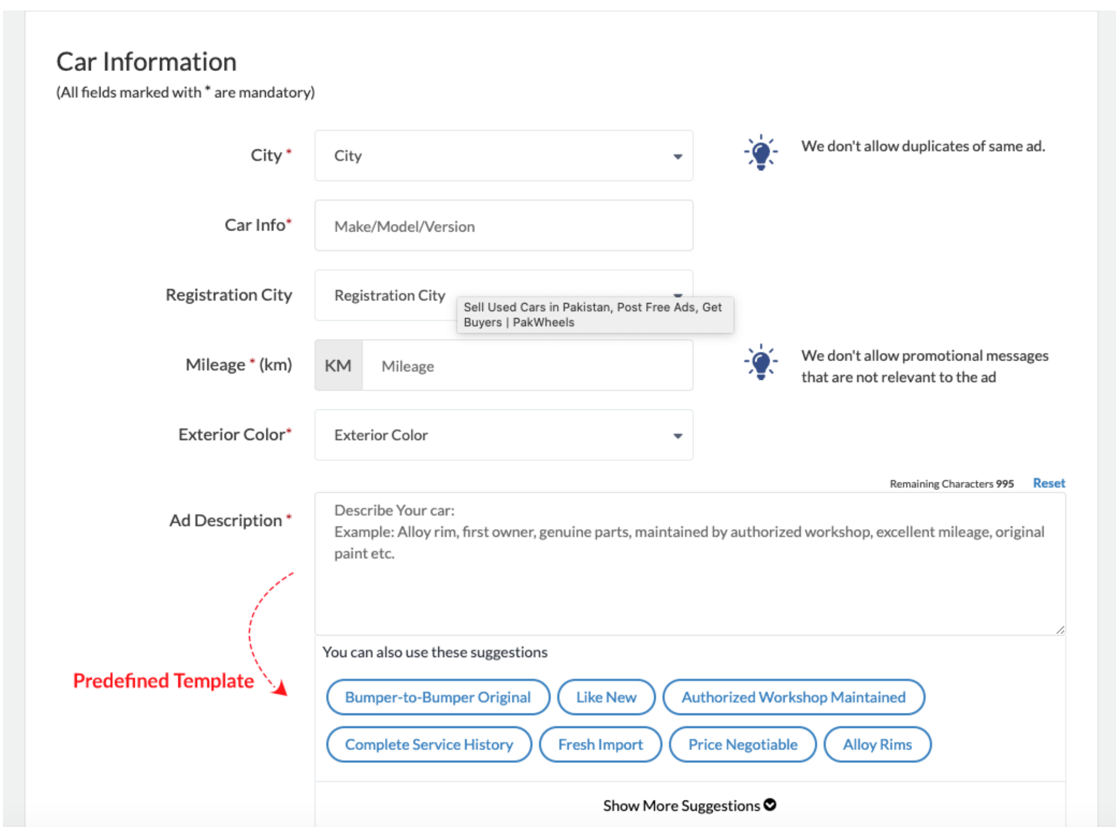



Figure 3: Making of a listing: Vehicle information

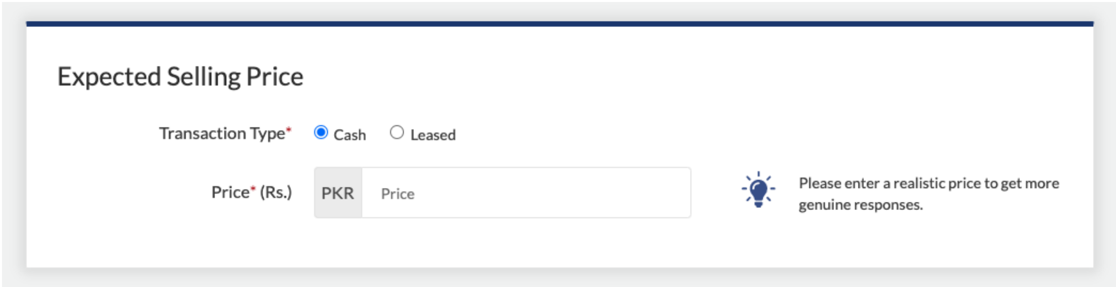



Figure 4: Making of a listing: Vehicle price

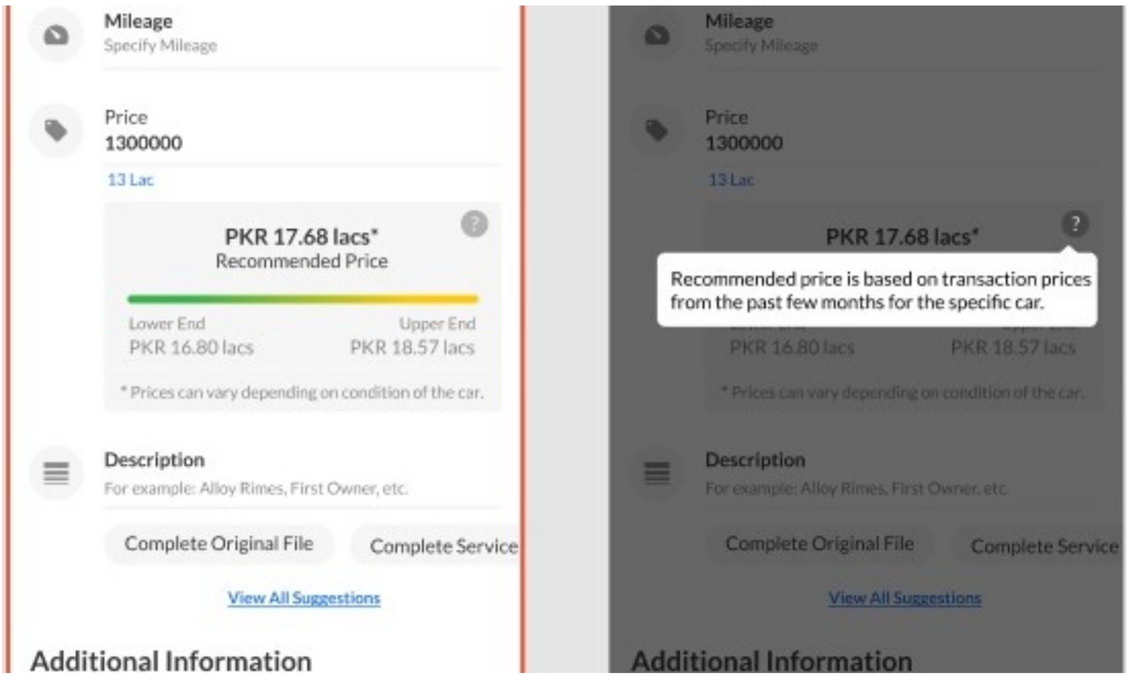



Figure 5: Display of the Price Calculator estimate

# A Data

## A.1 Posts

PakWheels' database tracks every post on the platform. Once a post is created, it is vetted against spam or fraud, made publicly available on the platform, then removed after 90 days or once the user asks for it to be taken down. We collect the following measures from the database:

- timing of the post's creation, approval, and closure
- vehicle characteristics
  - basic information such as make, model, model year, mileage, sellers' location, and registration city
  - additional information about vehicle characteristics, such as version, assembly, engine size, and capacity
- listing price
- self-reported transaction outcome (e.g., sold to a customer on the platform, sold through other means, decided not to sell)
- self-reported transaction price, if sold.

The database also tracks any updates to variables over the course of posts' active status. This allows us to capture sellers' initial choice of the listing price before and after exposure to the Price Calculator estimate.

## A.2 Advertising tools and vehicle inspection services

PakWheels' database also tracks users' platform-credit purchases and usages, which we consider to be measures of sellers' advertising efforts. Users on PakWheels have two primary tools for advertising: "bump" and "feature" credits. A "bump" credit allows sellers to bring their post to the top of the result page in the default, reverse-chronological listing order. This effectively increases the post's visibility as more people look at the first pages of listings. On the other hand, a "feature" credit would put their post in a few reserved spots at the top of the result page and label it as a "featured ad", in a similar way as promoted ads on Google searches. Posts are otherwise listed in the reversed chronological order within the class of featured ads.

Another way for sellers to attract buyers' attention to their posts is to provide signals of vehicle quality. In order to do so, sellers can request in-person inspections by PakWheels' mechanics, who give scores (out of 100) on eight dimensions (engine, brakes, suspension, interior, AC, electrical, exterior, and tires) based on a pre-specified rubric. The vehicles will pass the inspection if the unweighted average of scores over these eight dimensions is above a threshold. They can then be marked as "PakWheels certified" on the platform for an additional fee. Because certification is endogenous to vehicle quality, we use the data on whether or not the vehicle was ever inspected, as opposed to certified.

### A.2.1 Expenditures on advertising tools

Another way of expressing sellers' advertising choices would be in terms of the amount paid to the platform for advertising. This is made difficult, however, by the fact that credits for bumps and features are purchased in bundles, and users can apply them to any posts that they own. We, therefore, do not plan on using this measure as a primary outcome. Nonetheless, we collect data on advertising expenditures for robustness checks where the unit of analysis is the user. The data set contains information on purchase timing, descriptions of the bundles or services, quantity, and prices.

## A.3 Buyer-attention measures

One of our hypotheses is that the price information intervention and resulting changes to pricing and advertising would affect buyers' attention to certain posts. In order to construct measures of buyer attention, we access PakWheels' data on views and clicks at the post level. We are able to collect cumulative measures of the following:

- page views (i.e., clicks on the post)
- clicks on the "Show Phone Number" button within the post to contact the seller.

We also capture the number of times each post appears on search listings. We run an analysis including this measure in the index as an alternative specification, as well.

## A.4 Endline survey

We use self-reported transaction prices to train and test the Price Calculator estimates. These self-reported data are collected in an online form whenever sellers choose to take their posts down. There are concerns about the accuracy of reported transaction prices, particularly for the following reasons:

1. Transaction prices may be selectively reported (e.g., if those who fetched a higher price are more likely to report).
2. Conditional on reporting, sellers may obfuscate the true value, leading to more noise in the price estimate.

3. Conditional on reporting, sellers may feel that the reported price should be inflated or deflated (because of, for example, their beliefs about a fair transaction price or the desire to appear successful).

4. Conditional on reporting, sellers may find it easy just to repeat the listing price they have already given for their post.

5. Conditional on reporting, sellers may simply put a random number down to "get it out of the way."

We can identify the extent of number 4. by comparing transaction prices with listing prices and address 5. via data cleaning. Concerns like number 2. may introduce noise but should not bias the Price Calculator estimate or our empirical analysis.

We are unable to directly address concerns 1. and 3. from the data, nor is it realistic for us to request sales receipts or access other independent sales records. Instead, we randomly selected 3,000 listings (stratified over the vehicle model) and conducted a short phone survey of the sellers. We survey owners of the posts between 3 to 5 weeks after their initial listing. The primary objective of the survey is to collect an independent measure of self-reported transactions and transaction prices from the platform-collected counterparts. In addition, we ask about sellers' expectations about transaction prices, beliefs about search and information frictions, and demand for, and perceived effectiveness of, the Price Calculator and advertising tools. [6]

## A.5 Search engine logs

Aggregated search engine logs tell us which combinations of terms are used most frequently by viewers on PakWheels. We use these aggregate statistics for our justifications for market

[6] The survey questionnaire is included in Appendix Section G.

cluster groupings. Our objective is to minimize concerns about inter-cluster interference but also retain as many randomization clusters for the step as possible. Our aggregate search logs data are taken from the month of August 2020. They represent tens of millions of searches over the month, and our data contain numbers of searches per combination of search terms (e.g., make, model, model-years in range, city, range of listing prices). We capture 35,000 most common search combinations, which account for 93% of all searches. We use these data for our definition of clusters in Section 3.2.1.

## A.6 Listing orders

Beyond the primary analysis, in which we measure the average spillover effects on treated clusters, we plan to assess the extent to which the spillover effects depend on the "proximity" to treated posts, such as how close a given ad is to treated peers in ad listings. For this, we web-scrape listing orders in their default, reverse-chronological order on a daily basis for each make-model cluster in the sample.

## A.7 Use of an old Price Calculator

We also track the usage of a previous version of the Price Calculator, which our intervention replaces. The previous iteration of the Price Calculator was designed and implemented prior to the beginning of our research collaboration with PakWheels. It was contained in a separate module on PakWheels' website and mobile apps, unintegrated with the posting process, and was discontinued at the end of December 2020. The old Price Calculator offered predictions to only a handful of make-model-model years of certain colors, locations, and mileage. PakWheels keeps a log of all price estimates the old Price Calculator provided at each instance. This dataset contains the user ID, search inputs (make, model, model year, location, mileage, and if seller or buyer), the price estimates, and the time stamp.

# B Secondary outcomes listed in the pre-analysis plan

### B.0.1 Survey data

The first-order objectives of the endline phone survey are to confirm reported transaction outcomes on PakWheels' platform and elicit sellers' beliefs about transaction prices. Those outcomes are listed in bold below. We also collect the following measures from survey respondents:

1. validation of self-reported transaction outcome
   - **1 if the vehicle is sold**
   - **transaction price (if sold)**
   - reasons for not selling the vehicle (if not sold)
   - relationship with the buyer (if sold)
2. price elicitation (stated beliefs)
   - **Expected transaction price at the time of initial posting**
   - lower and upper bounds of the expected transaction price
3. purchase price
4. number of vehicles previously traded
5. recall and salience of the Price Calculator instrument
   - 1 if the seller recalls seeing the Price Calculator estimate
   - recall of the Price Calculator estimate
   - beliefs about Price Calculator's accuracy

6. search and information acquisition
    - if the seller searches for other posts on PakWheels
    - terms used for the search
    - other sources of information
7. stated beliefs about challenges and frictions on the market
    - if the seller believes it is difficult to receive enough inquiries on PakWheels
    - if the seller believes it is difficult to receive acceptable price offers on PakWheels
8. stated beliefs about the usefulness of the advertising tools offered by PakWheels (i.e., bumps and features)
9. stated willingness to pay for the Price Calculator estimates

#### B.0.2 Post's duration on the platform

PakWheels' database reports when each post is created and taken down, so we can calculate the duration of the post's active status on the platform. One challenge is that posts may be left inactive for a period of time, so this would not be a measure of sellers' active participation in the market. This makes it difficult to interpret the meaning of any causal effect on this variable other than in aggregate as a measure of market congestion. For this reason, we consider this as a secondary outcome.

#### B.0.3 Price changes, and convergence to estimated price

We have chosen the logged absolute difference between the listing price and the Price Calculator estimate as a primary outcome variable. It is possible, however, that the treatment effect on the list price may be better captured if the Price Calculator induces a level shift in price or affects whether sellers ever adjust their initial listing prices. It is also possible that the treatment effect on the listing price is asymmetrical around the Price Calculator estimate. We intend to address this possibility with following alternative outcomes pertaining to the listing price as robustness checks:

- log(list price)
- 1 if the listing price is ever modified
- difference between the initial and final listing prices.

### B.0.4 Cluster-level outcomes

Our two-stage randomization procedure allows us to estimate the impact on cluster-level outcomes because of the two-stage design. Part of our secondary analysis focuses on cluster-level aggregate measures of moments of prices, page views, and post duration. We construct the following variables at the cluster-day level:

- number of new posts
- number of active posts
- standard deviation and kurtosis of the listing price
- standard deviation and kurtosis of page views.

### B.0.5 Spillovers based on listing order

For our primary analysis, we use the post-level data to identify the treatment effects of treatment and spillover assignment when the posts are initially created. An alternative conception of potential spillover, however, maybe that it is a function of exposure to treated posts over the listing space over time. To create a measure of potential spillover intensity based on proximity to treated posts, we web-scrape data on the listing order from PakWheels.

We define outcomes from PakWheels' scraped data as follows:

- number of days a post is on the first page of the make-model level search result
- average page number of the search results over the course of its active status.

We also construct the following variables as proxies of exposure to treated posts:

- number of days spent being adjacent to at least one treated post
- average number of treated posts within its listing result page (i.e., if the post resides on page 5 in a given day, then we take the number of treated posts on page 5), throughout the post's active status.

# C Details on Conceptual Framework

## C.1 Set-up

Suppose that we have a seller $i$, who is endowed with an asset. The asset- and seller-characteristics are denoted as $s_i$, and information set $\mathcal{I}_i$. The search and transaction process is as follows:

- Seller $i$ forms a prior belief about the demand distribution for their asset based on the information set $\mathcal{I}_i$ and characteristics $s_i$.
- Some sellers are provided with an information signal, i.e., the Price Calculator estimate denoted as $x_i$.

- If treated, seller $i$ forms a posterior belief about the demand distribution based on the information signal $x_i$, their belief in the quality of the signal, and their prior.
- Seller $i$ chooses a listing price $p^l$ and amount of advertisements $a$, based on their posterior belief about the demand distribution and their characteristics $s_i$.
- Choices of $p^l$ and $a$ affect the distribution of potential buyers with whom seller $i$ is matched via a Poisson process.
- Once a match occurs, seller $i$ makes a take-it-or-leave-it (TIOLI) offer $p^t$ below $p^l$ to the potential buyer.
- Transaction occurs if matched buyer's WTP is higher than $p^t$.

We denote the probability density function (PDF) of true buyer WTP as $f(\theta)$, and the distribution of potential buyers that get matched to the seller, conditional on $p^l$, as $g(\theta; p^l)$ and their cumulative equivalents, $F$ and $G$. The distinction between $F(\theta)$ and $G(\theta; p^l)$ is key since we assume that the seller's choice of the listing price $p^l$ skews the distribution of potential buyers (who may click on the post depending on the listing price) towards $p^l$ itself. Setting too high of $p^l$ also comes at a cost, as we make the following assumption:

$$\int_{-\infty}^{\infty} g(\theta; p^l) d\theta \leq 1 \tag{4}$$

$$\frac{\delta}{\delta p^l} \int_{0}^{\infty} g(\theta; p^l) d\theta < 0 \tag{5}$$

In other words, the distribution $g$ is a subset of $f$ and does not add up to one. In other words, $g$ does not add up to one, and high values of $p^l$ effectively reduce the pool of buyers to match with. The relationship between $p^l$ and $g(\theta; p^l)$ are also described schematically in Figure C.1:

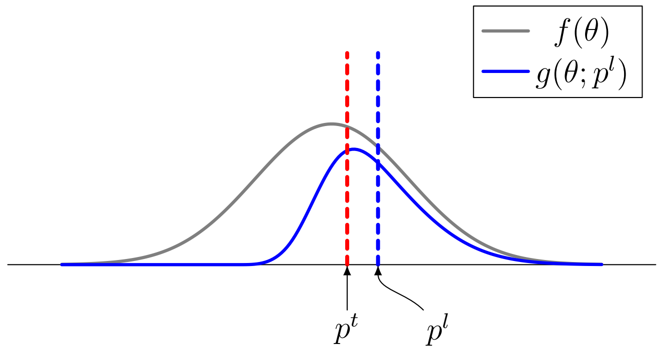



Figure C.1: Relationship between list price, the buyer it draws, and the TIOLI price

## C.2 The objective function

Under no information friction, seller $i$ chooses the listing price and advertisements to maximize the following:

$$V(p^l, a; s_i) = -c - k(a) + \gamma(a) \int \max_{p^t} [\mathbb{E}\pi(p^t; p^l, s_i)] g(\theta; p^l)\, d\theta \tag{6}$$

Sellers incur a constant cost of search, denoted as $c$. They also incur a variable cost $k()$, based on the amount spent on advertising, $a$. The term $\gamma(a)$ is a Poisson match rate between the seller and a potential buyer, and it is an increasing function with respect to $a$. We denote the seller's utility from the transaction as $\pi(p^t; s_i)$. This is a function not of the listing price but of the eventual offer price $p^t$, which is discussed below. $\pi$ is also not strictly a profit term, as sellers may also have preferences over how quickly to sell the vehicle, as captured in $s_i$. We assume the function is continuously differentiable and concave with respect to its only choice variable $p^t$, so that there is a single global maximum that is conditional on individual characteristics $s_i$.

### C.2.1 The TIOLI price $p^t$

Seller $i$ sets the listing price $p^l$, keeping in mind the distribution of buyers the list price attracts and the (TIOLI) offer price $p^t$ that seller $i$ would then select. We assume that there is a one-to-one correspondence between $p^l$ and $p^t$ conditional on seller $i$'s characteristics. We also assume that potential buyers cannot perfectly infer $p^t$ from $p^l$, because this depends on the seller's individual characteristics $s_i$ as well as $\mathcal{I}_i$. This allows us to express the seller's problem of maximizing the value function $V()$ as choices of $p^l$ and $a$.

Based on the mapping we assume between $p^l$ and $p^t$, we can also express Equation 6 as follows:

$$V(p^l, a; s_i) = -c - k(a) + \gamma(a)\pi(p^t(p^l; s_i))\Omega(p^t(p^l), s_i), \tag{7}$$

$$\text{where } \Omega(p^t(p^l), s_i) = \int_{p^t(p^l)}^{\infty} g(\theta; s_i, p^l)d\theta \tag{8}$$

$\Omega$ is a function that represents the probability that a potential buyer's willingness to pay is greater than the TIOLI offer price, given the listing price $p^l$ chosen by the seller. In order to ensure a unique and interior solution to the problem, we assume that $\Omega$ is decreasing and concave with respect to $p^l$; As $p^l$ increases, fewer buyers are drawn to the listing and have a WTP greater than the TIOLI price associated with $p^l$. This ensures that the objective function in Equation 7 is quasiconcave with respect to its argument $p^l$.

## C.3 Identifying optimal $p^l$ and $a$

Taking the first-order condition of Equation 7 with respect to $p^l$ gives the following expression, where we see that the choice of optimal $p^l$ is independent of $a$ under no information friction.

$$0 = \frac{dV}{dp^l} = \gamma(a)[\pi'(p^t)\frac{dp^t}{dp^l}\Omega(p^t(p^l), s_i) + \pi(p^t(p^l; s_i))\frac{d\Omega(p^t(p^l), s_i)}{dp^l}\frac{dp^t}{dp^l}] \tag{9}$$

Rearranging and simplifying Equation 9, we get:

$$\Omega(p^t(p^l), s_i)\pi'(p^t)\frac{dp^t}{dp^l} = -\frac{d\Omega(p^t(p^l), s_i)}{dp^l}\frac{dp^t}{dp^l}\pi(p^t(p^l; s_i)) \tag{10}$$

The left-hand side of Equation 10 is an expression of "marginal benefit" of price adjustment, i.e., the marginal change in the seller's payoff ($\pi'(p^t)\frac{dp^t}{dp^l}$) times the probability that a matched buyer accepts the TIOLI price ($\Omega(p^t(p^l), s_i)$). The right-hand side is an expression of the "marginal cost" of price adjustment, i.e., the marginal effect of the changes in listing price on the probability of TIOLI price's acceptance ($\frac{d\Omega(p^t(p^l),s_i)}{dp^l}\frac{dp^t}{dp^l} < 0$) times the payoff ($\pi(p^t(p^l; s_i))$). As for the second order conditions, we have made assumptions about the functional forms of $\pi()$ and $\Omega$ such that we can show that the "marginal benefit" from Equation 10 is decreasing and "marginal cost" increasing.

Similarly, taking the first-order condition of Equation 7 with respect to $a$ and rearranging gives the following expression that identifies the optimal $a$ is conditional on a choice of $p^l$.

$$\frac{d\gamma}{da}\pi(p^t(p^l; s_i))\Omega(p^t(p^l), s_i) = k'(a) \tag{11}$$

A component of the left-hand side of Equation 11 is the marginal gain from advertising, which is a product of changes in the Poisson match rate ($\frac{d\gamma}{da}$) and expected payoff ($\pi(p^t(p^l; s_i))\Omega(p^t(p^l), s_i)$). This marginal gain is equal to the right-hand side term $k'(a)$, i.e., the marginal cost of advertising. As for the second-order condition, we assume the functional forms of the Poisson matching function $\gamma()$ and the cost function $k()$ such that a unique solution of $a$ exists.[7]

## C.4 Information friction and beliefs

The solutions above hinge on the assumption that sellers have accurate beliefs about buyers' WTP, other parameters, and functional forms (e.g., Poisson match rate function). However, if there is noise in sellers' beliefs about buyers' WTP, how would it affect sellers' decisions? We explore this possibility while assuming that beliefs on other parameters and functional forms are accurate.

Suppose that seller $i$ possesses noisy information about the distribution of buyers' WTP. We assume that their beliefs are accurate on average over all sellers to focus on a point about noise rather than bias. Individual sellers hold beliefs over $f(\theta)$, and the distribution of buyers they get matched to conditional on $p^l$(i.e., $g(\theta)$) also depends on their belief over $f(\theta)$. We denote seller $i$'s belief on $f$ as $\hat{f}(\theta|\mathcal{I}_i)$ and their resulting belief over $g$ as $\hat{g}(\theta_0|\mathcal{I}_i)$, where $\mathcal{I}_i$ denotes information quality individuals possess to form a prior belief. The resulting optimality conditions then simply replace $f$ with $\hat{f}(\theta_0|\mathcal{I}_i)$ and $g$ with $\hat{g}(\theta_0|\mathcal{I}_i)$.

The idea behind our intervention is that a randomly selected subset of sellers would update their beliefs based on the information signals contained in the Price Calculator estimates. Signal $x_i$ is drawn from the true distribution of the WTP, $f$. If treated sellers engage in a rational Bayesian updating process, their posterior beliefs $\hat{f}(|x_i, \mathcal{I}_i)$ and $\hat{g}(|x_i, \mathcal{I}_i)$ from their equivalents under no information friction. The schematic representation of Bayesian

[7] It is likely reasonable to assume that the Poisson match rate function $\gamma()$ is concave given the diminishing returns to advertising. The potential issue is with the cost function $k()$, including the financial cost of advertising. PakWheels offers quantity discounts of advertising credits, making the per-unit cost of advertisement use cheaper as sellers use more. We will check with data to see if the use of advertising tools in excess (e.g., bumping their ads at a high frequency) is a concern.

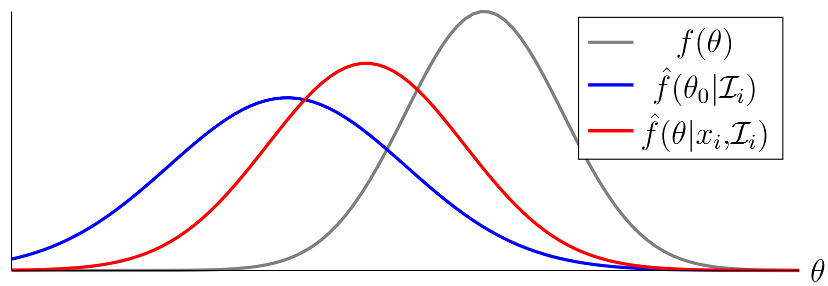



Figure C.2: Beliefs about$f$ based on information set $\mathcal{I}_i$ and signal $x_i$

belief updating is shown in Figure C.2.

### C.4.1 Bayesian belief updating

We assume that sellers engage in a Bayesian belief-updating process when they receive information signals in the form of Price Calculator estimates. We note that, in reality, some sellers may not be Bayesian and exhibit behavioral deviations (e.g., motivated beliefs). We stay away from such complications and focus on a rational framework, which we believe is more relevant to the main treatment effects we expect to see. Furthermore, we may expect sellers to have heterogeneous strategic responses to the information signal. Formalizing the belief updating process thus allows us to separate the strategic responses (expressed in the functional form of $\pi()$) from the changes in beliefs (parameters of $\hat{f}()$, which we elicit in the endline survey.

We assume that both buyers' WTP and sellers' prior beliefs about it are normally distributed. We make this assumption to simplify the distributional forms of prior and posterior beliefs, as the normal distribution is its own conjugate prior. We also note that the sellers' prior beliefs are based on information they already have access to, i.e., $\mathcal{I}_i$. We express the prior beliefs and true distributions as follows:

- Prior belief about demand distribution: $\hat{f}(\theta_0|\mathcal{I}_i) \sim \; N(\mu_{i,0}, \sigma_0^2)$

- True demand distribution: $f(x) \sim \; N(\mu, \sigma^2)$

Signals that sellers receive are drawn from the true demand distribution $x$. If sellers are Bayesian, they will update $\theta$ based on $x$. Both the prior belief as well as the signals are continuous, so the posterior belief function is as follows:

$$\hat{f}(\theta|x_i, \mathcal{I}_i) \sim \; N(\frac{a\mu_0 + bx}{a+b}, \frac{1}{a+b}), \tag{12}$$

where $a = \frac{1}{\sigma_0^2}$, and $b = \frac{1}{\hat{\sigma}^2}$. We assume that sellers have *perceptions* about the quality of information signals they receive, whether that is the variance of $f$ and/or the standard error of the information signal we deliver in practice. We therefore use $\hat{\sigma}^2$ instead of $\sigma^2$ to include an individual's perception about the credibility, or variance, of the information signal. Furthermore, we could have specified that $\hat{\sigma}^2$ is a function of some argument (e.g., the difference between data and prior mean: $\sigma^2 = \phi(|x - \mu_{i,0}|)$). Instead, we stay agnostic about factors that correlate with $\hat{\sigma}^2$ and leave this as an empirical exercise after estimating $\hat{\sigma}^2$.

## C.5 Model predictions: direct treatment effects

### C.5.1 Information intervention reduces deviation of $p^l$ from $p^{l*}$)

The optimality condition for $p^l$ in Equation 10, under noisy beliefs, can be rearranged as follows.

$$\frac{\pi'(p^t)}{\pi(p^t(p^l; s_i))} = -\frac{\frac{d\Omega(p^t(p^l; \hat{f}(\theta_0|\mathcal{I}_i), s_i)}{dp^l}}{\Omega(p^t(p^l; \hat{f}(\theta_0|\mathcal{I}_i), s_i))} \tag{13}$$

Following the logic from Equation 10, Equation 13 shows that the seller sets their listing price $p^l$ such that their *beliefs* about the expected payoff equals their *beliefs* about the cost. Their choice of $p^l$ based on their belief about $\hat{f}()$, however, does not necessarily equal that based on true $f()$. In other words, it is generally true that given a choice of $p^l$ made under information friction (with access only to $\mathcal{I}_i$) and prior belief $\hat{f}()$:

$$\frac{\frac{d\Omega(p^t(p^l;\hat{f}(\theta_0|\mathcal{I}_i),s_i)}{dp^l}}{\Omega(p^t(p^l;\hat{f}(\theta_0|\mathcal{I}_i),s_i))} \neq \frac{\frac{d\Omega(p^t(p^l;f(\theta),s_i)}{dp^l}}{\Omega(p^t(p^l;f(\theta),s_i))} \tag{14}$$

We make assumptions about the structure of the belief-updating process in Section C.4.1 to show how the Price Calculator estimate could help sellers update beliefs about the demand distribution $\hat{f}()$, on average toward the truth $f()$. Given our assumption that the form of the objective function with respect to the choice variable $p^l$ is quasiconcave, we make the following prediction:

- Prediction 1.: Information intervention brings $p^l$ closer to what it would be under no information friction about the demand distribution (call this $p^{l*}$), if the updated belief brings the posterior distribution $\hat{f}(\theta|x_i,\mathcal{I}_i)$ closer to $f(\theta)$ from $\hat{f}(\theta|\mathcal{I}_i)$. (Research question 1.1.)

### C.5.2 Information intervention increases *ex post* payoffs

If information friction results in beliefs about $f()$ and the objective function are quasiconcave with respect to $p^l$, then the choice of $p^l$ under information friction is *ex-post* suboptimal. It follows that the Price Calculator information signal would increase the *ex-post* payoff, as posterior beliefs about $f()$ are more accurate and would result in $p^l$ closer to $p^{l*}$ on average. In other words, we can show that:

$$\pi(p^t(p^l;\hat{f}(\theta|x_i,\mathcal{I}_i),s_i))\Omega(p^t(p^l;\hat{f}(\theta|x_i,\mathcal{I}_i),s_i)) \geq \pi(p^t(p^l;\hat{f}(\theta_0|\mathcal{I}_i),s_i))\Omega(p^t(p^l;\hat{f}(\theta_0|\mathcal{I}_i),s_i)) \tag{15}$$

This leads to the next prediction of our conceptual framework:

- Prediction 2.: Information intervention increases sellers' *ex-post* returns from the platform if the updated belief brings the posterior distribution $\hat{f}(\theta|x_i,\mathcal{I}_i)$ closer to $f(\theta)$ from $\hat{f}(\theta|\mathcal{I}_i)$. (Research question 1.2.)

### C.5.3 Information intervention may increase $a$

We have so far shown that the choice of listing price can be affected by noise in sellers' beliefs about the demand and that *if* the Price Calculator estimate leads to an updated belief that is closer to the truth, then it would bring the listing price toward the optimum and improve their payoff from engaging with the marketplace. How would their choice of advertising then be affected by information friction? From Equation 11, we see that under no information friction, sellers use advertising up to the point where the expected marginal benefit of its use equals its marginal cost. Under information friction, however, sellers consume advertising tools to the point where their *beliefs* about the expected marginal benefit equals marginal cost. The following equation makes this point by modifying Equation 11, and putting the term corresponding to beliefs about expected payoffs in a (very) wide hat:

$$\frac{d\gamma(a;s_i,\mathcal{I}_i)}{da}\widehat{\pi(p^t(p^l;\hat{f}(\theta_0|\mathcal{I}_i),s_i))\Omega(p^t(p^l;\hat{f}(\theta_0|\mathcal{I}_i),s_i))} = k'(a;s_i,\mathcal{I}_i) \tag{16}$$

The information intervention improves sellers' *ex-post* payoffs (Equation 15). The Price Calculator intervention may also shift sellers' expectations, i.e., they themselves believe that their *ex-post* payoffs would improve, meaning:

$$\widetilde{\pi(p^t(p^l;\hat{f}(\theta|x_i,\mathcal{I}_i),s_i))\Omega(p^t(p^l;\hat{f}(\theta|x_i,\mathcal{I}_i),s_i))} \geq \widetilde{\pi(p^t(p^l;\hat{f}(\theta_0|\mathcal{I}_i),s_i))\Omega(p^t(p^l;\hat{f}(\theta_0|\mathcal{I}_i),s_i))} \tag{17}$$

If Equation 17 is true, then, combined with Equation 16 we see that

$$\frac{d\gamma(a;s_i,\mathcal{I}_i)}{da}\widetilde{\pi(p^t(p^l;\hat{f}(\theta|x_i,\mathcal{I}_i),s_i))\Omega(p^t(p^l;\hat{f}(\theta|x_i,\mathcal{I}_i),s_i))} \geq k'(a;s_i,\mathcal{I}_i) \tag{18}$$

Then, it follows that $a^*(s_i,x_i,\mathcal{I}_i) \geq a(s_i,\mathcal{I}_i)$. In other words:

- Prediction 3.: Information intervention increases $a$ if sellers' expectations about their *ex-post* returns from the platform are updated upward when they receive the price information signal.(Research question 1.3.)

## C.6 Model predictions: spillovers and their mechanisms

The optimality conditions and predictions above are based on the assumption that exogenous information shocks via the experiment only affect individual choices. We also hypothesize that individuals' access to information and their choices may generate spillovers, happening through multiple mechanisms. In this sub-section, we discuss three possibilities: information spillovers, distribution of buyer attention, and congestion.

### C.6.1 Information spillovers

The first possibility is that sellers' choices of $p^l$ may generate changes to the quality of information signals available in the market, therefore affecting $\mathcal{I}_i$ for all seller $i$ in the market segment. This point is captured in research question 2.1.1.. An exogenous shift in the information set available in a market segment would affect sellers' prior beliefs about the distribution of buyers' WTP. The choices of $p^l$ and $a$ made in a treated market segment would therefore be closer to those under no information friction than in an untreated market segment.

In other words, suppose that part of a market segment is exposed to the Price Calculator treatment, and their choices of $p^l$ are closer to the values they would choose under no information friction. Define the resulting information set in this market segment to be a union of the existing information set and information contained in treated $p^l$'s, i.e., $\mathcal{J}_i \equiv \mathcal{I}_i \cup I(\bigcup_{i \in T} p_i^l)$, where $I$ is a function that maps a set of prices and $T$ is a set of treated individuals. Then we get:

$$\pi(p^t(p^l;\hat{f}(\theta|\mathcal{J}_i),s_i))\Omega(p^t(p^l;\hat{f}(\theta|\mathcal{J}_i),s_i)) \geq \pi(p^t(p^l;\hat{f}(\theta_0|\mathcal{I}_i),s_i))\Omega(p^t(p^l;\hat{f}(\theta_0|\mathcal{I}_i),s_i)) \quad (19)$$

- Prediction 4.: Information spillovers from treated individual sellers in a given market segment would weakly improve the information set of all sellers in the market segment and would bring the prior beliefs about $f()$ and $p^l$ closer to those under no information friction, increase *ex-post* returns, and would increase $a$ if expected returns increase. (Research question 2.1.1.)

### C.6.2 Congestion and the match rate

Lastly, the treatment may affect quantities of sellers and buyers actively participating in the market, affecting congestion and the speed at which sellers and buyers are matched. The match rate is expressed via a function $\gamma()$. The spillover effect of changes in congestion levels may depend on what types of sellers and buyers are taken out of the market as a result of the Price Calculator intervention. Complex assumptions about the resulting composition of sellers and buyers are outside the scope of this framework.

One simple scenario we explore is what would happen if the intervention relaxes congestion in the matching process overall and the match rate increases for all sellers. In other words, treated market segments would have $\tilde{\gamma}(a) \geq \gamma(a)$, $\forall a$. Then we get:

$$a^*|_{\tilde{\gamma}} \geq a^*|_{\gamma} \tag{20}$$

This is because the marginal benefit of advertising is now higher in treated market segments for a given $a$, while the cost function is unchanged. This would lead to further consumption of $a$ to the point where marginal cost equals the benefit. This does not affect the choice of $p^l$, as it is a separate problem from the choice of $a$.

- Prediction 5.: A higher match rate as a result of reduced congestion in treated market segments results in higher consumption of advertising tools than in untreated market segments. (Research question 2.1.3.)

# D Pre-specified research questions

We pre-specified the research questions in the pre-analysis plan, which was initially submitted and modified both prior to the beginning of the intervention. We divided our research questions into two; the first set of questions pertains to direct treatment effects of price information on listing prices, transaction outcomes, and mechanisms at the individual level. Our hypothesis is that the information intervention reduces noise in sellers' beliefs about the distribution of demand, affects their pricing decisions, and improves their market outcomes. We also posit that sellers do not make strategic choices beyond pricing, such as advertising, and those choices are contingent upon their pricing decisions and beliefs. We, therefore, hypothesize that contingent strategic choices like advertising could be affected by price information intervention.

The second set of questions concerns spillovers and other market-level impacts of the information intervention. Possible channels include a) diffusion of information itself via shifts in the distribution of listing prices, b) competing sellers' pricing and advertising choices to treated individuals' strategic choices, and c) reduction in search friction and congestion in the market. Our empirical objectives, therefore, are to identify spillover effects on our primary outcome variables and narrow down on channels of such spillovers.

Following is the list of primary (in bold) and secondary questions, with links to the theoretical predictions in Section 6.

1. Does the price information intervention induce direct effects on pricing, advertising, and transaction outcomes?

    1.1. **Do sellers adjust their listing prices toward the price signal they receive?** (Prediction 1.)

        1.1.1. Does the intervention affect sellers' stated beliefs about the distribution of transaction prices?

1.2. **Does the price information intervention improve sellers' returns from the platform?** (Prediction 2.)

1.2.1. **Does it increase page views?**

1.2.2. **Does it increase the transaction probability?**

1.2.3. **Does it affect the transaction price?**

1.3. **Do sellers respond to the intervention by making strategic adjustments in advertising?** (Prediction 3.)

1.4. Across what characteristics do we observe heterogeneous treatment effects?

- sellers' experience
- product heterogeneity in market clusters
- availability and variation of price information at baseline

2. Does the price information intervention create spillovers and other knock-on effects?

2.1. **Does the intervention induce spillovers in terms of listing prices, transaction outcomes, and the use of advertising?**

2.1.1. Are these spillovers induced by changes in listing prices and advertising by competing, treated sellers? (Prediction 4.)

2.1.1.1. Are there spillover effects on the stated belief about the distribution of transaction prices?

2.1.2. Do spillovers occur through a zero-sum shift in buyer attention toward treated sellers?

2.1.3. Do spillovers occur through changes in congestion? (Prediction 5.)

# E Statistical Power

We make choices on the following dimensions to maximize the statistical power of detecting treatment, spillover, and saturation effects:

- shares of clusters assigned to control, high treatment, and medium treatment groups
- share of posts into treatment assignment for both high- and medium groups.

We take as given the cluster sizes, as it depends on a fixed experimental duration of 8 weeks. We also take as given the number of clusters, as it depends on the number of models PakWheels could offer Price Calculator estimates without risking providing noisy information to infrequently traded vehicle models.

We take a hybrid approach based on theoretical optimal design and Monte Carlo simulations. For the latter, we use real historical data with assumptions about the reduced-form structure and relative effect sizes between direct treatment, spillovers, and saturation. First, we set the share of control clusters to 0.5, and the rest split evenly between high- and medium-treatment groups, based on insight from Baird et al. (2018). Their setup and assumptions are similar to ours, such as that they allow for intracluster correlation and only partial interference (i.e., within clusters but not across). We deviate from the procedure by Baird et al. (2018) on our choices of saturation levels. We assign second-stage randomization based on the last digit of the sellers' user ID, and we expect some level of treatment non-compliance as discussed in Section 3.3. As such, we have chosen the high treatment assignment to be 9 out of 10 digits and middle treatment 5 out of 10. With a conservative assumption on treatment take-up of about 70 percent, then treatment intensities would be symmetrical around 0.5, as recommended by Baird et al. (2018).

Based on the saturation levels and the range of control group size chosen by the process above, we run Monte Carlo simulations to estimate power under several assumptions. We use actual data from PakWheels and estimate the statistical power of detecting a range of effect sizes for direct impact, spillovers, and saturation. We use different data samples and specifications for direct and spillover effects, as described in Section 4.1. We bootstrap-sample the data 100 times, stratified over the make-model. We then assign treatment according to the method described in Section 3.2, and construct outcome variables conditional on cluster and individual assignments into treatment. We assume that direct and spillover treatment effects are linear and additive, except for the transaction outcome. [8] Spillovers are assumed to occur within the make-model cluster evenly for both treated and untreated posts. Using real historical data, we assume that intra-cluster correlation is already built in. We assume no inter-cluster interference.

The outcome variables, which are standardized and identical to the primary outcomes described in Section 4, are the following:

- log(absolute difference between listing price and Price Calculator estimate)
- 1 if reported as sold
- log(self-reported transaction price)
- advertising index[9]
- buyer-attention index.

We estimate the power of detecting the intend-to-treat (ITT) effects of direct treatment and spillovers for a range of relatively small effect sizes (0.025 to 0.2 standard deviation). We explore two scenarios of spillover and saturation effect sizes relative to direct treatment effects:

[8]Given that the transaction outcome is binary, we assumed that assignment into treatment would increase the probability of transaction by X%, where X is a standardized effect size based on the standard deviation of the binary variable.

[9]There is a minor difference in definitions of constituent variables due to limitations in the pre-intervention data

1. spillover effect in high-saturation is 50% of the direct treatment effect, and in medium saturation 25%.

2. spillover effect in high-saturation is 100% of the direct treatment effect, and in medium saturation 50%.

We identify the optimal division of clusters into treatment arms based on the following proposition by Baird et al. (2018):

$$\psi^* = \frac{-\kappa + \sqrt{\kappa^2 + (1-\rho)\kappa}}{1-\rho} \tag{21}$$

where $\psi$ is share of control clusters, $\kappa \equiv 1 + (n-1)$, $\rho$ the intracluster correlation, and $n$ cluster size. The boundary values of $\psi^*$ are $\sqrt{2}-1$ and 0.5. Plugging in our parameter values to Equation 21 resulted in a control share close to 0.5.

We use the identical estimating equations to estimate intent-to-treat effects as in the main analysis in Section 4.2. In other words, we run the *logit* model for the binary outcome and linear regressions for all other outcomes. These models include the same set of controls as the ones used for the primary analysis. We use data from an 8-week period that approximates the actual experimental timing. We also present results that include data from 8 weeks prior in addition to data from the experimental period. This is to gauge how much power gains we could make in detecting spillovers by a larger sample and with cluster fixed effects, as described in Section **??**. In both approaches, we report the false-discovery-rate-adjusted q-values based on five p-values corresponding to the main outcomes. These adjustments are made separately for direct treatment, spillover, and high-saturation effects.

## E.1 Results of power calculations

The results of power simulations from specifications containing only data from the 8-week experimental period are shown in Figures E.3 and E.4, corresponding to scenarios 1. and 2., respectively. These figures reveal that the power to detect direct treatment effects of 0.05 SD is 80% or greater for all five primary outcomes. The effect size of 0.05 SD translates into 11,594 PKR (65.73 USD at 176.4 PKR to USD) in an absolute difference between the listing price and Price Calculator estimate (level mean: 305,434 PKR), 2.44 percentage-points in transaction probability (mean: 0.394), and 55,473 PKR (314.47 USD) in transaction price (level mean: 1,893,626 PKR).

Figures E.3 and E.4 also show that we are able to detect some spillover and saturation effects at 80% power or greater with the specification for primary analysis, depending on the effect sizes and assumptions about their relative sizes to direct effects. Figure E.3 suggests that under assumption 1. we would have greater than 80% power to detect a spillover effect of 0.05 SD on advertisement, as well as saturation effects of 0.1 SD on advertisement and demand. Figure E.4 suggests that under assumption 2., we would have greater than 80% power to detect spillover effects of 0.1 SD on transaction, demand, and advertisement, and saturation effect of 0.2 SD on all outcomes. Figures E.5 and E.6 also show that using the two-way fixed-effect specification from secondary analysis in Section **??** would improve power on some of the spillover outcomes, as compared to Figures E.3 and E.4, respectively.

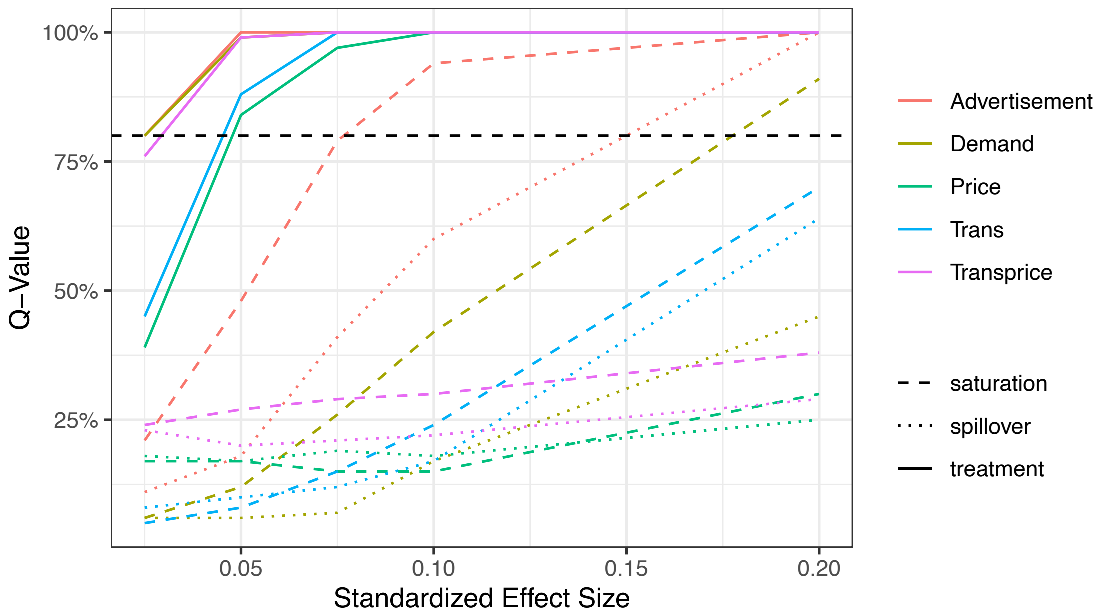



Figure E.3: Power estimates: Scenario 1. and data over 8 weeks

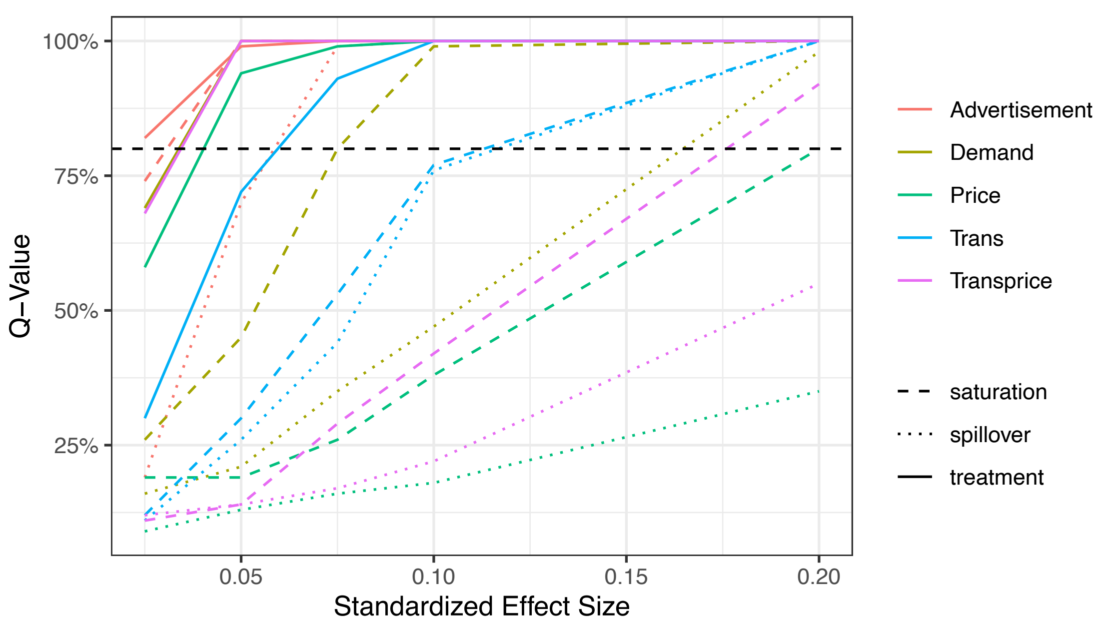



Figure E.4: Power estimates: Scenario 2. and data over 8 weeks

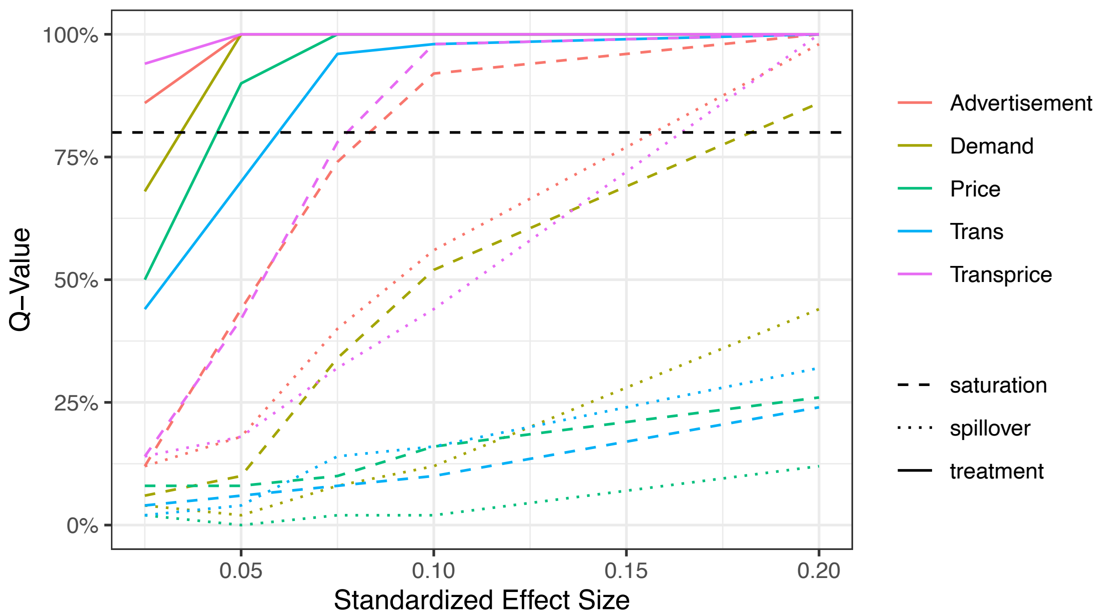



Figure E.5: Power estimates: Scenario 1. and data over 16 weeks

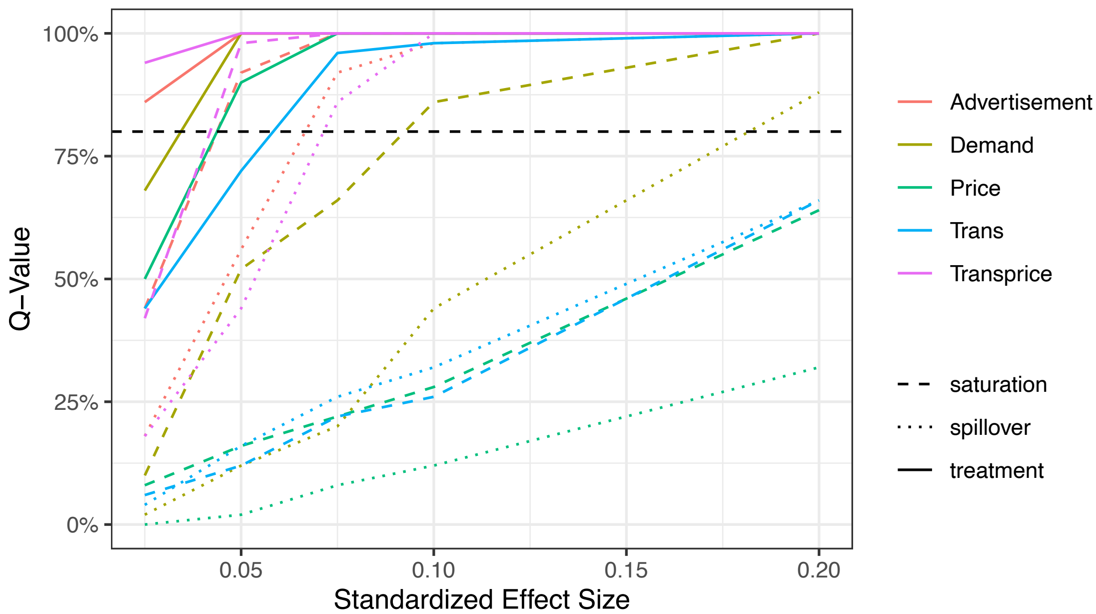



Figure E.6: Power estimates: Scenario 2. and data over 16 weeks

# F Additional Tables

Table F.1: **ITT estimates on price-related outcomes**

| | log(Listing price) | Price updated | N. price updates | log(Abs. price change) |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | OLS | Logit | OLS | OLS |
| Assignment | -0.0005 | -0.0144 | 0.0092 | -0.0585 |
| | (0.0014) | (0.0194) | (0.0096) | (0.0511) |
| Spillover | -0.0211 | -0.0230 | -0.0401** | -0.0451 |
| | (0.0256) | (0.0222) | (0.0166) | (0.0505) |
| Spillover (high) | -0.0034 | 0.0587*** | 0.0361* | 0.1356*** |
| | (0.0418) | (0.0226) | (0.0194) | (0.0488) |
| Observations | 117,891 | 117,891 | 117,891 | 117,891 |
| Squared Correlation | 0.93107 | 0.02790 | 0.03762 | 0.03023 |
| Pseudo $R^2$ | 1.2508 | 0.02253 | 0.01069 | 0.00500 |
| BIC | -61,139.6 | 144,326.4 | 420,509.3 | 721,911.2 |

*Notes:* OLS estimates. "log(Listing price)": log of the final listing price. "Price updated": a binary outcome that is 1 if the seller ever updates the listing price. "N. price updates": number of times the seller updates the listing price. "log(Abs. price change)": log of the absolute difference between the initial and final listing price. The specification is as shown in Equation 1. Standard errors are reported in parentheses and clustered at the make-model level. The stars show the two-tailed significance in p-values: p<0.1*; p<0.05**; p<0.01***.

Table F.2: **ITT estimates on transaction-related outcomes**

| | 1 if sold | log(Transaction price) | log(Seller revenue) |
|---|---|---|---|
| | (1) | (2) | (3) |
| Assignment | -0.0110*** | -0.0008 | -0.0457 |
| | (0.0038) | (0.0037) | (0.0363) |
| Spillover | 0.0105*** | -0.0401 | 0.0185 |
| | (0.0030) | (0.0335) | (0.0407) |
| Spillover (high) | -0.0033 | -0.0031 | 0.0307 |
| | (0.0071) | (0.0457) | (0.0367) |
| Observations | 111,312 | 14,084 | 117,891 |
| $R^2$ | 0.01478 | 0.92874 | 0.01112 |
| Within $R^2$ | 0.00465 | 0.75562 | 0.00343 |

*Notes:* OLS estimates. "1 if sold": a binary outcome that is 1 if the seller reports the car as sold. "log(Transaction price)": log of the self-reported transaction price. "log(Seller revenue)": log of the self-reported transaction price if sold, and 0 otherwise. The specification is as shown in Equation 1. Standard errors are reported in parentheses and clustered at the make-model level. The stars show the two-tailed significance in p-values: p<0.1*; p<0.05**; p<0.01***.

Table F.3: **ITT estimates on variables included in the page-view index outcome**

| | Page views<br>(1) | Phone number views<br>(2) | Page-view index (not winsorized)<br>(3) | Page-view index<br>(4) |
|---|---|---|---|---|
| Assignment | -59.91*** | -0.4050** | -0.0335*** | -0.0172*** |
| | (12.37) | (0.2041) | (0.0070) | (0.0039) |
| Spillover | 57.39*** | 1.151*** | 0.0499*** | 0.0332*** |
| | (18.23) | (0.3713) | (0.0145) | (0.0106) |
| Spillover (high) | 8.431 | -0.3183 | -0.0054 | -0.0092 |
| | (25.02) | (0.7047) | (0.0245) | (0.0169) |
| | | | | |
| Observations | 117,891 | 117,891 | 117,891 | 117,891 |
| $R^2$ | 0.10683 | 0.06462 | 0.09405 | 0.12322 |
| Within $R^2$ | 0.01673 | 0.00938 | 0.01082 | 0.01652 |

*Notes:* OLS estimates. "Page views': The number of times the individual listing page is viewed. "Phone number views": The number of times the button in the individual listing page that shows the seller's phone number is clicked. "Page-view index (not winsorized)": The standardized index consisting of page views and phone number views, non-winsorized. "Page-view index": The standardized index consisting of page views and phone number views, winsorized. The specification is as shown in Equation 1. Standard errors are reported in parentheses and clustered at the make-model level. The stars show the two-tailed significance in p-values: p<0.1*; p<0.05**; p<0.01***.

Table F.4: **ITT estimates on variables included in the advertising index outcome**

| | N. bumps<br>(1)<br>OLS | 1 if featured<br>(2)<br>Logit | 1 if certified<br>(3)<br>Logit | Advertising index (not winsorized)<br>(4)<br>OLS | Advertising index<br>(5)<br>OLS |
|---|---|---|---|---|---|
| Assignment | -0.0456*** | -0.0948** | -1.182*** | -0.0578*** | -0.0095*** |
| | (0.0153) | (0.0398) | (0.1601) | (0.0108) | (0.0025) |
| Spillover | 0.0146 | -0.0525 | 0.0984 | 0.0165 | -0.0005 |
| | (0.0159) | (0.0488) | (0.0933) | (0.0108) | (0.0042) |
| Spillover (high) | 0.0472*** | 0.0638 | 0.7654*** | 0.0415*** | 0.0062 |
| | (0.0174) | (0.0618) | (0.1381) | (0.0136) | (0.0059) |
| | | | | | |
| Observations | 117,891 | 116,346 | 91,159 | 117,891 | 117,891 |
| Squared Correlation | 0.07426 | 0.29549 | 0.11528 | 0.21337 | 0.29329 |
| Pseudo $R^2$ | 0.02057 | 0.29896 | 0.26361 | 0.08105 | 0.24067 |
| BIC | 435,153.9 | 50,519.6 | 7,744.6 | 322,880.4 | 131,198.5 |

*Notes:* OLS estimates. "N. bumps": the number of times the seller applies the "bump" tool on the listing. "1 if featured": a binary outcome that is 1 if the seller ever features the listing. "1 if certified": a binary outcome that is 1 if the car in the listing is certified. "Advertising index (not winsorized)": The standardized index consisting of N. bumps, 1 if featured, and 1 if certified, non-winsorized. "Advertising index": The standardized index consisting of N. bumps, 1 if featured, and 1 if certified, winsorized. The specification is as shown in Equation 1. Standard errors are reported in parentheses and clustered at the make-model level. The stars show the two-tailed significance in p-values: p<0.1*; p<0.05**; p<0.01***.

# G Endline telephone survey questions

| | | | |
|---|---|---|---|
| begin group | section_1 | Background | |
| select_one yn_noad | s1_q1 | Our records show that you recently listed [make] [model] [model year] in [city location] on PakWheels. Have you already sold this vehicle you posted? | کیا آپ نے اپنی گاڑی بیچی ہے. ہمارے ریکارڈز کے مطابق آپ نے ${make_b}, ماڈل ${model_b} سال ${year_b} شہر ${city_b} کا اشتہار لگایا تھا . کیا آپ نے جس گاڑی کا اشتہار پاک ویلز پر لگایا تھا وہ بیچ دی ہے؟ |
| integer | s1_q2 | What was the price you sold this car at? We would like to remind you again that your answers will stay anonymous and be used for research purposes only | اگر گاڑی بیچ دی ہے تو اسکی قیمت کیا تھی. آپ نے کس قیمت پر گاڑی بیچی ؟<br>ہم آپکو دوبارہ یاد کروا دیں کے آپ کے جوابات کو گم نام رکھا جائے گا اور صرف تحقیقی مقاصد کے لیے استعمال کیا جائے گا. |
| integer | s1_q3 | What was the "expected" price?<br>At what price did you expect to sell this car at, when you initially posted it on PakWheels? | جب اپنے ابتدائی طور پر اِس گاڑی کا اشتہار پاک ویلز پر لگایا تھا تو آپ کو کیا تواقع تھی کے یہ گاڑی کتنے کی بک جائے گی ؟ |
| integer | s1_q4 | What is realistically the highest price you could have gotten for your car? | آپ کے خیال میں، آپ کی گاڑی زیادہ سے زیادہ کتنی قیمت پر بیچی جا سکتی تھی؟ |
| integer | s1_q4a | What is realistically the lowest price you could have gotten for your car? | آپ کے خیال میں، آپ کی گاڑی کم سے کم کس قیمت پر بیچی جا سکتی تھی؟ |
| integer | s1_q5 | How much did you pay for this vehicle when you first bought it? | آپ نے یہ گاڑی کتنی قیمت پر خریدی تھی؟ |
| select_multiple reason_sell | s1_q6 | Why have you not sold the car? (Allow the respondent to elaborate and ask follow-up questions to determine which of the following apply. You can choose more than one options.) | آپ نے گاڑی کیوں نہیں بیچی ؟ |
| text | s1_q6_o | Please specify other | دیگر کی وضاحت کریں |
| select_one reasons_who | s1_q7 | How and to whom did you sell the car? (Allow the respondent to elaborate and ask follow-up questions to determine which of the following apply.) | آپ نے یہ گاڑی کس کو اور کس طرح بیچی؟ |
| integer | s1_q8 | In the past 12 months, how many cars did you try to sell in total, not just on PakWheels? | پچھلے 12 مہینوں میں ، آپ نے مجموعی طور پر کتنی گاڑیاں فروخت کرنے کی کوشش کی؟ کل گاڑیاں ,صرف وہ نہیں جن کا اشتیار آپ نے پاک ویلز پہ لگایا ہو |
| end group | section_1 | | |
| begin group | section_2 | Price Calculator | |
| calculate | treat_2nd | | کچھ لوگوں کو گاڑی کا اشتہار پاک ویلز کی ویب سائٹ پر بناتے وقت ایک پوپ اپ یا گرافک باکس دکھایا گیا تھا ، جس میں نئے پرائس کیلکو لیٹر کی مدد سے |
| select_one yesno_dk | s2_q1 | When creating the post for a car (in the "Post an Ad" process), a random subset of people were shown a pop-up or graphic box containing price estimate from the new Price Calculator, along with higher and lower end estimates. If you were selected, you would have seen this when you first selected your listing price while creating the post. Do you remember seeing this particular Price Calculator estimate? | کچھ لوگوں کو گاڑی کا اشتہار پاک ویلز کی ویب سائٹ پر بناتے وقت ایک پوپ اپ یا گرافک باکس دکھایا گیا تھا ، جس میں نئے پرائس کیلکو لیٹر کی مدد سے گاڑی کی اندازاً قیمت اور گاڑی کی کم سے کم اور زیادہ سے زیادہ اندازاً قیمت دی گئی تھی۔ اگر آپ ان کچھ لوگوں میں شامل ہیں توآپ نے یہ پرائس کیلکولیٹر سب سے پہلے تب دیکھا ہوگا جب آپ پوسٹ لکھتے وقت لسٹ پرئس سیلیکٹ کر رہے ہوں گے۔ کیا آپ کو یہ پرائس کیلکو لیٹر کی مدد سے لگائی گئی گاڑی کی یہ اندازاً قیمت یاد ہے؟ |
| integer | s2_q2 | What was the estimate you were given? | اگر آپکو پرائس کیلکو لیٹر دیکھنا یاد ہے ، آپکو گاڑی کی اندازاً قیمت کیا دی گئی تھی ؟ |
| select_one enum_note | s2_q2_e | Did the respondent give you estimates from the Price Calculator provided to them on the sell-form (the "Post an Ad" process) from the intervention? Or did they give you something else? | کیا جواب دھندہ نے پرائس کیلکو لیٹر پوسٹ این ایڈ پروسیس والے کے مطابق جواب دیا یا کچھ اور جواب دیا ؟ |
| select_one too_hl | s2_q3 | Did you think that this Price Calculator box during the "Post an Ad" process gave you a reasonable estimate of transaction price? Or was it too low, or high? | آپ کے خیال سے پرائس کیلکو لیٹر پوسٹ این ایڈ پروسیس والے کے ذریعے جو گاڑی کی قیمت پاک ویلز کی جانب سے بتائی گئی وہ مناسب تھی یا بہت زیادہ تھی یا بہت کم تھی ؟ |
| select_one yesno_dk | s2_q5 | Do you know of any other sellers who have gotten the Price Calculator estimates from PakWheels? | کیا آپ کوئی اور گاڑی فروخت کرنے والوں کو جانتے ہیں جن کو پاک ویلز استعمال کرتے وقت پرائس کیلکو لیٹر کی مدد سے لگائی گئی اندازاً قیمت ملی ہو ؟ |
| end group | section_2 | | |
| begin group | section3 | | |

| | | | |
|---|---|---|---|
| select_one yesno_dk | s3_q1 | Now I would like to ask you a few questions about searching for similar posts and the choice of listing price. Did you searched for, or looked at other posts that are similar to your vehicle on PakWheels? | اب میں آپ سے کچھ سوال آپ کی گاڑی کی پوسٹ سے ملتی جُلتی پوسٹ اور لسٹ پرائس کے بارے میں پوچھنا چاہتا ہوں. کیا آپ نے اپنی گاڑی سے ملتے جلتے دوسرے اشتھار پاک ویلز پر سرچ کیے ؟ |
| select_one search | s3_q2 | Did you search for posts only for ${model_b}, or did you also search for other models? | کیا آپ نے صرف ${model_b} پوسٹس کی تلاش کی یا دوسرے ماڈلوں کی بھی تلاش کی؟ |
| text | s3_q2_o | Please specify which models | آپ نے کون سے ماڈل کی سرچ کی ؟ |
| select_multiple search_m | s3_q3 | Did you restrict your search by any other terms? For example, your model year, version, or your city? | آپ نے اپنی سرچ کو درج زیل میں سے کن چیزوں پر محدود کیا ؟ جیسے کے ماڈل کا سال ، ماڈل ورژن یا آپکا شہر وغیرہ |
| text | s3_q3_o | Please specify other | دیگر کی وضاحت کریں |
| select_multiple search_op | s3_q4 | Besides other listings from PakWheels, what other information or experience did you base your initial listing price on? | پاک ویلز پر دوسری پوسٹس دیکھنے کے علاوہ ، اور کون سی معلومات یا تجربہ کی بنیاد پر آپ نے اپنی گاڑی کی ابتدائی قیمت کو مقرر کیا؟ |
| text | s3_q4_other | Please explain this | پاک ویلز کی دی ہوئی قیمت کے علاوہ کوئی دیگر معلومات کی وضاحت کریں |
| text | s3_q4_o | Please specify other | دیگر کی وضاحت کریں |
| select_one yesno_dk | s3_q5 | Have you changed your listing price on PakWheels after you have set it initially | کیا آپ نے اپنی گاڑی کی جو شروع میں قیمت پاک ویلز پر مقرر کی تھی اُسے بعد میں تبدیل کیا. |
| select_multiple price_adj | s3_q6 | Could you tell us why you did so? | آپ نے اپنی گاڑی کی ایک قیمت مقرر کی تھی، لیکن بعد میں اِس قیمت کو تبدیل کر دیا۔ کیا آپ ایسا کرنے کی وجہ بیان کر سکتے ہیں؟ |
| text | s3_q6_o | Please specify other | دیگر کی وضاحت کریں |
| end group | section3 | | |
| begin group | section4 | | |
| integer | s4_q1 | How much (in PKR) were you willing to bargain from the listing price you chose when you created the listing? | آپ اپنے اشتہار میں مقرر کردہ قیمت میں کتنی کم و بیشی کرنے پر رضامند تھے؟ |
| begin group | s4_preamble | Please tell us if you agre with the following statements | مندرجہ ذیل بیانات سے آپ کتنا متفق یا غیرمتفق ہیں؟ بلکل متفق سے بلکل غیر متفق کے پیمانے پر جواب دیں. |
| select_one likert_agreedk | s4_q2 | It was difficult to get enough inquiries for your post on PakWheels | آپنی پوسٹ سے متعلق مطلوبہ انکوائریز پاک ویلز پر حاصل کرنا مشکل تھا |
| select_one likert_agreedk | s4_q3 | It was difficult to get a price offer that you would accept for your car on PakWheels. | آپنی پوسٹ سے متعلق، قابل قبول قیمت پاک ویلز پر حاصل کرنا مشکل تھا |
| select_one likert_agreedk | s4_q5 | Most (around 3 out of 4 or more) of potential buyers of ${make_b} ${model_b} have good information about what are fair used car prices. | گاڑی ${model_b}$ {make_b} خریدنے والے زیادہ تر (چار میں سے تِین ) افراد کو مناسب قیمتوں کا اندازہ ہوتا ہے۔ |
| select_one likert_agreedk | s4_q6 | Most (around 3 out of 4 or more) of other sellers of ${make_b} ${model_b} have good information about what are fair used car prices. | گاڑی ${model_b}$ {make_b} بیچنے والے زیادہ تر (چار میں سے تِین ) افراد کو مناسب قیمتوں کا اندازہ ہوتا ہے۔ |
| end group | s4_preamble | | |
| begin group | s4a_preamble | Please tell us if you agre with the following statements | مندرجہ ذیل بیانات سے آپ کتنا متفق یا غیرمتفق ہیں؟ بلکل متفق سے بلکل غیر متفق کے پیمانے پر جواب دیں. |
| select_one helpful_scale | s4_q7 | As you may know, sellers like you can feature your ad, use "bumps" to get your post to be more visible, or requested for vehicle inspections by PakWheels. In a scale of 1 to 5, 5 being the most useful, how useful are these features, bumps, and inspections to get people to buy your car at a higher price? | جیسا کہ آپ جانتے ہیں، جو لوگ آپ کی طرح پاک ویلز کی ویبسائٹ پر اپنی گاڑی کو بیجنے کےلئے اشتہارلگاتے ہیں وہ اپنے اشتہار کو فیچر کر سکتے ہیں یا مشہور بنانے کے لئے "بمپس" کا استعمال کرسکتے ہیں یا پاک ویلز کی طرف سے اپنی گاڑی کی انسپیکشن کرواسکتے ہیں۔ ایک سے پانچ کے پیمانے پر بتاین کے ان سہولیات کا استعمال آپ کو اپنی گاڑی کو زیادہ قیمت پر بیچنے میں کتنا مدد گار ثابت ہو سکتا ہے ؟ ایک کا مطلب ہے بلکل مدد کار نہیں اور پانچ کا مطلب ہے بہت زیادہ مدد کار |
| select_one helpful_scale | s4_q8 | Again in a scale of 1 to 5, how useful are these features, bumps and inspections to increase the chance that it sells, or sells faster? | ایک سے پانچ کے پیمانے پر بتاین کے آپ کی گاڑی کے بکنے کے امکانات میں اضافہ کرنے میں بمپز اور فیچر کا استعال کتنا مدد گار ثابت ہو سکتا ہے ؟ ایک کا مطلب ہے بلکل مدد کار نہیں اور پانچ کا مطلب ہے بہت زیادہ مدد کار |
| select_one yesno_dk | s4_q9 | The Price Calculator is currently provided for free. But in the future if it were offered for 100 rupees per post, would you be willing to pay for it? | فل حال پرائس کیلکولیٹر فری ہے، لیکن اگر مستقبل میں اس کی قیمت ایک سو روپے فی پوسٹ مقرر کردی جائے تو کیا آپ اس کے استعمال کے لئے پیسے دیں گیں؟ |
| integer | s4_q10 | What is the maximum amount (PKR) that you would be willing to pay for the Price Calculator, per post? | پرائس کیلکولیٹر کے استعال کے لئے آپ زیادہ سے زیادہ کتنی قیمت فی پوسٹ ادا کرنا چاہتے ہیں؟ |
| end group | s4a_preamble | | |
| end group | section4 | | |

# Chapter 3: Belief formation, signal quality and information sources—Experimental evidence on air quality from Pakistan

Matthew Gibson* Isra Imtiaz† Shotaro Nakamura‡ Sanval Nasim§

Arman Rezaee¶

August 11, 2023

**Abstract**

Government and private services often compete in public services in developing countries, including environmental information. The efficacy of private alternatives in services traditionally provided by the government may depend on citizens' preferences for providers and beliefs about service quality. We conduct a field experiment to study how the information source affects citizens' preferences and beliefs in the context of ambient air quality forecast services in Lahore, Pakistan. We provide day-ahead air pollution forecasts and make salient one of the information sources—the government vs. a private citizens group. We find that respondents have a high willingness to pay

*Department of Economics, Williams College, MA.
†Mahbub ul Haq Research Centre, Lahore University of Management Sciences, Lahore, Pakistan.
‡Department of Economics, University of California, Davis, CA; email: snnakamura@ucdavis.edu
§Department of Economics, Colby College, Waterville, ME
¶Department of Economics, University of California, Davis, CA

for the forecast service. We also find that they have a significantly higher revealed preference for the assigned source against the other. Respondents' beliefs about air pollution levels are not statistically different across treatment groups, but their belief about the government forecast error is 12% higher than for the private alternative. Our findings suggest that respondents have weak priors and malleable preferences for information sources yet expect lower service quality from the government.

# 1 Introduction

An emerging body of empirical work emphasizes the demand for effective mitigation measures against severe seasonal ambient air pollution in developing cities (e.g., Freeman et al. 2019; Ito and Zhang 2020). One such measure that could yield considerable public benefit is accessible and reliable air-quality information. Our previous experimental study in Lahore, Pakistan, revealed that citizens are willing to pay an average of PKR 91 per month for day-ahead air pollution forecasts (Ahmad et al. 2022). Access to reliable information may allow citizens to form accurate beliefs about air pollution and take mitigation measures. This is evidenced by an increased demand for protective masks in our study (Ahmad et al. 2022).

Governments in developing countries, however, often struggle to provide consistent and reliable air quality information.[1] Consequently, various stakeholders, including citizens' groups, international and bilateral agencies, and research institutions have begun providing air quality information in the absence of reliable local government services.[2] For example, in Lahore, a citizens' group called Pakistan Air Quality Initiative (PAQI) crowd-sources low-cost monitors across the city and publishes their readings on Twitter and a mobile app at no charge to the users.

Private alternatives to government services may improve citizens' access to air quality information. However, the efficacy of private alternatives may depend upon citizens' preferences for and beliefs about the accuracy of the information sources. First, there are signif-

[1] There are several reasons behind the unreliable government service provision. First, government agencies may suffer from resource and capacity constraints that prevent them from providing consistent information, as many of the high-quality air quality monitors require regular maintenance. In Lahore, data by the local environmental regulator (Punjab Environment Protection Department, or EPD) are only made public as reports on their website in PDF (URL: `https://epd.punjab.gov.pk`). Second, government agencies may also face perverse incentives to obscure the true extent of environmental degradation by omitting unfavorable information (e.g.Ghanem and Zhang 2014)

[2] One notable example is AirNow by the U.S. Department of State, which has installed monitors at U.S. embassies and consulates and provides readings via their website (URL: `https://www.airnow.gov/international/us-embassies-and-consulates/`) and via Twitter.

icant variations in the availability of air quality information and readings between sources, as shown in Figure 1. Second, we document heterogeneity in citizens' preferences for air quality information sources in our baseline survey, as shown in Figure 2. It is, however, unclear how much of this variation is driven by their beliefs about the accuracy of the sources or their affinity toward each of the information sources. As such, we study how citizens in a developing city form beliefs about air quality and modify their behavior as they infer the quality of information from its attributed source. We conduct a randomized information intervention to address the following research questions; First, are citizens willing to pay for air quality information, regardless of the sources to which it is attributed? Second, holding quality constant, is there differential willingness to pay for air quality information by the source? Third, how are beliefs about the service quality and the state of air pollution levels affected by the sources to which they are exposed? And lastly, does exposure to information from various sources induce differential policy preferences for environmental services?

We conduct an intervention in which we send daily air quality forecasts via SMS to a sample of residents in one of Lahore's working-class neighborhoods. We developed an ensemble forecast model of day-ahead air pollution using data inputs from multiple sources, including government and private monitors. We provide the identical day-ahead forecast to all 1,010 of our sample households but experimentally vary the salience of information sources. In one arm, respondents are told that the forecasts are constructed using data from Punjab Environmental Protection Department (EPD), a government agency responsible for reporting on air quality. On the other arm, respondents are told that the forecasts are constructed using data from a citizens' group called Pakistan Air Quality Initiative (PAQI). In other words, we vary the attributed sources but hold constant the information quality, which is correlated in naturally occurring setups. We do not have a control arm in which we do not provide air quality forecasts and focus on the differential effects of source attribution instead.

The experimental setup allows us to measure not only if and how citizens value air quality information but also how they value and trust the sources from which the information comes. We conduct a series of incentivized games in which we measure their willingness to pay for air quality forecasts, elicit their beliefs about air quality levels and the accuracy of our forecast service, and their preferences for monetary donations to government and private agencies.

We find that the recipients of our service in a working-class neighborhood are willing to pay PKR 238 on average for two additional months of service after the free trial during our experimental period. Yet, when we randomize the source to which the air quality information we provide is assigned, we do not find significant differences in either their willingness to pay or their forecast ability. We hypothesize that the recipients are equally well-informed about air pollution due to the identical air quality readings they received and the relatively consistent and reliable SMS forecast service we provide.

We still find, however, that the recipients' preferences for sources shift significantly toward the one to which they are exposed from the baseline of relatively weak preference. At baseline, most respondents choose to split their endowment 50:50 between the two sources to which they can donate. However, at the endline, they shift 75:25 toward the source to which they are assigned. This result may indicate that citizens' preferences for sources are relatively malleable. At the same time, we also find that recipients believe that the government service has a higher error of approximately 12% relative to the private alternative. This may indicate that citizens may hold beliefs of lower government service quality.

We contribute to several strands of literature, including environmental public services in developing economies, public-private competition, and belief formation. First, our work contributes to the emerging body of work on the demand for, and the challenges with the public provision of, environmental services (e.g., Ghanem and Zhang 2014 Freeman et al. 2019; Ito and Zhang 2020). We hope to understand the importance of information sources and consumers' beliefs in shaping the demand for such services. Second, we hope our findings will provide relevant insights to policymakers and stakeholders on the supply side of environmental service markets and contribute to the literature on accountability and competition for publicly provided services in developing economies (e.g., Muralidharan and Sundararaman 2015; Das et al. 2016; Jha and Nauze 2022). Third, our work relates to the literature on news media, particularly around mechanisms behind polarization of beliefs and trust in information sources (Gentzkow et al. 2023; Baysan 2022; Chopra et al. 2022). We hope to understand a) the role of beliefs and trust in shaping the demand for environmental information and b) the importance of prior beliefs and conditions under which beliefs about the state of the world and preferences for information services might diverge.

The rest of the paper is organized as follows; In section 2, we discuss the information environment for air quality in Lahore, Pakistan. In Section 3, we discuss the experimental design. In Section 4, we discuss the pre-specified outcome variables and discuss the identification strategy in Section 5. In Section 6, we discuss the specified results and conclude in Section 7.

## 2 Air quality information in Lahore

Lahore often ranks as one of the worst cities in terms of air quality around the world.[3] There are several sources of air quality information providing daily readings of PM2.5 concentration, yet access to the readings from these monitors is limited for the average citizen. Table 1 shows that approximately 9 percent of the representative sample of a working-class neighborhood state have received air pollution information from the main two sources: the Environment Protection Department (EPD) and an AirVisual app. We collect major on-the-ground, and publicly available air quality information sources, construct a forecast model based on these

[3] https://www.iqair.com/world-air-quality-ranking.

inputs, and provide forecast measures to our study sample. The following are the four main ground sources, plus a satellite-based measure (SPRINTARS) that we use in our model.

1. **Environment Protection Department (EPD)**: EPD is a department operating under the provincial government of Punjab, Pakistan. We collect their daily readings data from their website. The daily PDF reports include readings from three to four monitor locations, each reporting one of the scheduled pollutants. During our intervention period in the first half of 2023, the reports include readings on carbon monoxide ($CO$), particulate matter smaller than $2.5\mu m$ ($PM2.5$), and particulate matter smaller than $10\mu m$ ($PM10$). Each of these readings is reported in two different indices: (i) pollutants' concentration; and (ii) AQI (Air Quality Index). All of these indices are reported as 24-hour averages uploaded on the EPD's website every day around 9–11 AM. The reports also contain a disclaimer that "[any] other data from any source presenting ambient air quality of any city of Punjab is neither verified nor approved by the EPA Punjab."

2. **Pakistan Air Quality Initiative (PAQI)**: PAQI is a citizens' initiative that crowd-sources the collection of air quality readings and provides it via social media and other platforms. Started in 2016, PAQI crowd-sourced several low-cost air quality monitors (IQAir and PurpleAir) that were originally designed for indoor use. PAQI, among other operators, uploads their PM2.5 readings to an online platform named AirVisual. The platform reports both monitor-level and city-level readings at the hourly and daily concentration, going back as far as one month (e.g., Lahore and e.g. Lahore American School). Furthermore, PAQI reports city-level hourly readings from AirVisual on Twitter (such as @LahoreSmog). As such, we take the latter readings posted on Twitter as air quality measures provided by PAQI even though the reading is an average of

monitors, including ones that do not belong to PAQI.[4]

3. **U.S. Consulate**: The U.S. Consulate General in Lahore hosts an air quality monitor funded by U.S. EPA. The program, called AirNow International, places air quality monitors at U.S. embassies and consulates in mostly developing countries and provides hourly historical readings of $PM_{2.5}$ concentration. The monitor is located within the U.S. Consulate's compound in Shimla Hills, Lahore. The readings can be accessed via AirNow International's website. The U.S. consular services in Pakistan also share their readings from each consulate via Twitter (e.g., @Lahore_Air).

4. **Urban Unit**: The Urban Unit is a government-owned yet privately operated entity that addresses urban issues using data in Punjab Province. It was launched as part of a unit in the Planning and Development Department of the provincial government of Punjab in 2005 and was spun off to the private sector with full government ownership in 2012. The unit works on a range of issues pertaining to sustainable urban development, primarily in the realm of environmental services and management. The department owns a high-quality air quality monitor and had previously provided its readings on the banner of their website, but had stopped providing this daily information publicly prior to the beginning of our intervention in early 2023. They have an Environment Dashboard that individuals can sign up for and gain access to historical data on PM2.5 readings, but this data is updated at a lag of 10-15 days. We receive hourly average readings of PM2.5 concentration from the Unit's staff members on a daily basis.

5. **SPRINTARS**: Spectral Radiation-Transport Model for Aerosol Species (SPRINT-ARS) is a numerical model which estimates the effect of aerosols on the climatic system via simulations based on an atmosphere-ocean general circulation model called MIROC.

[4] This is because PAQI considers the city-level aggregate measure to be the most comprehensive of air quality information in Lahore and associates itself with it.

The model and estimates have been developed by the Climate Change Science Section at the Research Institute for Applied Mechanics, Kyushu University (Fukuoka, Japan). SPRINTARS considers both natural and anthropogenic sources of aerosols and categorizes them into suspended particulate matter (SPM), PM2.5, and PM10. Through a collaboration with the model's developers at Kyushu University, we access the hourly forecasts generated by SPRINTARS and construct the $t+1$ average forecast.

# 3 Research design

## 3.1 Sampling

The intervention is conducted in lower-middle-class neighborhoods of National Assembly (N.A.) constituencies 123 and 124 in northern Lahore. We divide the two constituencies into 200m$\times$200m blocks and randomly select 100 of them, weighted by population density. Figure 3 shows the selected blocks plus 20 backups. We then sample 1,010 households from the block centroids by following the left-hand rule: survey every ten households by spiraling out from the centroid counterclockwise.

## 3.2 Randomization

### 3.2.1 Treatment arms

Figure 4 shows that the sampled households are divided into the following two treatment arms:

- T1: SMS forecasts are attributed to a government agency (EPD)
- T2 SMS forecasts are attributed to a citizens' group (PAQI)

### 3.2.2 Stratified randomization

We stratify the randomization process into the two treatment groups (T1 vs. T2) on a set of baseline variables that either a) we considered as potential outcome variables, b) proxies of potential outcome variables that we were unable to collect at baseline due to the experimental design, c) candidate dimensions of heterogeneity, or d) the household asset index. [5] We use the optimal-greedy algorithm and generate blocks using the Minimum Volume Ellipsoid (MVE) estimator. We are primarily concerned about balance on outcome variables at baseline, as well as the "take-up" in terms of exposure and comprehension of our SMS forecast messages. We follow the advice from Athey and Imbens (2017) that each block contains two units per treatment arm. We then assign subjects to T1 and T2.

## 3.3 Forecast model

We design an ensemble model to forecast PM2.5 concentration for the next day (t+1) by building on Ahmad et al. (2022). In this subsection, we discuss a) how we define and measure the ground "truth" of air quality levels and b) how we construct the ensemble model.

[5]The final set of stratified variables are as follows:

1. absolute error of incentivized $t+1$ forecast of PM2.5 concentration (i.e., primary outcome 2.)
2. share of donations to government vs. citizens' group (i.e., primary outcome 4.)
3. time spent outdoors (i.e., secondary outcome 5.1.)
4. index: perceived accuracy and approval of government's services on air quality
5. index: perceived accuracy and approval of citizens' groups' services on air quality
6. 1 if comprehended a mock-up of the SMS forecast message without further explanation
7. 1 if reported to have received air pollution information from the EPD
8. 1 if reported to have received air pollution information from the AirVisual app (on which PAQI posts air quality readings)
9. Indicators of respondents' main T.V. news source
10. Asset index: a count of assets (electricity, appliances, vehicles, and number of rooms)

### 3.3.1 Defining the ground "truth"

For both the forecast model and the intervention, we first define the measure (which we refer to as the "truth") that the model predicts and how the readings are collected. The measure of interest is the average concentration of PM2.5 (in $\mu g/m^3$) between 12:00 AM and 4:00 PM for the day on which it is reported. This is because we send out the daily readings and the $t+1$ forecast between 6:00–8:00 PM, as we learned in our previous study, Ahmad et al. (2022), that most respondents make plans for the next day in the evening.

The daily readings that we provide as the "truth" and on which the forecast model is trained are from the U.S. Consulate monitor, as it is presumably of highest quality using the "reference method" in compliance with the U.S. EPA standards.[6]. On days where the U.S. Consulate monitor is missing data, we use the Urban Unit readings, which are also based on a high-quality monitor (BAM-1020 by MET). If both sources are missing, we use readings from PAQI, which are consistently available. As of 24th May 2023, the U.S. Consulate monitor is missing readings for 16 out of the 97 intervention days. Out of 16 days where U.S. Consulate is missing data, the Urban Unit is missing data on 4 days.

### 3.3.2 Constructing the ensemble forecast

We use an ensemble model that combines the following $t+1$ forecasts of the "truth":

1. predictions using data from the U.S. Consulate data
2. predictions using data from the Urban Unit
3. predictions using data from EPD
4. predictions using data that PAQI posts on Twitter

[6] https://www.epa.gov/system/files/documents/2022-12/List_of_FRM_and_FEM.pdf

5. *t+1* predictions from the SPRINTARS air pollution model

Combining these prediction models into an ensemble achieves two goals; first, it improves our overall predictive ability, and second, it allows us to attribute our predictions to different sources (i.e., EPD or PAQI) when the information is provided to the treatment households.

1. **Constructing individual predictions**

   Each model, except for SPRINTARS uses the following inputs:

   - the lagged readings from a given source on days $t-6$ to $t$
   - AccuWeather's *t+1* forecasts for minimum temperature, maximum temperature, and precipitation in inches, as well as their squared values
   - Historical weather data on a daily average, minimum, and maximum temperature, dew point temperature, wind speed and direction, visibility, and relative humidity from ASOS
   - Historical weather data on pressure and precipitation from Weather Underground

   Using the Adaptive Lasso model, we predict *j+1* PM2.5 concentration (i.e., the "truth") using a model trained on data from Day 1 to Day $j$, for $j$ going from Day 20 to $t$.

   SPRINTARS gives a model-based forecast, so we do not construct our own forecasts.

2. **Combining the forecasts to construct an ensemble model:**

   We estimate the root-mean-square error (RMSE) of each model over the period in which we have forecasts. We then weight the forecast based on the sum of RMSE across five models to their own (i.e., $w_i = \frac{\sum_{s \in S} RMSE_s}{RMSE_i * W}$ for a source $i$ in a set of sources

$S$, and $W$ is the sum of all $w_i$'s).

The ensemble forecast is the weighted sum of the individual forecasts.

## 3.4 Intervention: SMS forecast messaging

The main element of our intervention is the daily provision of the day-ahead (i.e., *t+1*) forecasts of PM 2.5 measures in $\mu g/m^3$ via SMS. In these messages, one of the sources (EPD or PAQI, chosen via the randomization procedure) is made salient. The daily messages also contain the readings from time $t$. The subjects also received an introductory message before the start of the daily SMSs, and a reminder message every two weeks over the course of the intervention. The daily messages are sent out around 6:00–8:00 PM starting on 18th February 2023 and continue through to the end of the endline survey period (currently expected in mid-to late June 2023). All of these messages are sent out using *OpenCodes*, an API-based system using a short-code service. All messages were in Urdu in the Urdu alphabet (Nastaliq script).

### 3.4.1 Introductory message

The following messages were sent to the subjects, depending on the assigned treatment arm:

- T1: "Assalam u alaikum! We visited your residence last month and did a survey on Air Pollution in Lahore where you agreed to receive air quality forecast information messages. You will be receiving these messages every day for the next 2 months.

  These messages are based on PM 2.5 data which is measured in micrograms per meter cube. The data is collected from the Punjab government's Environmental Protection Department (EPD) which is tasked with collecting information on Air Pollution. If you have any queries or questions about these messages, please contact the following number [telephone number]."

- T2: "Assalam u alaikum! We visited your residence last month and did a survey on Air Pollution in Lahore where you agreed to receive air quality forecast information messages. You will be receiving these messages every day for the next 2 months. These messages are based on PM 2.5 data which is measured in micrograms per meter cube. The data is collected from a non-governmental organization (NGO) called Pakistan Air Quality Initiative (PAQI [insert phonetic for PAQI in Urdu alphabet]) which collects data on air pollution. If you have any queries or questions about these messages, please contact the following number [telephone number]. "

### 3.4.2 Daily forecast messages

The daily messages are sent around 6:00–8:00 PM after collecting the day's data and estimating the forecast for *t+1*. We use the shorthand "NGO" to refer to organizations of a type, such as PAQI, for the purpose of familiarity with our subjects. The message on, for instance, 18th February 2023 would look as follows:

- T1: "Actual Air Quality (PM 2.5) on 18-02-23: 179
  Air Quality Forecast (PM 2.5) for 19-02-23 using data From Punjab Government (EPD): 231

- T2: "Actual Air Quality (PM 2.5) on 18-02-23: 179
  Air Quality Forecast (PM 2.5) for 19-02-23 using data From NGO (PAQI [insert phonetic for PAQI in Urdu alphabet]): 231

Figure 5 shows screenshots of the daily messages for T1 and T2. Because the text messages are sent from the same number every day, it is easy to compare the forecast values for Day $t$ provided on Day $t$-1 to the realized value provided on Day $t$.

### 3.4.3 Fortnightly reminder messages

Starting on Saturday, 4th March 2023, reminder messages are sent every two weeks on Saturday about the source and the unit of measurement. The messages by the treatment groups are as follows:

- T1: "The following messages on air pollution (PM 2.5) are based on data from the Punjab Governments Environment Protection Department (EPD). The data is measured in micrograms per meter cube."
- T2: "The following messages on air pollution (PM 2.5) are based on data from a non-government organization (NGO) named Pakistan Air Quality Initiative (PAQI [insert phonetic for PAQI in Urdu alphabet]). The data is measured in micrograms per meter cube."

## 3.5 Project timelines

The project timelines are as follows:

1. Design Phase: -January 2023
2. Pilot baseline survey: December 2022 - January 2023
3. Baseline survey: January 2023 - February 2023
4. SMS intervention: February 2023 to May 2023
5. Endline survey: May 2023 to June 2023

# 4 Primary outcome variabless

Following the hypotheses listed in Section 1, we identify primary outcomes of interest. There are four outcomes, over which we test five primary hypotheses. The primary outcomes are constructed from incentivized games in the endline survey. They are defined as follows:

1. Demand for air quality information as the willingness-to-pay (WTP) for SMS-based air quality forecasts

    - The outcome is defined as the amount respondents are willing to pay in PKR. We elicit respondents' willingness to pay for the SMS forecast using the Becker-DeGroot-Marshak (BDM) method (Becker et al. 1964). In the endline survey, we ask for the respondent's willingness to pay for the SMS-based air quality forecast messages. They have been receiving these messages for the past three months, and we ask for their willingness to pay for an additional two months. In the prompt, we make the experimentally assigned source salient by reminding them that the forecast is built using data from the said source. The bid's ceiling is set at PKR 400.

2. Beliefs about air quality levels as the absolute error of incentivized $t + 1$ forecast of PM2.5 concentration

    - The outcome is defined as the absolute difference between the actual PM2.5 concentration and the respondent's forecast, divided by the actual PM2.5 concentration. In both baseline and endline surveys, we ask respondents to make an incentivized guess of the air pollution level on day $t + 1$. In the baseline survey, we show respondents a table containing the average, minimum, and maximum of the average daily PM2.5 concentration over the last calendar week. We then ask them to forecast tomorrow's average PM2.5 concentration. Respondents receive

PKR 250 if their guess falls within 5% of the actual levels, PKR 150 if within 10%, and PKR 50 if within 20%. In the endline, we first elicit the forecast without the table containing the information from the previous calendar week. We then allow the respondents to revise their forecast after showing them the table.

3. Perceived accuracy of air-quality information source as the absolute error of incentivized guess of the SMS's forecast

    - The outcome is defined as the absolute difference between the respondent's guess of the PM2.5 forecast generated by our model and their own forecast for $t + 1$. In the endline survey, we not only ask respondents to forecast the actual PM2.5 concentration for tomorrow but also the value of our SMS forecast. The guess is financially incentivized, as in the guess for the actual PM2.5 concentration for tomorrow.

4. Preference for information source as the share of donations to government vs. citizens' group

    - The outcome is defined as the share of PKR 100 donated to a government agency for an environmental cause, as opposed to the citizen's group. We offer an opportunity to donate PKR 100 between two sources for environmental protection purposes: a government institution and PAQI.

# 5 Identification strategy

## 5.1 Exogenous variable

Our main exogenous variable is treatment assignment between the arm where the government (EPD) was made salient as the source, as opposed to the citizens' group (PAQI). We refer to being in the citizens' group arm as being in the "treatment," and the government arm as being in the "control" for the rest of this document. Let $Z_g$ denote treatment assignment as a vector, whose inputs are equal to 1 if the respondent is assigned to the government arm and 0 if assigned to the citizens' group arm.

## 5.2 Pre-specified hypotheses

The following are the five hypotheses that we test and for which we correct for multiple testing.

1. The demand for air quality information is greater than zero regardless of the treatment assignment group (tested on outcome 1.)
2. The demand for air quality information is different between the treatment (citizen's group) and control (government) groups (tested on outcome 1.)
3. Treatment affects beliefs about air quality differentially relative to control (tested on outcome 2.)
4. Treatment affects the perceived accuracy of air-quality information source relative to control (tested on outcome 3.)
5. Treatment affects policy preferences for air quality relative to control (tested on outcome 4.)

The above hypotheses correspond, in order, to the research questions specified in Section 1.

### 5.3 Test of positive willingness to pay for air quality information

To test for hypothesis 1., we simply use a t-test to see if the willingness to pay for the SMS forecasts is higher than 0. We pool the two treatment arms and conduct a one-tail test.

### 5.4 Treatment Effects

#### 5.4.1 Intent to treat

We estimate the treatment effects between subjects as follows;

$$Y_i = \alpha + {Z_{g_i}}'\beta + \boldsymbol{X_i'}\boldsymbol{\gamma} + \varepsilon_i$$

The matrix $\boldsymbol{X}$ includes control variables selected through a double-post-selection method using LASSO, as in Belloni et al. (2014). Given that we are agnostic as to which information source is more likely to shift beliefs, preferences, and beliefs related to air quality, our hypothesis tests are two-tailed: $\beta \neq 0$.

With the above estimating equation, we test hypotheses 2. and 4..

We estimate the treatment effects within subjects as follows;

$$Y_i = Z_{g_i}'\beta + \gamma Y_{0i} + \boldsymbol{X_i'}\boldsymbol{\delta} + \varepsilon_i$$

We denote $Y_0$ as the baseline measure of the outcome variable $Y$. Much of the details about the specification and inference are the same as in the between-subject model; we select the vector of controls $\boldsymbol{X}$ via a double-post-selection method with LASSO and estimate p-values using randomization inference. Our hypothesis tests are also two-sided, i.e., $\beta \neq 0$.

With the above estimating equation, we test hypotheses 3. and 5.. We also pre-specified a treatment-on-the-treated identification strategy in the pre-analysis plan. However, we do not find significant first-stage results and therefore put this identification strategy to Appendix Section D.1.

### 5.5 Heterogeneous effects

We consider dimensions of heterogeneity that we expect to drive the preferences for air quality information sources. We are primarily interested in (a) baseline beliefs about, and preferences for, information sources and (b) baseline beliefs about air quality levels and their deviation from the truth.

The first dimension is informed by an emerging body of work on media bias, trust for information sources and polarization. Theoretical and empirical work in this literature shows that agents may place heavier weights on information from a source that aligns with their priors, leading to polarization in preferences and beliefs (e.g., Gentzkow et al. 2023; Chopra et al. 2022).[7] If, on the other hand, agents do not exhibit belief confirmation or do not hold strong priors about the sources' quality, they may shift their priors more strongly to information from a source that they are less exposed to at baseline. As such, it is *a priori* unclear how the demand for the sources evolves based on their baseline preferences and beliefs. The second dimension is of more standard Bayesian concern in that individuals who are less well-informed about air quality levels may hold priors with more deviations from the truth. Those individuals may therefore update their beliefs more strongly toward the truth based on the signals they receive and value the SMS forecasts more.

[7] This may be driven by "belief confirmation," i.e., they prefer sources that distort information toward their prior beliefs (Mullainathan and Shleifer 2005), or driven by uncertainty about accuracy of information sources, inducing an individual to put heavier weights on their preferred source (Gentzkow and Shapiro 2006).

### 5.5.1 Dimensions of heterogeneity

For the dimension of heterogeneity on baseline preferences for, and beliefs about, the sources of air quality information, we use the following proxies:

1. donation share of PKR 100 between government's environmental agency vs. citizens' group that tackles air pollution

2. categorical variable of overall approval: "Government-leaning" if the respondents' Likert-scale approval measure for the government is greater than that for the citizens' group, "Citizens' group-leaning" if vice versa, and "neutral" if they equally approve the two sources

3. categorical variable of accuracy: same as above, except the Likert-scale measure captures respondents' beliefs about the accuracy of the sources' air quality information services.

For robustness, we also consider other definitions of baseline preferences and beliefs, such as the original Likert scales used to construct the proxies above, as well as the respondents' primary news sources' political leanings.

For the dimension of heterogeneity on baseline beliefs about air quality and its deviation from the truth, we use the following proxy:

- baseline outcome variable 2.: absolute error of incentivized $t+1$ forecast of PM2.5 levels.

We also use several other definitions of baseline beliefs to test, for instance, asymmetry based on the direction of the error.

### 5.5.2 Estimating equations

For brevity, we present the specification for within-subject analysis of the ITT effects. The between-subject analysis and TOT effects follow a similar structure. The estimating equation is as follows:

$$Y_i = \alpha + Z_{g_i}\boldsymbol{H_i'}\boldsymbol{\beta} + \boldsymbol{X_i}\boldsymbol{\gamma} + \varepsilon_i$$

$H_i$ is the relevant dimension of heterogeneity, coded as a matrix consisting of vectors of dummies that may represent bins of an underlying continuous or categorical variable. We include all bins so that we separately estimate treatment effects treatments for each bin (i.e., $\boldsymbol{\beta}$ is a vector), and $Z$ does not enter the equation separately.

We also estimate specifications where the underlying dimension of heterogeneity is a continuous variable (such as the donation share and the absolute forecast error). The estimation equation in such a case would be as follows:

$$Y_i = \alpha + Z_{g_i}H_i{}'\beta + H_i\beta_h + \boldsymbol{X_i}\boldsymbol{\gamma} + \varepsilon_i$$

In this specification, $H$ also enters separately to control for the baseline level of the dimension of heterogeneity.

# 6 Results

## 6.1 Checks on balance

We test the balance of variables used for blocking between the two treatment arms as well as other additional variables. The statistics we present include means for the two treatment arms, differences between the two treatment arms, and t-tests of the null hypothesis of zero difference.

Table 1 shows the balance on the variables used in the blocking procedure. We do not find statistically significant differences in any of the primary outcomes for which we have baseline measures or other variables over which we stratified our randomization.

In the next version of the working paper, we will run a regression of the following form and estimate the F statistic and the p-value with heterogeneity-robust standard errors.

$$Z_{g_i} = \boldsymbol{X_i'}\boldsymbol{\eta_1} + \boldsymbol{W_i'}\boldsymbol{\eta_2} + \epsilon_i$$

We will also evaluate the balance of the attrition rate by assessing if attritors and non-attritors differ on observables when interacting with treatment assignments. First, we report attrition rates by the two experimental arms. We will then compare attritors and non-attritors on observables as follows, where $D$ is an attrition dummy;

$$D_i = Z_{g_i}\kappa_1 + Z_{g_i}\boldsymbol{X_i'}\boldsymbol{\kappa_2} + Z_{g_i}\boldsymbol{W_i'}\boldsymbol{\kappa_3} + \boldsymbol{X_i'}\boldsymbol{\kappa_4} + \boldsymbol{W_i'}\boldsymbol{\kappa_5} + \omega_i$$

Again, we estimate the F statistic and the p-value obtained with heterogeneity-robust standard errors.

## 6.2 Prespecified primary outcomes: Intent-to-treat

Table 2 shows the coefficients and their standard errors of the intend-to-treat estimates for the five prespecified primary hypotheses using post-double-selection LASSO. Here, by "treatment," we mean being assigned to the government arm, as opposed to the citizen's group arm. Table 3 shows the p- and q-values of the corresponding columns.

Column 1 in Table 2 shows that the respondents are willing to pay PKR 238 for two months of air quality forecast services. In comparison, a standard prepaid mobile cell and data plan as of July 2023 costs PKR 87.[8] This means that the individuals who received the air pollution services are willing to pay approximately 1.4 times a typical prepaid mobile plan. Figure 6 also shows the distribution of the willingness-to-pay for air quality forecasts as demand functions, indicating considerable heterogeneity.

We find, however, that there are no statistically and economically significant differences between the treatment arms in their willingness to pay for the forecasts. Column 2 in Table 2 shows that those assigned to the Government arm are willing to pay only PKR 0.33 more on average, and the difference is not statistically significant from zero. The small coefficient and standard error also exclude any economically meaningful difference between the two treatment arms.

We do not find that the treatment arms lead to differential beliefs about air quality levels, as measured through their own forecast of air quality levels the next day. Column 3 shows that there are small differences in the air quality forecast error for the next day that are statistically indistinguishable from zero. However, we find that respondents assigned to the Government arm report that they believe the SMS forecast they receive for tomorrow will be more significantly different from their own forecast for tomorrow. Column 4 shows that this difference is 2.8 points here for the Government arm, relative to the mean of 22.7 for the Private arm.

We also find that the respondents donate a larger fraction of their endowment to the source to which they are assigned than to the other source. Figure 9, which shows the amount donated to the government out of PKR 100 in the donation game, demonstrates this point. More than 90 percent of respondents who are assigned to the Government arm

[8] A standard plan offered by Jazz includes 3GB data, free WhatsApp, 1000 minutes of in-network calls, 1,000 SMS, and 30 minutes of out-of-network calls (https://jazz.com.pk/prepaid/bemisaal-offer-1).

donate more to the government, as opposed to the private alternative. On the other hand, more than 90 percent of respondents assigned to the Private arm donate more to the private alternative. The average ratio between the assigned source and the other is around 3:1. Column 5 in Table 2 confirms that those assigned to the Government arm donate PKR 54 more to the government, on average, relative to the respondents in the Private arm.

Tables A.1 and A.2 also show the results using winsorized outcomes at the 1st and 99th percentiles. The tables correspond to Tables 2 and 3, respectively.

Results using alternative definitions of the primary outcome variables are shown in Tables, A.3, A.4, A.5, and A.6.

### 6.3 Prespecified primary outcomes: Treatment-on-treated

Table 4 reports the first-stage results, where our preferred measure of treatment exposure is "days read," i.e., the number of days in an average week in which the respondents report to have read the air pollution forecast message.

## 7 Conclusion

We study how residents living under uncertainty about the state of air pollution and information quality provided by multiple sources form beliefs and demand for information services through a randomized control trial. We find a high level of willingness to pay among a working-class population in Lahore, Pakistan.

Yet, when we randomize the source to which the air quality information we provide is assigned, we do not find significant differences in either their willingness to pay or their forecast ability. We hypothesize that the recipients are equally well-informed about air pollution due to the identical air quality readings they received and the relatively consistent and reliable SMS forecast service we provide.

We still find, however, that the recipients' preferences for sources shift significantly toward the one to which they are exposed from the baseline of relatively weak preference. This result may indicate that citizens' preferences for sources are relatively malleable. At the same time, we also find that recipients believe that the government service has higher errors relative to the private alternative. This may indicate that citizens may hold beliefs of lower government service quality.

We plan on conducting further analysis by exploring heterogeneous responses and effects on secondary outcomes and discussing our findings through a theoretical framework in the next update to the draft.

# References

Husnain F Ahmad, Matthew Gibson, Fatiq Nadeem, Sanval Nasim, and Arman Rezaee. Forecasts: Consumption, Production, and Behavioral Responses. 2022.

S. Athey and G.W. Imbens. The Econometrics of Randomized Experiments. In *Handbook of Economic Field Experiments*, volume 1, pages 73–140. Elsevier, 2017. ISBN 978-0-444-63324-8. doi: 10.1016/bs.hefe.2016.10.003. URL `https://linkinghub.elsevier.com/retrieve/pii/S2214658X16300174`.

Ceren Baysan. Persistent Polarizing Effects of Persuasion: Experimental Evidence from Turkey. *American Economic Review*, 112(11):3528–3546, 2022. ISSN 0002-8282. doi: 10.1257/aer.20201892. URL `https://www.aeaweb.org/articles?id=10.1257/aer.20201892`.

Gordon M. Becker, Morris H. Degroot, and Jacob Marschak. Measuring utility by a single-response sequential method. *Behavioral Science*, 9 (3):226–232, 1964. ISSN 1099-1743. doi: 10.1002/bs.3830090304. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/bs.3830090304`. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/bs.3830090304.

A. Belloni, V. Chernozhukov, and C. Hansen. Inference on Treatment Effects after Selection among High-Dimensional Controls. *The Review of Economic Studies*, 81(2):608–650, April 2014. ISSN 0034-6527, 1467-937X. doi: 10.1093/restud/rdt044. URL `https://academic.oup.com/restud/article-lookup/doi/10.1093/restud/rdt044`.

Felix Chopra, Ingar Haaland, and Christopher Roth. CESifo Working Paper no. 9673. 2022.

Jishnu Das, Alaka Holla, Aakash Mohpal, and Karthik Muralidharan. Quality and Accountability in Health Care Delivery: Audit-Study Evidence from Primary Care in India. *Amer-*

ican Economic Review*, 106(12):3765–3799, December 2016. ISSN 0002-8282. doi: 10.1257/aer.20151138. URL https://www.aeaweb.org/articles?id=10.1257/aer.20151138.

Richard Freeman, Wenquan Liang, Ran Song, and Christopher Timmins. Willingness to pay for clean air in China. *Journal of Environmental Economics and Management*, 94:188–216, March 2019. ISSN 0095-0696. doi: 10.1016/j.jeem.2019.01.005. URL https://www.sciencedirect.com/science/article/pii/S0095069617308902.

Matthew Gentzkow and Jesse M. Shapiro. Media Bias and Reputation. *Journal of Political Economy*, 114(2):280–316, 2006. ISSN 0022-3808. doi: 10.1086/499414. URL https://www.jstor.org/stable/10.1086/499414. Publisher: The University of Chicago Press.

Matthew Gentzkow, Michael B Wong, and Allen T Zhang. Ideological Bias and Trust in Information Sources. 2023.

Dalia Ghanem and Junjie Zhang. 'Effortless Perfection:' Do Chinese cities manipulate air pollution data? *Journal of Environmental Economics and Management*, 68(2):203–225, September 2014. ISSN 0095-0696. doi: 10.1016/j.jeem.2014.05.003. URL https://www.sciencedirect.com/science/article/pii/S0095069614000400.

Koichiro Ito and Shuang Zhang. Willingness to Pay for Clean Air: Evidence from Air Purifier Markets in China. *Journal of Political Economy*, 128(5):1627–1672, May 2020. ISSN 0022-3808. doi: 10.1086/705554. URL https://www.journals.uchicago.edu/doi/full/10.1086/705554. Publisher: The University of Chicago Press.

Akshaya Jha and Andrea La Nauze. US Embassy air-quality tweets led to global health benefits. *Proceedings of the National Academy of Sciences*, 119(44):e2201092119, November 2022. doi: 10.1073/pnas.2201092119. URL https://www.pnas.org/doi/10.1073/pnas.2201092119. Publisher: Proceedings of the National Academy of Sciences.

Sendhil Mullainathan and Andrei Shleifer. The Market for News. *American Economic Review*, 95(4):1031–1053, September 2005. ISSN 0002-8282. doi: 10.1257/0002828054825619. URL https://www.aeaweb.org/articles?id=10.1257/0002828054825619.

Karthik Muralidharan and Venkatesh Sundararaman. The Aggregate Effect of School Choice: Evidence from a Two-Stage Experiment in India. *The Quarterly Journal of Economics*, 130(3):1011–1066, August 2015. ISSN 0033-5533. doi: 10.1093/qje/qjv013. URL https://doi.org/10.1093/qje/qjv013.

# 8 Figures

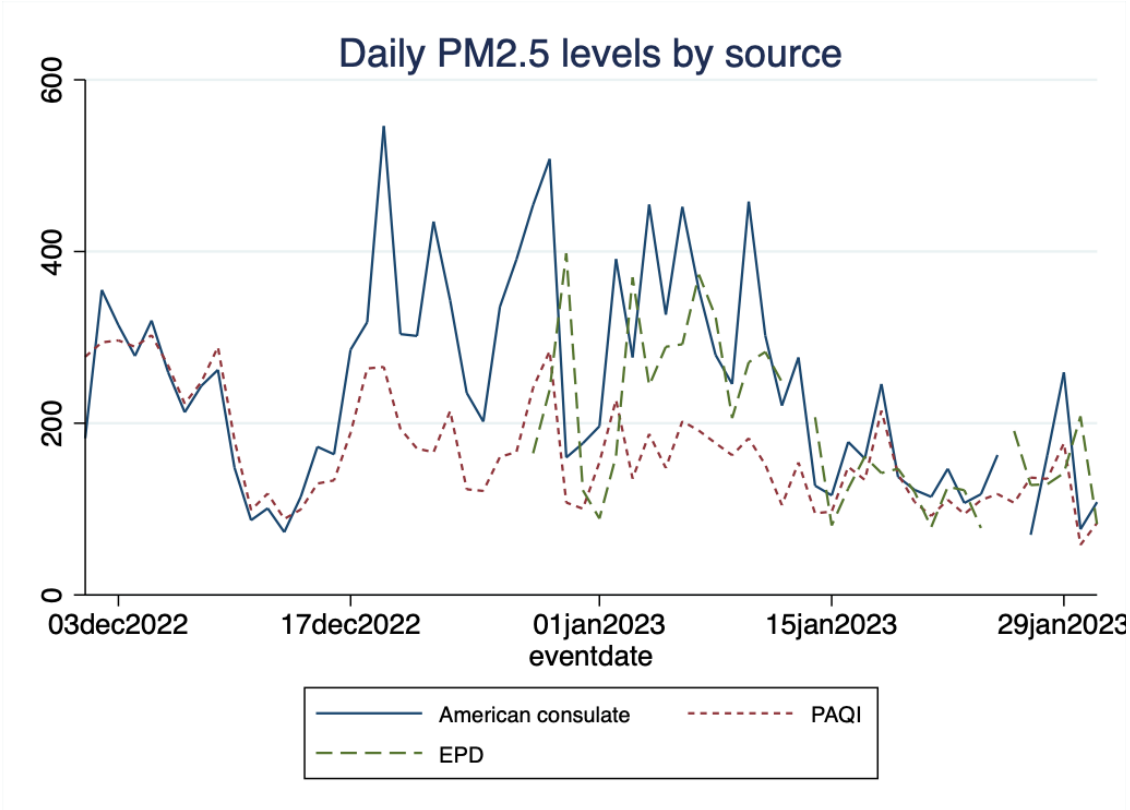



Figure 1: This figure shows the daily average PM2.5 concentration (in $\mu g/m^3$ ) levels by sources. "American consulate" refers to readings from the air quality monitor at the American consulate in Lahore. We treat this reading as the ground truth. "PAQI" refers to readings from the average of lower-cost air quality monitors managed by Pakistan Air Quality Initiative (PAQI) in Lahore. "EPD" refers to readings from air quality monitors managed by the Environmental Protection Department (EPD) of the Government of Punjab Province.

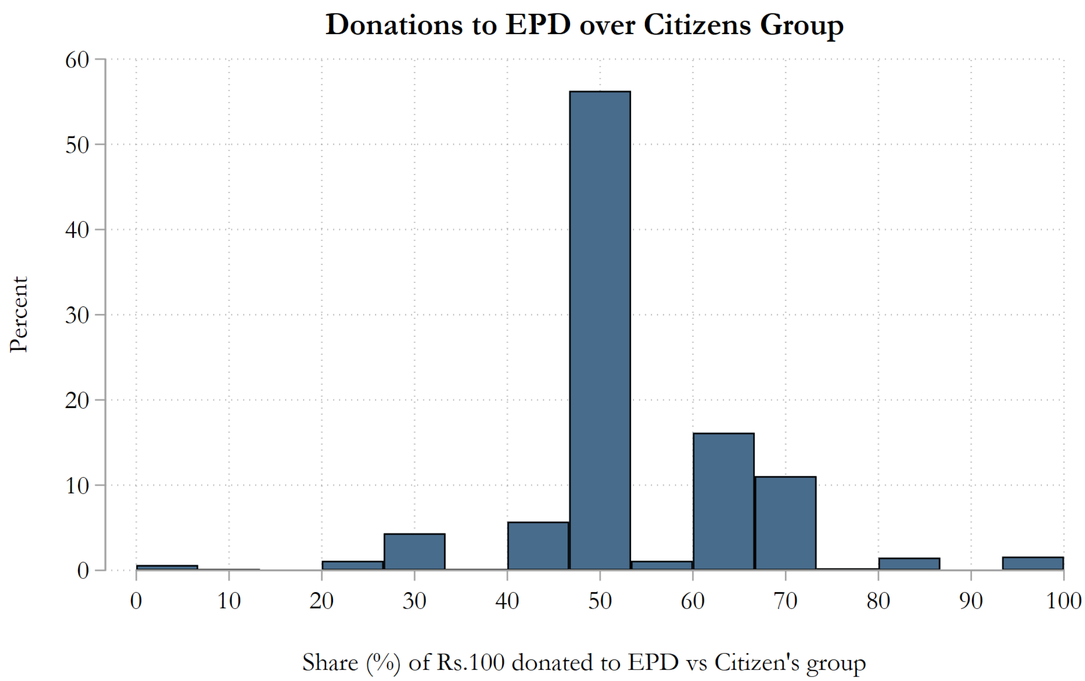



Figure 2: This figure shows the result from the donation game in our baseline survey, in which we asked respondents to split PKR 100 between government (EPD) and private (PAQI) sources.

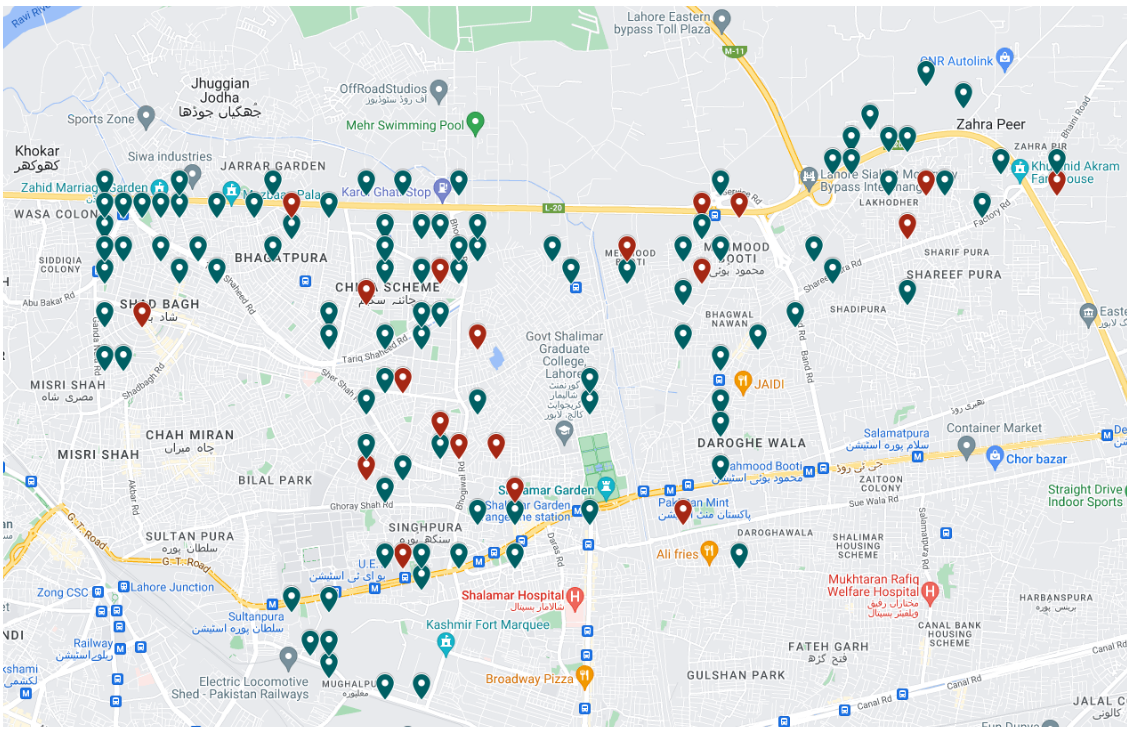



Figure 3: Sampling coordinates in NA-123 and NA-124 constituencies in Lahore, Pakistan

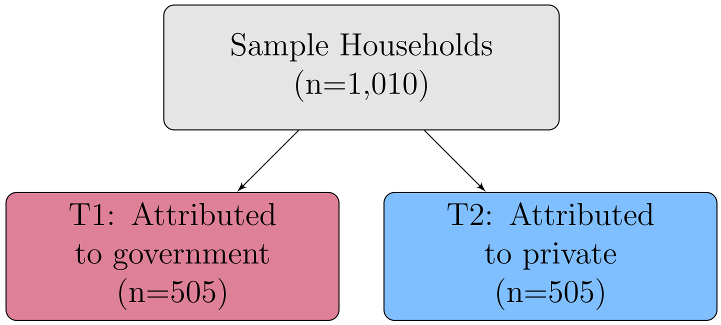
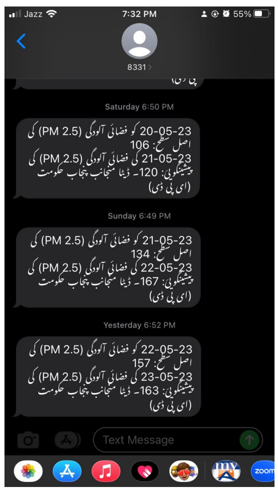



Figure 4: Treatment Groups

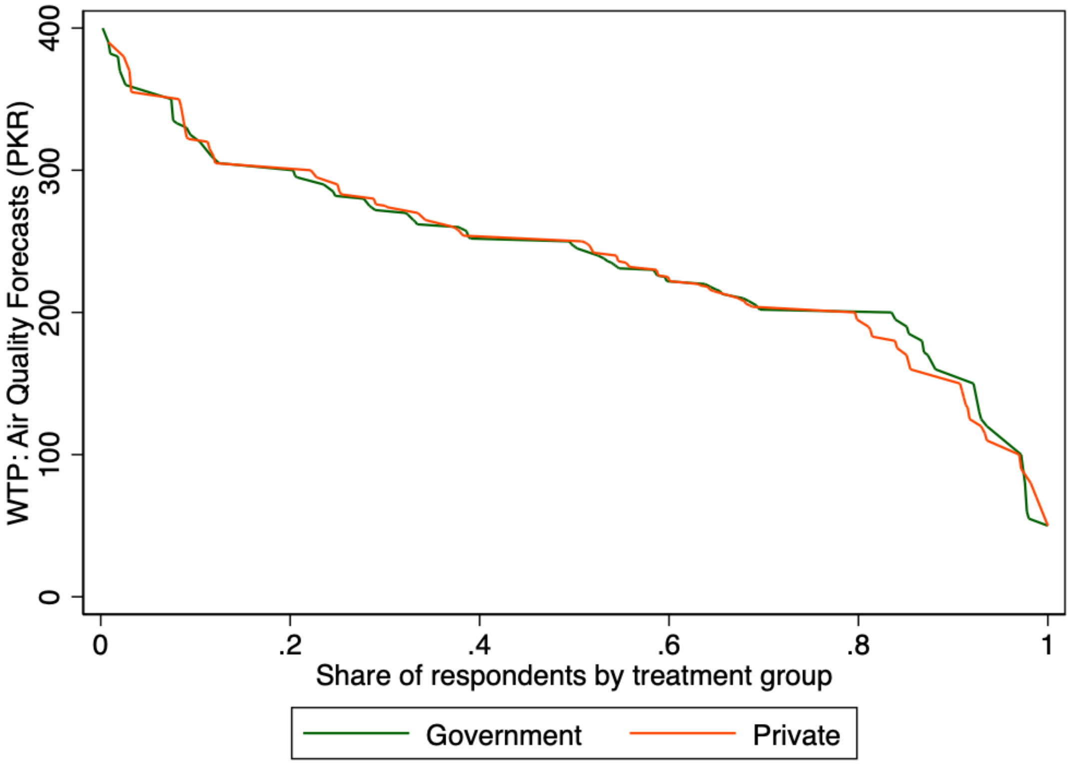



(a) T1: Daily messages


(b) T2: Daily messages

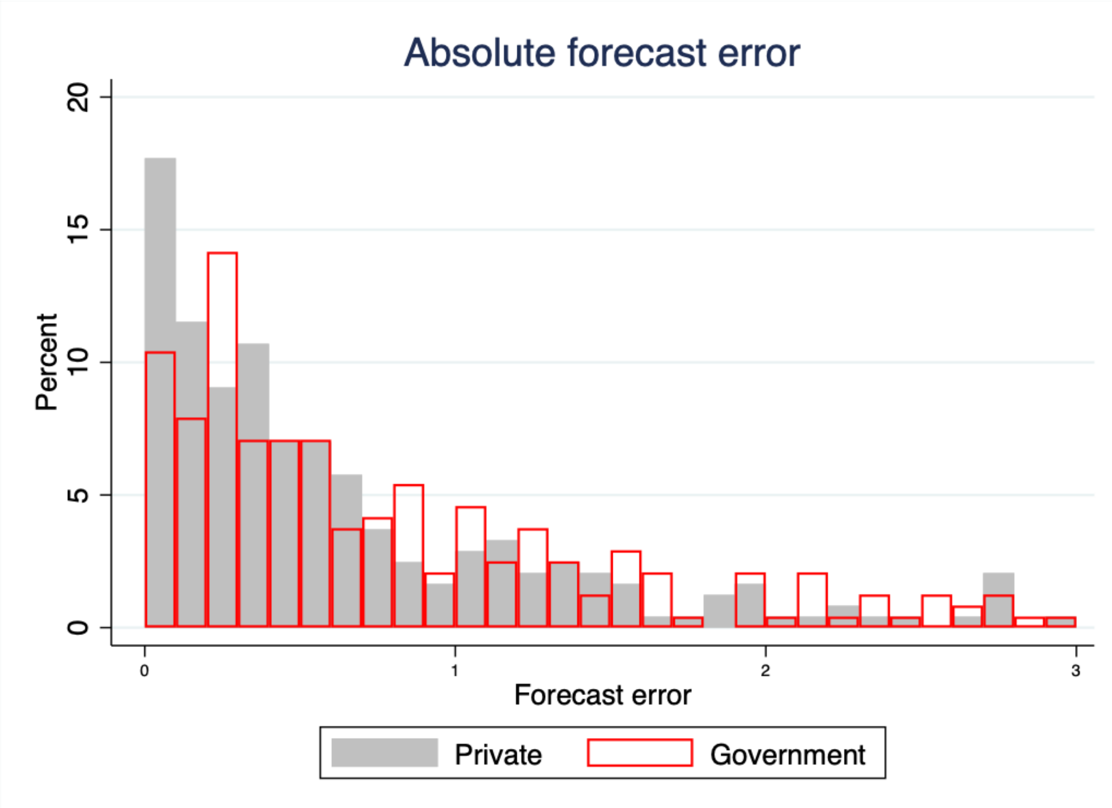
Figure 5: Sample messages to respondents

Figure 6: Demand curves for air pollution forecast by treatment



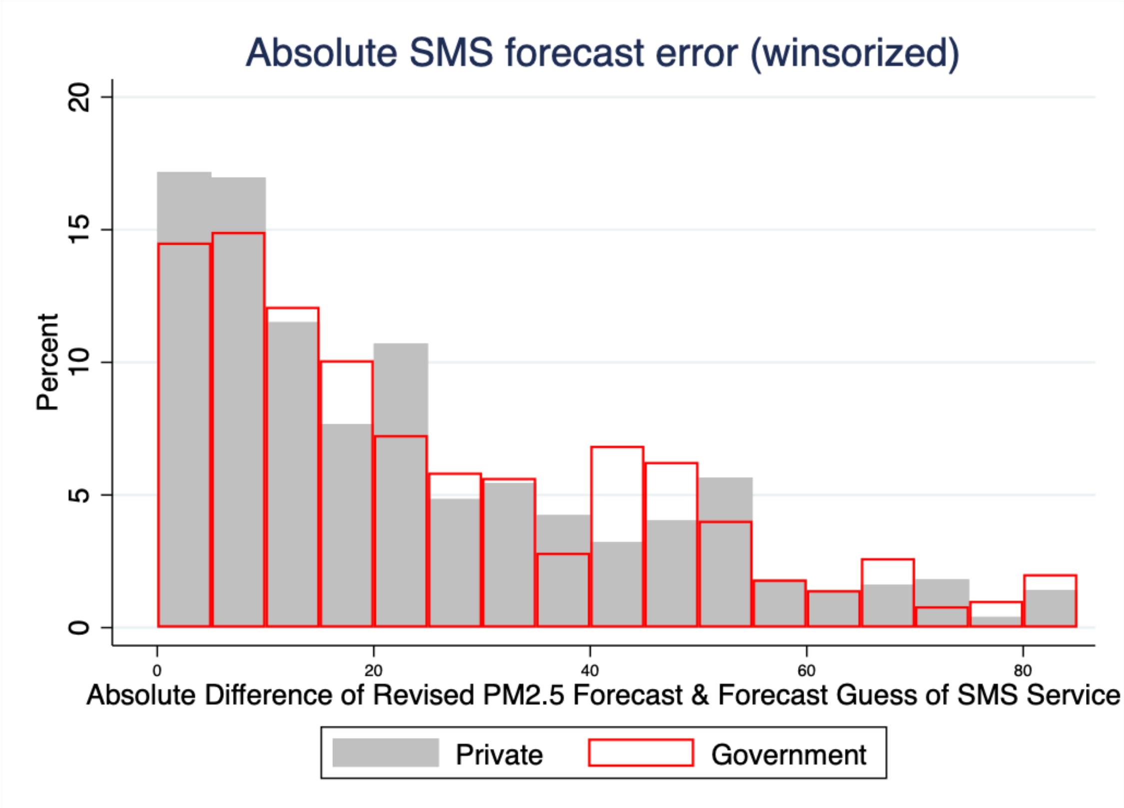

This figure shows the distributions of respondents' bids for two months of air pollution forecast service from the endline survey. "Government" corresponds to the arm in which the EPD source is made salient, and "Private" the PAQI source.

Figure 7: Absolute forecast error by treatment



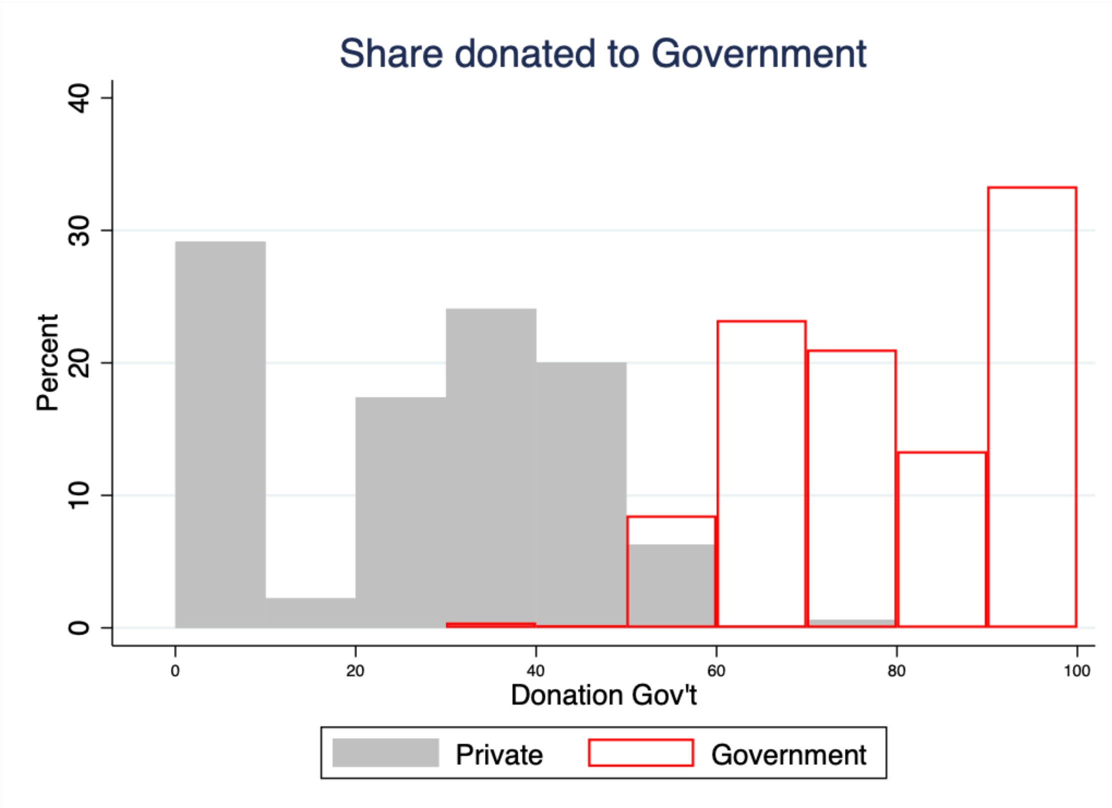

This figure shows the distributions of respondents' absolute forecast error in the endline survey. The measure is defined as the absolute difference between their forecast air pollution level on the next day and the actual reading, divided by the actual reading. "Government" corresponds to the arm in which the EPD source is made salient, and "Private" the PAQI source.

Figure 8: Estimate of SMS forecast error by treatment




This figure shows the distributions of respondents' beliefs about the absolute error of SMS forecasts, measured at the endline survey. The measure is defined as the absolute difference between their forecast air pollution level and their guess of the SMS forecast on the next day. "Government" corresponds to the arm in which the EPD source is made salient, and "Private" the PAQI source.

Figure 9: Donation to government sources vs private




This figure shows the distributions of respondents' donations to a government agency vs. a non-government entity for environmental protection, measured at the endline survey. The measure is defined as the amount out of PKR 100 donated to the government source. "Government" corresponds to the arm in which the EPD source is made salient, and "Private" the PAQI source.

# 9 Tables

Table 1: Balance Table

| Variable | (1)<br>No<br>Mean/(SE) | (2)<br>Yes<br>Mean/(SE) | (1)-(2)<br>Pairwise t-test<br>Mean difference |
|---|---|---|---|
| Absolute Difference of PM 2.5 Truth and Forecast | 96.014<br>(2.092) | 93.812<br>(2.183) | 2.202 |
| Ratio of Absolute Difference and Truth for PM 2.5 | 0.724<br>(0.019) | 0.713<br>(0.019) | 0.011 |
| Share (%) of Rs. 100 donated to EPD | 50.139<br>(0.681) | 50.059<br>(0.655) | 0.079 |
| Hours Spent Outdoors | 7.414<br>(0.204) | 7.446<br>(0.197) | -0.032 |
| Stated preference for citizens group | 0.009<br>(0.042) | -0.008<br>(0.043) | 0.017 |
| Stated preference for government | -0.013<br>(0.043) | -0.008<br>(0.043) | -0.005 |
| Comprehended the text message without explanation | 0.768<br>(0.019) | 0.766<br>(0.019) | 0.002 |
| Received air pollution info from: EPD | 0.087<br>(0.013) | 0.083<br>(0.012) | 0.004 |
| Received air pollution info from: AirVisual App | 0.097<br>(0.013) | 0.089<br>(0.013) | 0.008 |
| Index: Sentiment on air quality | -0.022<br>(0.032) | 0.008<br>(0.032) | -0.029 |
| Main TV channel: Geo News | 0.428<br>(0.022) | 0.457<br>(0.022) | -0.030 |
| Main TV channel: ARY News | 0.156<br>(0.016) | 0.139<br>(0.015) | 0.018 |
| Main TV channel: City 42 | 0.081<br>(0.012) | 0.095<br>(0.013) | -0.014 |
| Main TV channel: Express News | 0.081<br>(0.012) | 0.081<br>(0.012) | 0.000 |
| Index: Household asset ownership | 0.022<br>(0.046) | -0.022<br>(0.043) | 0.043 |
| F-test of joint significance (F-stat) | | | 0.359 |
| Number of observations | 505 | 505 | 1010 |

Significance: ***=.01, **=.05, *=.1. Errors are robust.

Table 2: Prespecified hypotheses: ITT

| | (1) WTP | (2) WTP | (3) Forecast error | (4) SMS error | (5) Donation gov't |
|---|---|---|---|---|---|
| Constant | 237.5*** | | | | |
| | (2.19) | | | | |
| Gov't arm | | 0.33 | 0.051 | 2.82** | 53.8*** |
| | | (3.68) | (0.040) | (1.29) | (1.04) |
| Observations | 993 | 993 | 993 | 991 | 989 |
| Endline mean of PVT | | 237.2 | 0.73 | 22.7 | 22.9 |

*Notes:* Model: PDSLASSO. "WTP": Willingness to pay for two months of SMS air quality forecasts. "Forecast error": the absolute difference between their forecast air pollution level on the next day and the actual reading, divided by the actual reading. "SMS error": the absolute difference between their forecast air pollution level and their guess of the SMS forecast on the next day. "Donation gov't": amount out of PKR 100 donated to the government source. Standard errors are reported in parentheses. All regressions include randomization-strata fixed effects, and heteroskedasticity-robust standard errors are used. Two-tailed significance: p<0.1*; p<0.05**; p<0.01***.

Table 3: Adjustments for multiple hypothesis testing on prespecified hypotheses

| | (1) WTP | (2) WTP | (3) Forecast error | (4) SMS error | (5) Donation gov't |
|---|---|---|---|---|---|
| P value | 0 | .927 | .208 | .029 | 0 |
| Q value | .001 | .351 | .116 | .03 | .001 |

*Notes:* We show the critical values for the "Constant" and "Gov't arm" coefficients in the corresponding columns of Table 2. "P value:" Unadjusted p-values. "Q value": Benjamini Krieger Yekutieli (2006) sharpened q-values.

Table 4: First-stage results for ToT

| | (1)<br>First-stage index | (2)<br>Days received | (3)<br>Days read |
|---|---|---|---|
| Gov't arm | 0.041 | 0.042 | 0.083 |
| | (0.031) | (0.038) | (0.078) |
| Observations | 993 | 993 | 993 |
| Endline mean of PVT | -0.028 | 6.59 | 4.88 |

*Notes:* "First-index": Index of created from "Days received" and "Days read." "Days received": Number of days in an average week in which forecasts are received. "Days read": Number of days in an average week in which forecasts are read. Standard errors are reported in parentheses. All regressions include randomization-strata fixed effects, and heteroskedasticity-robust standard errors are used. Two-tailed significance: $p<0.1$*; $p<0.05$**; $p<0.01$***.

# A Appendix tables

Table A.1: Prespecified hypotheses: ITT (winsorized)

| | (1) WTP | (2) WTP | (3) Forecast error | (4) SMS error | (5) Donation gov't |
|---|---|---|---|---|---|
| Constant | 237.4*** | | | | |
| | (2.18) | | | | |
| Gov't arm | | 0.31 | 0.051 | 1.89* | 53.8*** |
| | | (3.67) | (0.037) | (1.01) | (1.04) |
| Observations | 993 | 993 | 993 | 991 | 989 |
| Endline mean of PVT | | 237.2 | 0.73 | 22.7 | 22.9 |

*Notes:* We winsorize the outcome variables at the 1st and 99th percentiles. "WTP": Willingness to pay for two months of SMS air quality forecasts. "Forecast error": the absolute difference between their forecast air pollution level on the next day and the actual reading, divided by the actual reading. "SMS error": the absolute difference between their forecast air pollution level and their guess of the SMS forecast on the next day. "Donation gov't": amount out of PKR 100 donated to the government source. Standard errors are reported in parentheses. All regressions include randomization-strata fixed effects, and heteroskedasticity-robust standard errors are used. Two-tailed significance: p<0.1*; p<0.05**; p<0.01***.

Table A.2: Adjustments for multiple hypothesis testing on prespecified hypotheses (winsorized)

| | (1) WTP | (2) WTP | (3) Forecast error | (4) SMS error | (5) Donation gov't |
|---|---|---|---|---|---|
| P value | 0 | .932 | .177 | .061 | 0 |
| Q value | .001 | .284 | .113 | .065 | .001 |

*Notes:* We show the critical values for the "Constant" and "Gov't arm" coefficients in the corresponding columns of Table A.1. "P value:" Unadjusted p-values. "Q value": Benjamini Krieger Yekutieli (2006) sharpened q-values.

Table A.3: ITT: Alternative definitions of the WTP outcome

| | (1) WTP | (2) WTP (other) | (3) diff(WTP) | (4) diff(WTP) |
|---|---|---|---|---|
| Gov't arm | 0.33 | 0.55 | -0.21 | |
| | (3.68) | (3.66) | (0.54) | |
| Constant | | | | 15.9*** |
| | | | | (0.60) |
| Observations | 993 | 993 | 993 | 993 |
| Endline mean of PVT | 237.2 | 221.2 | 16.0 | |

*Notes:* "WTP": The prespecified outcome measuring the willingness to pay for two months of SMS air quality forecasts, where the assigned source is made salient. "WTP if other source": hypothetical WTP if the forecast were to come from the other source not assigned to them. "diff(WTP sources)": the difference between the willingness to pay for the assigned vs. the other sources. Standard errors are reported in parentheses. All regressions include randomization-strata fixed effects, and heteroskedasticity-robust standard errors are used. Two-tailed significance: p<0.1*; p<0.05**; p<0.01***.

Table A.4: ITT: Alternative definitions of the forecast outcome

| | (1) abs(own - truth)/truth | (2) (own - truth)/truth | (3) abs(own - truth) | (4) (own - truth) |
|---|---|---|---|---|
| Gov't arm | 0.051 | 0.060 | -0.48 | 3.39 |
| | (0.040) | (0.049) | (2.17) | (2.85) |
| Observations | 993 | 993 | 993 | 993 |
| Endline mean of PVT | 0.73 | 0.47 | 0.47 | 0.47 |

*Notes:* We present effects on forecast outcomes with different definitions, where "own" stands for the respondent's own forecast of the air quality level the next day, and "truth" the actual readings on the corresponding day. Standard errors are reported in parentheses. All regressions include randomization-strata fixed effects, and heteroskedasticity-robust standard errors are used. Two-tailed significance: p<0.1*; p<0.05**; p<0.01***.

Table A.5: ITT: Alternative definitions of the SMS forecast outcome

| | (1) abs(SMS - own) | (2) (SMS - own) | (3) abs(SMS - truth) | (4) (SMS - truth) |
|---|---|---|---|---|
| Gov't arm | 2.82** | 0.64 | 0.060 | 5.05 |
| | (1.29) | (1.89) | (2.02) | (3.09) |
| Observations | 991 | 991 | 991 | 991 |
| Endline mean of PVT | 22.7 | 4.60 | 46.1 | 10.4 |

*Notes:* We present effects on forecast outcomes with different definitions, where "SMS" stands for the respondent's guess of the SMS forecast they will receive for tomorrow, and "own" stands for the respondent's own forecast of the air quality level the next day. Standard errors are reported in parentheses. All regressions include randomization-strata fixed effects, and heteroskedasticity-robust standard errors are used. Two-tailed significance: p<0.1*; p<0.05**; p<0.01***.

Table A.6: ITT: Alternative definitions of the donation outcome

| | (1) Donation at endline | (2) Change from baseline to endline |
|---|---|---|
| Gov't arm | 53.8*** | 54.0*** |
| | (1.04) | (1.05) |
| Donation at baseline | 0.080 | |
| | (0.068) | |
| Observations | 989 | 989 |
| Endline mean of PVT | 22.9 | 22.9 |

*Notes:* Standard errors are reported in parentheses. All regressions include randomization-strata fixed effects, and heteroskedasticity-robust standard errors are used. Two-tailed significance: p<0.1*; p<0.05**; p<0.01***.

# B Data

## B.1 Survey data

### B.1.1 Survey frequency

We conduct the following surveys:

- Baseline survey (11th to 31st January 2023)
- Endline survey (29th May to mid/late June 2023)

### B.1.2 Survey modules

In the baseline survey, we ask for demographics, some of the outcome measures (i.e., outcomes that are not contingent on the subjects' having experienced the forecast service), and dimensions of heterogeneity. Detailed survey instruments are included in the appendix. We provide detailed descriptions of outcomes and other variable definitions in Section 4.

The baseline survey modules are as follows:

- Identification of a decision maker in the household as the respondent and consent
- Household roster and their demographics
- Awareness about air pollution in Lahore and access to information
- Donation game between EPD and PAQI, and stated preferences for the sources
- Stated beliefs in their trust in government services
- Incentivized forecast of air pollution (PM 2.5) concentration tomorrow

- Attitudes and behaviors regarding air pollution
- Time use survey and outdoor activities
- Participation in the local community and civil society
- Access to news sources and preferred channels
- Household assets

The endline survey modules are as follows:

- Identification of the same respondent as in the baseline and consent
- Incentivized forecast of air pollution (PM 2.5) levels tomorrow and incentivized guess of the SMS's forecast
- Value elicitation of the SMS forecast service through a bidding game using the BDM method
- Access to information about air pollution and stated satisfaction with the SMS forecast service
- Donation game between EPD and PAQI, and stated preferences for the sources
- Preferences for air quality-related policies via hypothetical scenarios
- Attitudes and behaviors regarding air pollution
- Time use survey and outdoor activities
- Stated mask usage
- Interest in filing complaints about air pollution to government authorities

## B.2 Air quality data

We collect air quality reading data from five different sources for the forecast model and for the intervention. We provide further detail on each of the data sources in Section 2.

## B.3 Weather Data

We also collect weather data as inputs for the forecast model, as described in further detail in Section 3.3.2.

- **AccuWeather:** We scrape daily forecasts on maximum and minimum temperatures and precipitation probability from AccuWeather for Lahore at `https://www.accuweather.com/en/pk/lahore/260622/daily-weather-forecast/260622`. AccuWeather uses NOAA's (National Oceanic and Atmospheric Administration) data and constructs its own forecasts.

- **ASOS**: We also collect detailed meteorological data collected by weather stations at airports. The data sources are called Automated Surface/Weather Observing Systems (ASOS/AWOS) or, more generically, METeorological Aerodome Reports (METARs). We use a web repository of these data sets hosted by Iowa State University's Iowa Environmental Mesonet and collect data for a station named "[OPLA] LAHORE(CIV/MIL)" via the following link: `https://mesonet.agron.iastate.edu/request/download.phtml?network=PK__ASOS`.

- **Weather Underground**: We also collect data on average and minimum atmospheric pressure and daily total precipitation from Weather Underground (URL: `https://www.wunderground.com/weather/pk/lahore`).

# C Power Calculations

We estimate the minimum detectable effect sizes on our primary outcomes at 80% probability, with $\alpha = 0.05$. We assume 15 percent attrition on our sample of 1,010. We also make conservative adjustments by dividing the $\alpha$ level by the number of tests for which we are identifying minimum treatment effect sizes.

There are two iterations to our power calculations. First, we identified the number of experimental arms and sample size based on the minimum detectable effect sizes during the design phase in June 2022. Out of the five hypotheses we present in this pre-analysis plan, we had only identified two of them during the design phase (and therefore divide $\alpha$ by 2). We then take sample means and standard deviations from survey data used in Ahmad et al. (2022). The outcomes, sample means, and standard deviations in parentheses are as follows:

1. Willingness-to-pay (WTP) for SMS-based air quality forecasts: 89.6 (45.2)

2. Absolute error of incentivized $t + 1$ forecast of PM2.5 concentration: 43.4 (43.0)

We find that we are able to detect impacts of 0.27 standard deviations, which is equal to PKR 12.3 in the willingness to pay, and 11.7 $\mu g/m^3$ for PM2.5 concentration.

Second, we re-estimate the minimum detectable effect sizes on the five hypotheses that we pre-specify in this document, using new data from the baseline survey when available. The outcomes, hypotheses, sample means, and standard deviations are:

1. Willingness-to-pay (WTP) for SMS-based air quality forecasts is greater than 0 regardless of the source to which the information is attributed: 89.6 (45.2)

2. Willingness-to-pay (WTP) for SMS-based air quality forecasts is differentially affected by treatment: 89.6 (45.2)

3. Absolute error of incentivized $t + 1$ forecast of PM2.5 concentration, divided by the truth, is differentially affected by treatment: 0.72 (0.42)

4. Perceived accuracy of air-quality information source as the absolute error of incentivized guess of the SMS's forecast is differentially affected by treatment: N/A

5. the amount out of PKR 100 donated to a government agency for an environmental cause, as opposed to the citizen's group, is differentially affected by treatment: 50.1 (15.0)

For hypotheses 1. and 2., we use the sample statistics from Ahmad et al. (2022) as we do not collect these outcomes in the baseline of this study. We do not have relevant statistics available from either the baseline or from Ahmad et al. (2022) for hypothesis 3., but we expect the outcome variable for it to have a similar distribution to the one for hypothesis 3..

We find that we are able to detect impacts of 0.43 standard deviations, which equals PKR 19.4 in the willingness to pay (for hypothesis 2.), 0.18 for hypothesis 3., and 6.4 for hypothesis 5.. For the test of means for hypothesis 1., we find that we are powered to detect that willingness to pay is greater than PKR 3.6.

Although the minimum detectable impact is fairly large in terms of standard deviations, the treatment effect sizes are relatively small in the outcomes' units. Furthermore, there are several reasons why our assumptions may not hold, or statistical precision could be improved. First, we plan to improve precision by including controls selected via a double-post-selection method using LASSO. Assuming a 30-percent reduction in standard errors, the minimum detectable effects would be 0.30 standard deviations. Second, the willingness-to-pay statistic from Ahmad et al. (2022) may be outdated after two years of high inflation.

# D Alternative identification strategies

## D.1 Treatment on the treated

We define takeup of our intervention as looking at our forecasts via the SMS, which we do not observe. Instead, we construct a proxy of this measure from the endline survey, where we ask, "[during] the service period, how many days out of the week did you read the message?" This question is asked to everyone in the sample as we send SMS forecasts to both treatment arms (i.e., no pure control group). We denote the number of days a subject $i$ reports to have seen the SMS as $R_i$. We code "not sure" and "refused to respond" as $R_i = 0$. A subject's takeup is $P_i = \frac{R_i}{7}$, i.e., the fraction of forecasts respondents report to have seen. We acknowledge that $R_i$ is likely measured with error and that the reported value may depend on the salience of the SMS forecasts and other factors that may be influenced by treatment. As such, we interpret $R_i$ as a measure of attention to the SMS forecasts, which we exogenously vary.

The treatment-on-the-treated (TOT) effects is estimated using 2SLS, with $Z_g$ or $\boldsymbol{A}$ instrumenting for $P$. We present the following first and second-stage specifications for a within-subject model with $Z_g$ as an instrument.

$$P_{Ti} = \eta_T + Z_g'\phi_T + \nu_T Y_{0i} + \boldsymbol{X_i'\theta_T} + \upsilon_{Ti}$$

$$Y_i = \alpha + \widehat{P'}\beta + \gamma Y_{0i} + \boldsymbol{X_i'\delta} + \varepsilon_i$$

$\widehat{P}$ is the instrumented "takeup." Much of the rest of the specification and testing remain the same as in the ITT; we include the same set of controls in the first- and second-stage regressions and carry out two-sided tests on the same set of outcomes. The between-subject models are analogous to the equations above, except for the latter in which we omit $\nu_T Y_{0i}$

and $\gamma Y_{0i}$.

## E Secondary outcomes

We present other variables that are of interest but for which we do not correct for multiple testing. We first present outcomes that are alternative definitions of, or otherwise related to, the primary outcomes. We then list other complementary outcomes.

1. Demand for air quality information (related to Primary Outcome 1.)

   1.1. Stated satisfaction with the SMS service

   - The outcome is defined as the Likert scale, with five being the most favorable. We ask respondents to rank their overall satisfaction with the SMS forecast service in the past three months.

   1.2. Stated belief in the reliability of SMS forecast service

   - The outcome is defined as the Likert scale, with five being "strongly agree." We ask respondents if they agree with the statement that the SMS forecasts have been provided frequently and on time.

   1.3. Approval of government and citizens' group's air quality information service

   - The outcome is defined as the Likert scale, with five being "strongly agree." We ask respondents if they agree with the statement that they approve of the job EPD or PAQI, respectively, does to address air quality in Lahore.

   1.4. Stated belief in the reliability of government and citizens' group's air quality information

   - The outcome is defined as the Likert scale, with five being "strongly agree."

We ask respondents if they agree with the statement that EPD or PAQI, respectively, provide air quality measurements frequently and on time.

1.5. Access to other forms of air quality information

- The outcome is defined as the number of air quality information sources the respondents have accessed in the past.

2. Policy preferences and collective action for air quality (related to Primary Outcome 4.)

2.1. Prefers the local government to invest in air quality vs. other policies

- The outcome is defined as 1 if they prefer the government invest in air quality v.s. other policy goals. We ask a hypothetical scenario in which the local government has PKR 100 million to allocate either towards improving air quality or towards investing in one of three other goals (education, health, and waste management, in three separate scenarios).

2.2. Takes a document on how to file a complaint to the local government

- The outcome is defined as 1 if the respondent takes a pamphlet. At the end of the endline survey, we prompt the respondent that EPD is a government agency responsible for addressing air quality issues in Lahore. We tell the respondents that we have a document that shows them how to file a complaint to the EPD and ask if they would like a copy.

2.3. Plans to file a complaint to the local government about air quality

- The outcome is defined as 1 if a respondent intends to file a complaint to the EPD about air quality.

3. Beliefs about air quality levels (related to Primary Outcome 2.)

3.1. Unincentivized guesses of air quality in comparison (endline only)

- 1 if correctly guesses that yesterday's air quality is better than the day before yesterday.
- 1 if correctly guesses that today's air quality is better than yesterday.

3.2. Number of days with satisfactory air quality

- The outcome is defined as the number of days in the last week with satisfactory air quality. What would constitute "satisfactory" air quality is subjective and is left to the respondents' interpretation.

3.3. Concern about air quality

- The outcome is defined as the Likert scale, with five being "strongly agree". We ask respondents if they agree with the statement that they are "concerned about air quality in general" in the last week.

4. Perceived accuracy of air-quality information source (related to Primary Outcome 3.)

4.1. Weight put on a government reading in a hypothetical scenario

- The outcome is a continuous value between 0 and 1, indicating the weight the respondents put on an EPD reading as opposed to a PAQI one. We present a hypothetical scenario in which there are readings of the PM2.5 concentration from two sources: government (EPD) and citizens' group (PAQI). One of the sources (chosen at random) is $50\mu g/m^3$, and the other is $100\mu g/m^3$. We then ask the respondent what they think the true concentration level is, between 50 and 100. We then construct $(\frac{|V_g - V_r|}{50})$, where $V_g$ is the value assigned to EPD and $V_r$ is the respondent's guess on the truth. This is a hypothetical scenario, and the game is not incentivized.

4.2. Stated belief in the accuracy of the SMS forecasts

- The outcome is defined as the Likert scale, with 5 being "strongly agree". We ask respondents if they agree with the statement that the SMS forecasts we have provided in the past three months are accurate. We make the experimentally assigned source salient by reminding them that the forecast is built using data from the said source.

4.3. Stated belief in the accuracy of government and citizens' group's air quality information

- The outcome is defined as the Likert scale, with five being "strongly agree". Aside from their beliefs in the accuracy of the SMS forecasts, we ask respondents if they agree with the statement that air quality measurements published by EPD or PAQI, respectively, are accurate.

5. Avoidance behaviors

5.1. Outdoor time use

- The outcome is defined as the number of hours spent outside. We ask respondents the type of activity (sleep, paid work, homemaking, leisure, travel, and other) they conducted for each hour of the previous day and whether it was indoors or outdoors. We aggregate the number of hours the respondent engaged in any outdoor activity.
- We also plan to estimate the impact by the type of outdoor activities (sleep, paid work, homemaking, and leisure), as well as the share of each type of activity spent outdoors. We conduct analysis on these outcomes to identify mechanisms but do not adjust for multiple testing.

5.2. Access to high-quality masks

- The outcome is 1 if the respondent shows a high-quality mask to the enumerator. We ask if the respondents have been given or purchased any masks for air pollution, and if so, to show one to the enumerator. We identify respondents who show an N90/95 mask. We also collect information on what other types of masks (e.g., surgical masks, cloth) the respondents show.

5.3. Adjust their time use because of air pollution

- The outcome is defined as the Likert scale, with five being "strongly agree." We ask respondents if they agree with the statement that they reduced the number of hours worked significantly in response to poor air quality.
- We also ask respondents how many hours they would have spent outdoors if the pollution level was, hypothetically, 150 on average. This is asked after we measure how many hours they usually spend outdoors on a typical day and is meant to capture behavior changes, if any, due to poor pollution levels.