How Do Gig Workers Respond to Unanticipated Income Shocks? Evidence from Food Delivery Riders in China

Yuxiao Fu Dandan Zhang*

Abstract: This paper studies how gig workers adjust their labor supply in response to unanticipated negative income shocks, and what these responses reveal about the underlying behavioral foundations of labor supply. Using transaction-level administrative data from China's second-largest food delivery platform, we examine how delivery riders react to transitory income losses caused by low ratings, complaints and cancellations based on a discrete stopping-choice framework that models both extensive and intensive margins of labor supply within work shifts. The results show that unanticipated shocks induce significant loss-aversion effects among gig workers, while riders' responses to anticipated fluctuations are fully consistent with predictions of the neoclassical model. Using historical complaint and low-rating data from merchants and customers as instruments, the results continue to show a significant extension in labor supply. Further analyses document that shock responses are strongest when losses occur recently and when riders are close to their income targets, suggesting dynamically adjusting reference points. Heterogeneity analyses indicate that full-time and less-experienced gig riders are particularly sensitive to negative shocks. These findings carry implications for platform incentive design and labor regulation in gig work environments.

Key words: Gig economy, Labor supply, Neoclassical model, Reference-dependent preferences, Online delivery

* Fu: Peking University. yxfu2023@nsd.pku.edu.cn; Zhang: Peking University. ddzhang@nsd.pku.edu.cn.

1 Introduction

The ongoing digital transformation has dramatically expanded China's gig economy, fundamentally reshaping the organization of labor. Today, an estimated 200 million individuals in China operate under flexible employment arrangements, including over 13 million food delivery riders who supply labor through digital platforms. Characterized by high flexibility in working hours and algorithmic, task-based compensation, this sector's rapid rise makes the determinants of gig-worker labor supply both a fascinating economic puzzle and a pressing policy question. Understanding how these flexible workers respond to high-frequency wage fluctuations is critical for optimizing taxation, designing welfare programs, and structuring platform incentives.

At the heart of understanding gig-worker behavior is a lively and protracted debate in labor economics: do flexible workers follow the textbook neoclassical model, or do they exhibit reference-dependent, income-targeting behavior? Under classical labor-supply theory (Becker, 1965; Lucas & Rapping, 1969), workers are modeled as intertemporal utility-maximizers. Because transitory wage increases raise the opportunity cost of leisure, the substitution effect should dominate, yielding a positive labor-supply elasticity. Early empirical studies of self-scheduled workers, such as stadium vendors (Oettinger, 1999), supported this standard view. However, a profound empirical anomaly emerged in the seminal study of New York City taxi drivers by Camerer et al. (1997) and Chou (2002), who documented a negative daily wage elasticity. Drivers appeared to quit earlier on high-wage days and work longer on low-wage days. Drawing on prospect theory (Tversky & Kahneman, 1979), researchers hypothesized that workers set a daily "target income": they experience sharp disutility when falling short of the target, and diminished marginal utility once the target is reached.

This behavioral proposition prompted a wave of methodological and theoretical advancements. Farber (2005, 2008, 2015) challenged the income-targeting hypothesis by employing discrete-choice hazard models on granular trip-level data. He demonstrated that cumulative hours worked, rather than cumulative income, strongly predicted the hazard of quitting, concluding that driver behavior largely conforms to neoclassical predictions. To reconcile these conflicting findings, Kőszegi & Rabin (2006) introduced a rational-expectations framework, highlighting a crucial theoretical insight: distinguishing between anticipated and unanticipated shocks is the fundamental key to testing these competing theories. They posited that if wage fluctuations are anticipated, they are endogenized into the worker's reference point, resulting in neoclassical labor-supply behavior; however, if wage fluctuations are unanticipated, they trigger reference-dependent loss aversion. Crawford & Meng (2011) successfully operationalized this insight into a dual-reference-point model, showing that workers evaluate their labor supply against rational-expectation targets for both income and hours. More recently, the frontier of this literature has shifted toward dynamic adaptation. Thakral & Tô (2021) revealed a powerful "recency effect", showing that quitting decisions are heavily influenced by wage shocks in the immediately preceding hours, rather than early in the shift, suggesting that reference points update dynamically.

Despite this rich theoretical evolution, the empirical evidence remains overwhelmingly concentrated in the taxi and ride-hailing sectors of developed economies. While recent experimental studies have begun to explore developing contexts, such as Ma et al. (2024) and Crawford et al. (2025), who elicited expectations among Shanghai drivers, and Andersen et al. (2025), who examined Indian street vendors, no prior study has investigated the labor supply of food-delivery riders. Food-delivery riders share similar job characteristics with taxi drivers, including highly flexible schedules and high-frequency order acceptance. However, the unique nature of their work provides a highly advantageous setting for testing the Kőszegi & Rabin (2006) theoretical dichotomy. On one hand, because meal consumption follows a highly regular daily schedule, the demand fluctuations delivery riders face are inherently more predictable than traditional ride-hailing. Consequently, their responses to these routine, anticipated market variations are more likely to closely align with neoclassical intertemporal optimization. On the other hand, the physical delivery workflow is inherently more complex, involving merchant coordination, navigation constraints, and customer interaction, which exposes riders to a higher frequency of sudden, unforeseen operational disruptions. This duality makes it uniquely possible to isolate the impact of unanticipated factors, thereby providing a robust foundation for cleanly distinguishing between neoclassical and reference-dependent behaviors.

To exploit this setting, we address two fundamental identification challenges that have plagued the literature. First, as highlighted by Thakral & Tô (2021), traditional empirical strategies relying on cumulative income suffer from a mechanical correlation with cumulative hours and labor intensity. Because the vast majority of cumulative income is anticipated and directly tied to hours worked, it is extremely difficult to isolate the true behavioral response to an income shock from the natural fatigue of working longer hours. Second, while recent frontier research has ingeniously exploited unexpected events, such as passenger cancellations or no-shows (Duong et al., 2023), these events inherently involve a "wasted time" component. The driver has already expended search time and driving effort (incurring labor disutility) but receives no financial return, consequently, the income loss is inextricably confounded with the frustration and physical cost of uncompensated time, making it difficult to isolate the pure mechanism of reference dependence.

We overcome these hurdles by leveraging a unique feature of the food-delivery environment: financial penalties triggered by low customer ratings, complaints, and rider-caused cancellations, a pure direct income shock. The rider has fully completed the delivery, expending the exact same physical time and effort as they would for a successful order, yet their earnings are unexpectedly slashed post-completion. Because these financial penalties do not alter the physical time spent or the fatigue incurred on the task, they perfectly isolate the income shock from any variance in time allocation or labor disutility. To further address any residual endogeneity in the occurrence of these shocks, we also construct instrumental variables based on the historical complaint and low-rating propensities of the specific customers and merchants associated with the orders.

Employing a discrete stopping-choice framework (Farber, 2005, 2008, 2015; Crawford & Meng, 2011; Thakral & Tô, 2021; Duong et al., 2023) supplemented by an instrumental variable (IV) approach on a comprehensive transaction-level administrative dataset from China's second-largest food delivery platform, we uncover three sets of key findings.

Firstly, we document a stark behavioral dichotomy in how gig workers respond to anticipated versus unanticipated income fluctuations. When facing anticipated market dynamics, riders' behavior perfectly aligns with classical intertemporal optimization: cumulative hours significantly drive their decision to stop working, while cumulative income show no clear trend. However, when struck by unanticipated negative income shocks, riders exhibit intense reference-dependent loss aversion. Specifically, for highly flexible riders, a single bad shock significantly lowers the quitting hazard by 2.04 percentage points on the extensive margin and reduces intra-hour rest time by 1.96 minutes on the intensive margin. In absolute economic terms, this compensatory effort is equivalent to working approximately 1.02 extra hours, confirming that riders aggressively adjust both the duration and intensity of their labor to recover unexpected financial losses.

Secondly, we provide rare high-frequency evidence that reference points are dynamically adaptive rather than static. The observed labor-supply adjustment displays recency: the compensatory effects on both the extensive and intensive margins are short-lived, effectively vanishing within 60 minutes post-shock. Furthermore, this loss-aversion effect is highly sensitive to target proximity, and the catch-up response becomes substantially stronger when shocks occur later in the shift as riders physically approach their daily income targets, aligning with the dynamic adaptive reference points proposed by Thakral & Tô (2021).

Thirdly, our heterogeneity analyses reveal that institutional constraints and worker experience critically mediate these behavioral responses. The loss-driven effort response is predominantly observed among the highly flexible riders, whereas regular riders who are bound by relatively fixed schedules and monthly targets exhibit no significant reaction to sudden daily shocks.

Within the gig rider population, full-time workers and less-experienced riders respond far more aggressively to income losses. In contrast, highly experienced riders exhibit a more muted response, suggesting that accumulated experience gradually pushes workers toward classical neoclassical optimization. Additionally, shorter shifts exhibit more pronounced responses on the extensive margin, highlighting how the escalating marginal cost of time and fatigue eventually constrains compensatory behavior in long shifts.

We make three primary contributions. Firstly, we achieve a methodological progress in identification. By leveraging pure direct income shocks alongside an instrumental variable approach, we yield a relatively clean test of loss aversion. Secondly, by demonstrating that gig workers behave neo-classically toward anticipated demand but exhibit loss aversion strictly toward recent, unanticipated income shocks, we effectively reconcile the divergent neoclassical and behavioral perspectives in labor supply. Thirdly, we expand the external validity and behavioral margins of reference-dependent labor supply. Moving beyond traditional ride-hailing, we validate the theory in China's massive food-delivery sector, and capture a more granular picture of gig workers' continuous effort adjustments.

The rest of the paper proceeds as follows. Section 2 details the institutional background of the Chinese food delivery market and describes the data. Section 3 outlines our empirical strategy and hazard-based model. Section 4 presents the core results regarding rider responses to unanticipated shocks. Section 5 provides further analyses on dynamic adaptation and heterogeneity. Section 6 discusses alternative psychological interpretations and potential concerns. Section 7 concludes with theoretical implications and policy recommendations.

2 Background, Data and Preliminary Statistics

A. The Background of Chinese Delivery Sector

China's food delivery market is dominated by two major platforms: Meituan and Ele.me. As of 2024, the disclosed numbers of active riders for each platform were approximately 7.45 million

(Meituan) and 4.01 million (Ele.me). The data in this article was sourced from Ele.me, which, as of 2024 when the data was utilized, was the second-largest food delivery platform in China. Founded in 2008, Ele.me, a key component of the Alibaba Business Group, served over 21.55 million daily active users in 2023. Through its mobile application, users can conveniently place orders with merchants from the comfort of their homes and have their purchases delivered right to their doorstep by delivery riders. This service covers a wide range of products, including meals, daily necessities from supermarkets, cakes, and flowers, which has significantly enhanced the convenience of people's lives.

To better manage and coordinate delivery capacity, the platform categorizes riders into two main modes: *fixed* and *flexible*, with the *flexible* mode further divided into two subcategories: *preferred* and *ordinary gig*. Among them, *fixed* delivery riders represent the most stable labor supply for the platform, accounting for 30% of the workforce and completing 50% of the platform's orders. *Flexible* riders, making up 70% of the workforce, help the platform manage fluctuations in delivery capacity and handle the other half of its orders. The three types of riders each exhibit distinct characteristics: First, the *fixed* riders are managed by teams and operate on relatively fixed schedules, with predictable work hours and a stable routine. Second, the *preferred* riders are not bound to fixed schedules but are required to meet certain performance criteria, such as achieving a minimum weekly order quota, imposing a competitive structure that can yield higher earnings for the most productive flexible workers. Finally, the *ordinary gig* riders enjoy the highest degree of scheduling flexibility, choosing their work hours without external mandates. They may have another job, or be women who are required to babysit. When compared to the well-studied driver group in the existing literature, *fixed* riders would be more like traditional taxi drivers, and *ordinary gig* riders would be more like Uber or Lyft drivers.

Figure 1 illustrates the operational process of a food delivery platform. First, a user places an order through the platform. The platform then notifies the merchant and immediately assigns a delivery rider, prompting the merchant to begin preparing the order while the rider sets off for the store. Once the order is ready, the rider starts the delivery, transporting the items from the merchant to the customer's specified address. Upon arrival, the rider calls the customer to notify them and sends a photo of the delivered goods through the app.

[Figure 1]

B. Negative Feedbacks and Abnormal Events

The food delivery sector features specific shocks that provide an ideal quasi-natural setting for identifying anticipated and unanticipated shocks. Firstly, if customers are dissatisfied with a rider's service, such as poor attitude or spilled food, they can give the rider a negative rating, a complaint, or even cancel the order directly. Merchants can also give riders a negative rating or a complaint. Those negative feedbacks serve as the source of the unanticipated shock used in this study. After verifying the validity, the platform will impose a fine on the rider: *fixed* riders incur a penalty of 150 to 200 yuan, while *flexible* riders face fines ranging from 3 to 10 yuan (vary by city).

In addition to the fine, the platform also deducts points from the rider's *service score*. If the score falls below a certain threshold, the rider loses eligibility to deliver high-value items such as flowers, cakes, and 3C electronic products, though delivery of other ordinary orders remains unaffected. We then examine the effect of these algorithmic penalties (see Appendix Table A1), and the results show that in the hour immediately following the shock, there was no significant change in the value of products delivered by the riders. This indicates that, apart from the immediate income loss, a single negative shock does not lead to other lasting impacts on riders.

Secondly, during delivery process, riders may encounter some *abnormal events*: merchant delay, incorrect customer address, or customer lose contact. In these three scenarios, the rider can submit evidence such as photos, videos, or chat records to the platform. The platform then verifies the report through a combination of multimodal algorithms and manual review. If the report is validated, the platform extends the delivery time limit for that order by 10 minutes, and the rider is exempted from penalties for negative feedbacks related to that specific order. However, the fulfillment quality of other orders the rider is currently delivering may still be affected.

According to platform data, in the hour following an abnormal event, the probability of order delay increases significantly by approximately 2 percentage points (a relative rise of 59.5%), while the probability of receiving negative feedback rises significantly by about 2.14 percentage points (a relative increase of 98.6%). Therefore, whether an abnormal event was filed in the hour prior to a given order t can be regarded as one source of short-term anticipated negative shock. After controlling this, we consider subsequent negative feedbacks encountered by the rider as unanticipated shocks.

C. Data

This study utilizes two distinct data sources: Ele.me Rider Quarterly Survey Data and Platform Transaction-Level Administrative Data. The survey was conducted from October 21 to 28, 2024, with 12,876 valid responses collected. The survey gathers basic rider characteristics such as educational background, urban/rural household registration, local residency status, part-time engagement, insurance coverage, and future career plans that supplements variables not captured by the platform's big data, providing rich insight into the demographic and socio-economic characteristics of the riders.

From the respondents, a subsample of 930 riders was randomly selected based on a gender ratio of 5:1. We then sampled the platform's administrative data for these riders. The dataset covers all orders completed by these riders during two peak weeks (July 15–28) and two off-peak weeks (October 21–November 3) in 2024. After data matching, the final sample consisted of 917 riders and 472,045 order records. Of these riders, 322 are *fixed* riders, 140 are *preferred* riders, and 449 are *ordinary gig* riders. Order-level variables include order date, city, product type, weather conditions, whether the order was delayed, and whether an abnormal event was filed. Rider-level variables consist of rider ID, contract type, age, gender, and first delivery date. Logistics timeline variables record order acceptance, arrival at the store, pickup, and delivery times. Feedback-related variables cover negative ratings, complaints, and exceptional cancellations. This panel dataset links riders with merchants and customers, and provides a solid foundation for subsequent descriptive analysis and empirical analysis.

Based on this dataset, we adopt the concept of a "shift", which refers to a rider's uninterrupted sequence of order-taking bounded by sufficiently long breaks. This concept follows the classic unit of analysis used in gig-labor supply studies (Farber, 2005, 2008; Crawford & Meng, 2011; Farber, 2015; Thakral & Tô, 2021; Duong et al., 2022; Ma et al., 2024). In our study, a new shift is defined whenever the time elapsed between two consecutive order groups exceeds five hours (see Figure 2). We also conducted sensitivity analyses using six-hour and seven-hour thresholds in robustness checks.

[Figure 2]

Based on this shift segmentation, we compute, for each order: (i) the rider's cumulative working time and cumulative number of orders within the current shift; (ii) a binary indicator for whether the focal order is the last one in that shift; and (iii) the length of the rider's rest period in the hour immediately following order completion. These derived variables form the core inputs to our stopping-time and rest-time hazard models.

In addition, given data availability, we use the cumulative number of orders as a proxy for cumulative income, based on two considerations: First, a rider's per-order compensation consists mainly of a base delivery fee and a premium. The same rider typically operates within a relatively fixed delivery range, with fixed riders assigned to specific stations and flexible riders to delivery zones (usually 3–5 km). Consequently, the base fee remains largely consistent across orders for a given rider. The premium, which compensates for additional delivery difficulty, generally constitutes a small portion of total earnings (usually less than 10% of the base fee), resulting in limited fluctuation in average earnings per order.

Second, most platform incentives and subsidies are tied to the number of orders rather than directly to earnings. Examples include an increase in the base fee after reaching a certain cumulative order threshold (e.g., from ¥6 to ¥6.5 per order after completing 500 orders) and short-term incentives (e.g., a ¥10 bonus for completing 10 orders). Both types of rewards are triggered by number of orders rather than monetary amounts. In summary, cumulative order volume serves as a feasible proxy for measuring riders' income reference-point targets.

D. Descriptive Statistics

Figure 3 offers an intuitive visualization of delivery activity over the 336-hour window from October 21 to November 3. Several patterns stand out: first, two dense "bars" of color occur each day at lunch and dinner peak hours, showing that riders take far more orders during these periods. Second, the coloration is noticeably thicker on weekend days than weekdays, indicating higher overall order volumes on weekends. Third, the work patterns of full-time and preferred riders appear more regimented, with more consistent block lengths and timing. Finally, there is a greater prevalence of overnight delivery activity, as evidenced by the extended colored segments late at night, showing the necessity of using the concept of shifts instead of calendar days.

[Figure 3]

To further characterize riders' delivery tendency within the week, Figure 4 plots average active delivery time per rider by hour of the week, aligning corresponding hours across two consecutive weeks. Two distinct patterns emerge: first, preferred riders exhibit the highest work intensity throughout the week, followed by full-time riders, and then by ordinary gig riders, reflecting a clear ranking in average hourly workload. Second, although overall intensity is lower for ordinary gig riders, their absolute delivery minutes during overnight hours exceed those of the other groups, revealing a heavier night-shift contribution among this flexible segment.

[Figure 4]

Figure 5 presents summary statistics at the shift level, showing the distribution of shift start times and their durations. Firstly, there are three “mini-peaks” in shift commencements at 07:00, 10:00, and 17:00, with the modal start time at 10:00 accounting for nearly one quarter of all shifts. Secondly, riders who begin their shifts in the early morning hours work the longest, typically 12–14 hours, while those starting mid-morning also log longer shifts than those starting in the afternoon or evening.

[Figure 5]

Table 1 reports detailed summary statistics at three levels: rider (Panel A), order (Panel B), and shift (Panel C). From the order-level descriptive statistics, we observe that the baseline incidence rate of bad shocks is only 2.2%. Even when orders are divided into deciles based on their predicted scores from observable variables, ranging from lowest to highest predicted risk, we find the actual incidence of negative shocks in the highest-risk decile, namely the top 10 percent of orders, remains only around 7 percent. Therefore, for riders’ real-time labor-supply decisions, whether a particular order will result in a negative shock is still a highly uncertain, low-probability event, making it difficult to rely on available information to predict or avoid such shocks.

[Table 1]

From Panel C, we can see that a typical shift comprises 29.3 orders, lasts about 8.5 hours, and includes 1.2 negative shocks.

3 Empirical Strategy and Baseline Results

A. Identification Strategy

In testing labor supply patterns, we argue that delivery riders are highly sensitive to fluctuations in order volume driven by changes in supply and demand, and most variations in order volume during routine delivery are predictable to riders. In this context, observing that "cumulative working hours significantly affect riders' probability of stopping work, while cumulative income does not" is insufficient to conclude that individuals adhere to the neoclassical labor supply model. This is because riders can anticipate a considerable portion of wage fluctuations in advance and adjust their labor supply behavior accordingly.

To examine this proposition, we follow the approach of Farber (2015) and categorize observable predictable fluctuations into two types: medium-term (inter-day) predictable fluctuations, captured by fixed effects for day-of-week and week-of-year; and short-term (intra-day) predictable fluctuations, captured by fixed effects for weather severity, hour-of-day, cumulative working hours within the shift, and whether an abnormal event was filed in the hour before order acceptance.

We first estimate a regression model of riders' cumulative orders at each observation point, controlling for rider fixed effects and using the above observable variables as explanatory factors. The results show R^2 values of 0.965, 0.952, and 0.966 for fixed, preferred, and ordinary gig riders, respectively, indicating that observable variables explain nearly all order volume variation during normal delivery.

We then conduct a Shapley variance decomposition of order volume fluctuations based on the regression framework. The decomposition reveals that short-term (intra-day) predictable fluctuations account for 91.03% of the variation in cumulative orders, medium-term predictable fluctuations explain 0.47%, and only 8.50% is attributable to unobservable or unanticipated factors. This residual share is very close to, and even slightly lower than the 12.4% reported in

Farber (2015), further supporting the conclusion that most fluctuations are anticipated by riders ex ante.

Based on the evidence above, we argue that the behavioral pattern observed by Farber (2015), which appears consistent with the neoclassical labor supply model, largely stems from workers' ability to anticipate temporal fluctuations in cumulative income and flexibly adjust their reference points and labor supply accordingly. Therefore, to more effectively distinguish between the neoclassical labor supply model and reference-dependent preferences, we adopt the research design of Duong et al. (2023) and employ exogenous shocks that are difficult to anticipate as a quasi-natural experiment.

Our empirical approach exploits those negative feedbacks mentioned in section 3: negative ratings, complaints, and rider-caused order cancellations ($badshock_{it}$). In addition, if a rider has an abnormal event within the previous hour (T–1), it is treated as one source of anticipated negative shock and is captured by the variable $report_{it(T-1)}$. After controlling for cumulative hours, cumulative orders, weather, hour-of-day and day-of-week effects, and rider fixed effects, any remaining impact of $badshock_{it}$ on the probability of quitting can be attributed to the negative income shock.

Compared with Duong et al. (2023), the shock used in our study offers a cleaner identification, as it exerts an immediate and direct impact on riders' income without altering their actual working time. The shocks employed by Duong et al., such as passenger cancellations and no-shows, represent an opportunity cost for drivers: they work the corresponding time but receive no income, which simultaneously affects both their time disutility and income target. In contrast, the negative ratings, complaints, and exceptional cancellations used in our setting leave the actual labor time of delivery riders unchanged.

Formally, we estimate the following linear probability model, closely following Duong et al. (2023):

$$Y_{it} = \beta_{unp} badshock_{it} + \beta_p report_{it(T-1)} + f(h_{it}) + g(y_{it}) + X_{it}\gamma + \alpha_i + \epsilon_{it} \quad (1)$$

where i indexes riders, t indexes orders and T the corresponding hourly interval. To clarify, we examine both the extensive margin and the intensive margin. In the extensive-margin specifications, the dependent variable Y_{it} is $Quit_{it}$, equals one if rider i ends a shift after completing order t and does not accept order t+1. In the intensive-margin specifications, Y_{it} measures $Rest_{it(T+1)}$, the rider's rest time (in minutes) during the hour immediately following order t. Riders may not necessarily quit outright after a negative shock but could also respond by altering their work intensity, so considering both margins gives a more complete picture of labor-supply adjustment.

The key regressors are $badshock_{it}$, a dummy for any negative rating, complaint, or rider-caused cancellation on order t, and $report_{it(T-1)}$, a dummy for any abnormal platform report in the prior hour. For riders, $badshock_{it}$ represents direct, platform-imposed income losses and penalties. The non-parametric functions $f(h_{it})$ and $g(y_{it})$ capture nonlinear effects of cumulative hours and cumulative earnings, respectively, each via ten indicator variables for decile bins. The matrix X_{it} includes controls for the hour of day (24 hourly dummies), day of week (7 dummies), week of year (4 dummies), and five levels of weather. Rider fixed effects α_i absorb unobserved heterogeneity in preferences or expectations, and ϵ_{it} is an error term. To explore heterogeneity, we estimate the above equation separately for each rider category.

B. Testing the Unpredictability

We also conduct two tests for the exogeneity of negative shocks (bad shocks). First, we perform balance tests across rider characteristics (see Figure 6). The results show that after controlling for the aforementioned predictable effects in the regression, negative shocks at the order level are only related to the rider's contract type (fixed, preferred, ordinary gig), while the coefficients for all the other observable rider characteristics, including gender, age, urban/rural household registration, local residency, education level, and platform delivery experience, are all statistically indistinguishable from zero. Moreover, the joint significance test yields an F-statistic of only 1.63 ($p = 0.07$), indicating that, conditional on contract-type subsamples and after accounting for observable conditions, the occurrence of negative shocks exhibits no systematic observable variation across riders and can be regarded as approximately random.

[Figure 6]

Second, we implement a within-individual time-series placebo test. Under a framework that absorbs rider fixed effects, we hold constant the total number of actual negative shocks experienced by each rider within each shift, randomly reassign these shocks across that rider's sequence of orders, and then re-estimate our main regressions using this permuted set of "pseudo negative shocks." The previously observed significant effects disappear entirely (see Appendix Table A2). The results from both tests jointly suggest that, within the set of observable conditions and fixed-effect controls we employ, negative shocks can largely be treated as relatively exogenous and are shown to exert a significant impact on riders' labor supply.

C. Baseline Results

Table 2 presents the results of the baseline regression. After experiencing the shock, overall, riders across all types riders reduce their stopping probability by 1.12 percentage points and shorten their rest time in the following hour by 1.5 minutes on average. Separating by rider type, flexible riders significantly increase their labor supply on both the extensive and intensive margins, whereas fixed riders show no statistically significant response. The reaction is strongest among ordinary gig riders, whose stopping probability decreases by 2.04 percentage points and subsequent rest time falls by 1.8 minutes. Preferred riders exhibit a smaller but still significant decrease in stopping probability (0.47 percentage points) and rest time (1.4 minutes). These patterns suggest that when an unexpected shock disrupts their cumulative income trajectory, crowdsourced riders compensate by extending work hours and increasing work intensity, which displays typical loss aversion and reference-dependent behavior. In contrast, dedicated riders, whose schedules and tasks are more rigidly managed by the platform, do not show a significant willingness to increase labor supply in response to such shocks.

[Table 2]

To fully grasp the behavioral impact of these unanticipated income shocks, it is crucial to translate the estimated probability changes into an intuitive economic metric. While our baseline results demonstrate a statistically significant reduction in the quitting hazard, we further quantify the economic magnitude of this response by calculating the “equivalent hours” of labor supply induced by the shock.

Following the methodological approach of Duong et al. (2023), we estimate an alternative specification that allows for rich, continuous-duration dependence by directly incorporating cubic polynomials of cumulative hours and cumulative orders into our baseline model (reported in Table 3). Under this flexible functional form, the estimates for *ordinary gig* riders reveal that experiencing a negative income shock reduces the stopping probability by an average of 2.06 percentage points. Concurrently, the marginal effect of working one additional hour increases the stopping probability by an average of 2.02 percentage points. By taking the ratio of these two marginal effects, we can benchmark the shock against the natural fatigue of working, showing that the absolute behavioral effect of a single negative income shock is equivalent to working approximately 1.02 extra hours.

This magnitude is both economically substantial and remarkably consistent with the existing literature. For context, Duong et al. (2023) report a corresponding estimate of 1.21 equivalent hours for ride-hailing drivers in Singapore facing passenger no-show and cancellation shocks. The fact that our estimate for Chinese food-delivery riders is so close in magnitude provides strong external validity to our findings and the reference-dependent framework.

[Table 3]

D. Instrumental Variable

To address endogeneity concerns that time-varying rider-level factors, such as health conditions or family circumstances that could simultaneously affect both income targets and negative shock exposure, we employ an instrumental variables strategy. We construct two order-level instruments: customer-specific risk ($CustRisk_{it}$), measured as the total number of low ratings or complaints issued by the customer toward riders in the past year, and shop-specific risk ($ShopRisk_{it}$), defined as the total number of low ratings or complaints issued by the shop toward riders during the same period. These instruments capture exogenous, order-specific sources of shock risk that are plausibly orthogonal to unobserved rider-level time-varying confounders, while being strongly correlated with the incidence of rider-side negative shocks.

We first assess the exogeneity of the instrumental variables by examining their correlation with observable covariates. The coefficients range from -0.04 to 0.02, indicating no strong systematic relationships. We then test their relevance, and the first-stage F-statistics, reported in Table 4, reject the weak instrument hypothesis for all subgroups except preferred riders.

The results in Table 4 show that for the full sample of riders, after instrumenting, the probability of quitting work decreases significantly by 71.95 percentage points. Among those who do not quit, rest time in the following hour is significantly reduced by 33.36 minutes. Specifically, ordinary gig riders experience a significant reduction in quitting probability of 31.77 percentage points, and among those who continue working, their rest time in the next hour decreases significantly by 19.63 minutes.

[Table 4]

E. Effects of Anticipated Factors on Stopping

We also empirically test whether the influence of predictable factors on labor supply aligns with the predictions of the neoclassical model. First, regarding the effects of cumulative working hours and cumulative income, we follow Farber (2015) by incorporating the interaction terms $f(h_{it})$ and $g(y_{it})$ from equation (1) into the regression. The results on the extensive margin are fully consistent with those of Farber (2015). Figures 7a and 7b present the regression coefficients for flexible riders and fixed riders, respectively. They visualize how a rider's probability of ending a shift is shaped by the interaction of cumulative hours (vertical axis) and cumulative orders (horizontal axis).

The results show that for flexible riders, the probability of stopping work increases significantly with cumulative hours, while the coefficient on cumulative orders shows no clear trend. For fixed riders, due to schedule constraints that impose a minimum work-hour threshold, neither cumulative hours nor cumulative orders have a significant effect on stopping probability until cumulative hours exceed 10. After reaching 10 hours, the stopping probability increases slightly but still significantly with further cumulative hours, while the coefficient on cumulative orders still shows no clear pattern.

Using unexpected shocks, we find that riders exhibit significant characteristics of reference-dependent preferences. However, when testing using cumulative income and cumulative hours, we obtain results consistent with the neoclassical model. This further supports the argument presented in subsection A: routine income fluctuations are highly likely to be anticipated ex ante by those workers.

[Figure 7a, Figure 7b]

Additionally, we examine the impact of other anticipated income shocks, such as the occurrence of abnormal reports in the previous hour, weather conditions, and day of the week, and find that riders generally exhibit responses consistent with neoclassical labor supply theory (see Table

5). In the face of anticipated negative income shocks, represented by an abnormal report in the preceding hour, the stopping probability of preferred riders and ordinary gig riders increases significantly by 0.18 and 0.87 percentage points, respectively, indicating a tendency to cease work immediately to avoid further exposure to low returns or high penalty risks. Regarding anticipated positive income shocks, riders tend to increase their labor supply. For example, compared with normal weather conditions, under Level-1, Level-2, and Level-3 severe weather, the stopping probability of ordinary gig riders decreases significantly by 0.65, 2.04, and 2.26 percentage points, respectively, while their rest time in the following hour is reduced significantly by 1.08, 3.18, and 3.62 minutes, respectively. Similarly, compared with Monday, on high-demand days such as Friday, Saturday, and Sunday, all three rider types significantly increase their labor supply on both the extensive and intensive margins, whereas the regression coefficients for Tuesday, Wednesday, and Thursday are not statistically significant.

[Table 5]

4 Further Analysis

A. The Persistence of Shock Effects

To test whether riders' reference points are dynamic, we examine in Table 6 how the timing of a negative income shock shapes their subsequent labor supply. For ordinary gig riders, responses on the extensive margin show clear recency dependence: shocks occurring within 0–20 minutes reduce the stopping probability by 1.07 percentage points, whereas shocks older than 20 minutes have no significant effect. This pattern aligns with reference-dependent models in which reference points adapt quickly to earlier shocks (Thakral and Tô 2021).

On the intensive margin, negative shocks reduce rest time for all rider types, with the absolute magnitude of the coefficients decaying over time and eventually becoming statistically insignificant. However, the pace and persistence of this decay differ. Fixed riders exhibit only a short-lived effort increase: the effect is −0.0459 in the first 20 minutes and becomes insignificant beyond 40 minutes. In contrast, preferred riders show a more gradual and persistent adjustment, with significant reductions in rest time lasting 60–80 minutes after the shock. Ordinary gig riders display an intermediate decay pattern, with coefficients of −0.0540 within 0–20 minutes and −0.0346 at 40–60 minutes. The prolonged response among preferred riders suggests slower updating of reference points, consistent with their hybrid contractual structure, which blends platform flexibility with schedule commitments. Together, these decay gradients, which vanish within an hour for fixed riders but persist longer for gig and preferred riders, support the view that effort reallocation is guided by reference-dependent mechanisms, with the speed of adaptation moderated by institutional incentives.

[Table 6]

B. Heterogeneity Analysis

Firstly, for the heterogeneity of time slots, we augment the discrete stopping specification by interacting the transitory negative shock indicator with riders' cumulative hours worked within a shift to test whether proximity to an intra-shift income target conditions behavioral response. Specifically, we estimate the effect of a bad shock on both extensive and intensive margins while including an interaction term between the shock and a set of cumulative-hours indicators in Table 7. The interaction coefficients are negative and statistically significant: as cumulative hours rise, i.e., as riders approach their within-shift earnings target, the inhibitory effect of a bad shock on stopping becomes stronger and the reduction in subsequent rest time becomes larger. In other words, riders are less likely to quit and take shorter rests after a negative income shock when they have already worked many hours in the shift, consistent with target-seeking behavior and reference-dependent loss aversion. This pattern is most pronounced for ordinary gig riders.

[Table 7]

Secondly, regarding heterogeneity in rider characteristics, we conduct subgroup analyses for riders of different types across three dimensions: full-time status, experience level, and gender. The results are presented in Table 8a and 8b. For fixed riders, on the extensive margin, female riders show a higher probability of stopping work compared with male riders. On the intensive margin, however, the opposite pattern emerges: female riders who continue working increase their effort more substantially. We interpret these results as follows: the female riders who exit on the extensive margin may have weaker reference-point dependence, while those with stronger reference-point adherence may face constraints, such as family responsibilities, that limit their ability to extend working hours freely. Consequently, they compensate by intensifying their labor supply within the available time.

Table 8b reveals significant heterogeneity in labor supply responses to bad shocks across ordinary gig rider subgroups. For definition, "scheme" reflects riders' self-reports on the survey, with full-time indicating exclusive work for Ele.me and no other platform or job, and "experience" denotes whether a rider's tenure lies above or below one year.

Full-time riders exhibit a substantially stronger shift continuation response: the interaction term Badshock×Fulltime registers -0.0166 on the extensive margin and -0.0170 on the intensive margin. Part-time riders display weaker responses. Furthermore, greater experience attenuates the effects, and experienced riders demonstrate a muted response than unexperienced riders on both extensive and intensive margins, suggesting experienced riders flexibly adjust reference points and behave closer to neoclassical predictions. Preferred-rider status exhibits no robust heterogeneity.

[Table 8a, Table 8b]

Thirdly, regarding the heterogeneity at the shift level, we present in Figure 8 the heterogeneity in the effect of bad shocks on riders' stopping decisions, conditional on the total number of hours in a shift. For each total-hour category we restrict the sample to shifts that end within a

30-minute window centered on that hour. The key insight from this figure is that the behavioral response to bad shocks is strongly moderated by shift length: riders working shorter shifts are more likely to stop in response to a bad shock, while those working longer shifts exhibit smaller or even null effects. Specifically, when shifts last only 2 to 5 hours, a bad shock raises the likelihood of stopping by approximately 1.5 to 2 percentage points (relative to a baseline stopping probability of 5.5% to 7.5%), representing a 25% to 35% increase, and these effects are statistically significant at the 5% level. By contrast, for shifts of 11–13 hours, the estimated change in stopping probability is close to zero. This declining sensitivity aligns with a reference-dependent model in which the utility cost of a bad shock is more salient when time disutility is lower.

[Figure 8]

C. Impact on Earnings

Extending our analysis beyond next-hour rest, we also examine riders' completed-order counts in the following hour and observe an upward trend for full sample and flexible rider cohorts (see Table 9). In particular, preferred riders increase their order volume by a statistically significant 0.283 orders, a finding that dovetails with reference-dependent preference pattern. This outcome also demonstrates that there is no significant algorithmic penalty, and the riders' labor supply efforts continue to translate into returns.

[Table 9]

5 Robustness Checks

To assess the sensitivity of our baseline estimates, we implement four robustness checks. First, we change the interval for shift divisions to 6 hours and 7 hours to test the sensitivity. Second, we replace our original dummy controls for cumulative order count and cumulative hours with more flexible functional forms like Duong et al. (2023). Third, we augment our temporal controls by adding fully interacted hour-of-day × weekday fixed effects, which absorb even finer-grained intra-week demand fluctuations. Fourth, to ensure that our extensive-margin results do not hinge on the choice of linear-probability specification, we re-estimate the stopping equation as a logit model. In each case, the estimated impact of bad shocks on riders' labor supply decisions remains quantitatively similar, confirming the robustness of our core findings.

A. Sensitivity Check for Shift Divisions

To demonstrate that our definition of a shift does not drive the observed labor supply responses, we perform a sensitivity analysis by redefining the shift threshold as six and seven hours between consecutive order groups. While the number of shifts decreases under these alternative thresholds, the baseline regression coefficients (see Appendix Table A3a and A3b) remain largely unchanged, confirming the robustness of our findings.

B. Add Fully Interacted Fixed Effects

In our baseline specifications, we include fixed effects for hour of day and weekday to flexibly absorb systematic intraday and day-of-week variation in market conditions and worker availability. To mirror the identification strategy of Farber (2015), we also estimate an alternative model that replaces these unidimensional controls with the full hour-of-day × weekday interaction, thereby allowing each hour of each day of the week to have its own baseline hazard. As reported in Appendix Table A4, the core coefficients remain stable. This exercise confirms that our main findings are not driven by mis-specification of time trends, but instead reflect genuine behavioral responses.

C. Change to Logit Model

In the taxi-driver labor-supply literature, the large sample sizes typically justify the use of linear probability models for binary stopping decisions. Nonetheless, to more appropriately model the dichotomous nature of the outcome and enforce probability bounds, we re-estimate our baseline specification using a logit model. As shown in Appendix Table A5, the logit estimates closely mirror our original findings: a bad shock reduces the probability of stopping for preferred riders by 35.1% (significant at the 5% level) and for gig riders by 41.9% (significant at the 1% level). These results demonstrate that our key inference, the sharply negative impact of bad shocks on quitting behavior, is robust to adopting a nonlinear, binary-choice framework.

D. Robustness to Overlapping Orders

Unlike traditional ride-hailing drivers who typically serve one passenger at a time, food-delivery riders frequently engage in "order batching", executing multiple overlapping orders simultaneously. Consequently, a negative shock to one specific order, or the decision to end a shift, likely affects the rider's mindset and pacing across all concurrent orders, rendering the exact chronological sequence of individual order completions within a batch somewhat arbitrary. To ensure our baseline estimates are robust to this intra-batch spillover, we maintain our transaction-level dataset but redefine our key indicators for overlapping order groups.

Specifically, we identify groups of temporally overlapping orders. If any individual order within a concurrent group experiences a negative income shock, we conservatively assign the shock indicator a value of 1 for all individual orders comprising that group. Similarly, for the extensive margin, if a rider terminates their shift after completing their final group of concurrent orders, the quitting indicator is uniformly set to 1 for every order within that terminal group. Re-estimating our discrete stopping-choice models under this adjusted order-level specification yields results that are virtually robust (see Appendix A6). This confirms that our identification of reference-dependent labor supply is not an artifact of the rider's complex mechanism of simultaneous order fulfillment.

6 Discussion and Interpretation

While our empirical findings consistently align with the predictions of reference-dependent preferences and income targeting, identifying the exact psychological drivers of labor supply using administrative data inherently poses challenges. Specifically, one might argue that a negative shock, such as a customer complaint or a low rating, does not merely reduce a rider's income, but also inflicts a significant emotional toll. In this section, we discuss two potential mood-based alternative hypotheses: the "demotivation" (or mood loss) hypothesis and the "happy ending" (emotional closure) hypothesis. We also address potential concerns regarding the exact timing of when riders become aware of these penalties. Although we cannot perfectly observe the inner psychological states of gig workers, we make the greatest possible effort to interpret the available empirical and qualitative evidence to evaluate these alternatives.

The first alternative hypothesis posits that negative interactions with customers or merchants impose a psychological cost, essentially shocking the rider's mood. Under standard economic intuition, such emotional frustration should increase the marginal disutility of labor, making further work feel significantly more tedious or painful. If this were the primary mechanism, we would expect riders to exhibit a rage quit response by ending their shifts earlier than usual, or at least to exhibit passive behavior by taking longer breaks to recover emotionally. However, both the extensive and intensive margins of our data firmly contradict this directional prediction. Instead of quitting early or slacking off, riders who experience a bad shock significantly delay their log-off times, compress their intra-hour rest periods, and increase their per-hour order completion rate. This aggressive, accelerated work pace directly opposes the behavioral profile of a demotivated worker. Furthermore, our heterogeneity analysis reveals that the impact of bad shocks grows as cumulative time worked within a shift increases, and that the magnitude of this effect differs across shifts of varying total duration. While a sudden bad mood should theoretically affect a worker regardless of the time of day, reference-dependent models explicitly predict that loss-aversion effects will intensify precisely as workers approach their daily income targets.

A second, related psychological interpretation is the "peak-end rule" or the desire for emotional closure. Under this "happy ending" hypothesis, a rider might decide to complete more orders to wash away the frustration of a complaint, aiming to end their workday on a positive note rather than actively targeting a specific income level. However, three key pieces of evidence make this explanation highly unlikely. First, as estimated in Section 3, a single negative shock induces 1.02 extra hours of work. If riders merely sought a psychological happy ending, the marginal emotional benefit of a single positive interaction would be highly unlikely to offset the substantial time disutility associated with such prolonged effort. Second, while the desire for emotional closure might predict a delay in quitting on the extensive margin), it offers no theoretical basis for why a rider would alter their pacing between orders on the intensive margin. This localized rushing behavior reflects a calculated time-money tradeoff to aggressively recover a financial deficit, which cannot be explained by a passive wait for a pleasant final interaction. Finally, this response is heavily concentrated among the highly flexible riders. If the response were a universal human desire for emotional closure, it should manifest across all rider types. The stark contrast suggests that the response is fundamentally an economically motivated, target-driven decision facilitated by institutional flexibility.

A separate but equally important concern pertains to the timing of information. One might question whether customers and merchants file complaints immediately, and consequently, whether riders are instantly aware of the financial penalty. We acknowledge that our administrative dataset records the timestamp of the problematic order but does not definitively capture the exact moment the platform pushes the penalty notification to the rider's application. Nevertheless, two pieces of evidence strongly suggest that riders are immediately aware of the negative shock. First, our dynamic recency analysis demonstrates that *ordinary gig* riders react intensely and specifically to shocks that occurred within the past 20 minutes. If there were a significant delay in information transmission, we would not observe such a precise, immediate, and localized adjustment in their labor supply. Second, supplementary qualitative interviews conducted with a subset of riders corroborate this finding. Interviewees frequently reported that they can instantly sense a looming complaint or bad rating based on hostile verbal exchanges or the poor attitude of the customer or merchant during the handover of the meal. Thus, even if the formal deduction is processed slightly later, the rider's expectation of the income loss is updated immediately.

7 Conclusions

This paper investigates how gig workers adjust their labor supply in response to high-frequency wage fluctuations, aiming to reconcile the divergent neoclassical and reference-dependent theories of labor supply. Using transaction-level administrative data from China's second-largest food-delivery platform, we exploit financial penalties triggered by customer complaints, low ratings, and rider-caused cancellations, as unanticipated negative income shocks. By employing a discrete stopping-choice model alongside intensive-margin regressions, we isolate the impact of these shocks from the mechanical fatigue of cumulative hours. This approach effectively overcomes the confounding "wasted time" and "uncompensated effort" effects prevalent in prior literature, yielding relatively clean causal estimates of reference-dependent behavior.

Our findings provide robust evidence of loss aversion among flexible gig workers. For the full sample, an unanticipated negative shock reduces the quitting hazard by 1.12 percentage points and shortens subsequent rest time by 1.5 minutes. This compensatory effort is predominantly driven by highly flexible gig riders, for whom a surprise penalty diminishes the probability of ending a shift by 2.04 percentage points and shrinks inter-order breaks by 1.8 minutes. Importantly, we find that riders' reactions to these sudden penalties are short-lived, showing that their reference points adjust dynamically over time. The effect on their decision to quit and their tendency to take shorter breaks fades after an hour. Moreover, riders work much harder to make up for the lost income when the shock happens closer to their daily earnings targets. In contrast, when dealing with predictable factors, like how many hours they have already worked, riders behave exactly as the standard neoclassical model predicts.

Heterogeneity analyses further reveal that outside options and worker experience critically mediate these behavioral responses. Full-time and less-experienced gig riders exhibit the most pronounced loss-aversion behaviors. In contrast, highly experienced riders display more muted responses, suggesting a learning process that gradually aligns worker behavior with classical intertemporal optimization.

Ultimately, our findings carry profound implications for algorithmic management and labor regulation in the gig economy. From a worker-welfare perspective, the intense catch-up behavior induced by unanticipated shocks, such as skipping breaks and extending shifts, can significantly exacerbate physical fatigue and elevate road safety risks. Crucially, because qualitative evidence suggests that riders immediately perceive negative interactions even before formal platform deductions are processed, simply delaying penalty notifications or offering grace periods is insufficient to prevent this compensatory overwork.

To mitigate these risks, platforms should rethink the structural design of algorithmic penalties. While flexible and fixed-schedule riders naturally self-select into their respective roles based on differing preferences, their contrasting responses highlight a key institutional insight: longer evaluation horizons may effectively dampen daily behavioral volatility. Platforms could transition from immediate cash deductions to aggregated, rolling-period metrics to dilute the acute sting of sudden shocks. Additionally, dispatch algorithms could facilitate safe financial recovery by temporarily routing orders with generous delivery windows to riders who recently experienced a penalty.

For policymakers and regulators, these findings suggest that effective gig economy regulation must address the psychological and behavioral impacts of algorithmic pay structures. Rather than solely focusing on maximum working hours, which target-driven gig workers may attempt to bypass, regulators should consider capping the maximum allowable daily financial penalties to limit extreme income volatility. Furthermore, occupational safety guidelines could mandate platforms to integrate fatigue-management protocols that detect and dynamically intervene during intense target-chasing spells, ensuring that algorithmic efficiency does not compromise public safety or worker well-being.

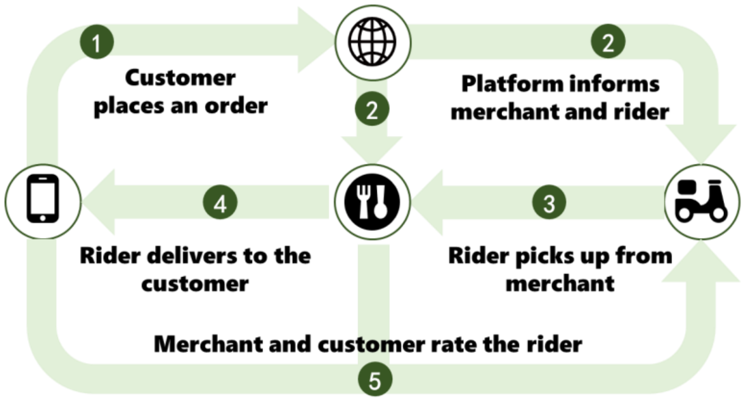

Figure 1: Delivery process

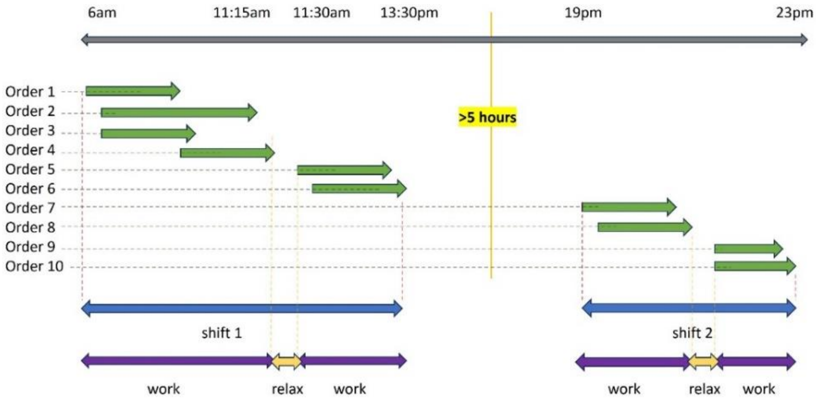

Figure 2: The division of shifts

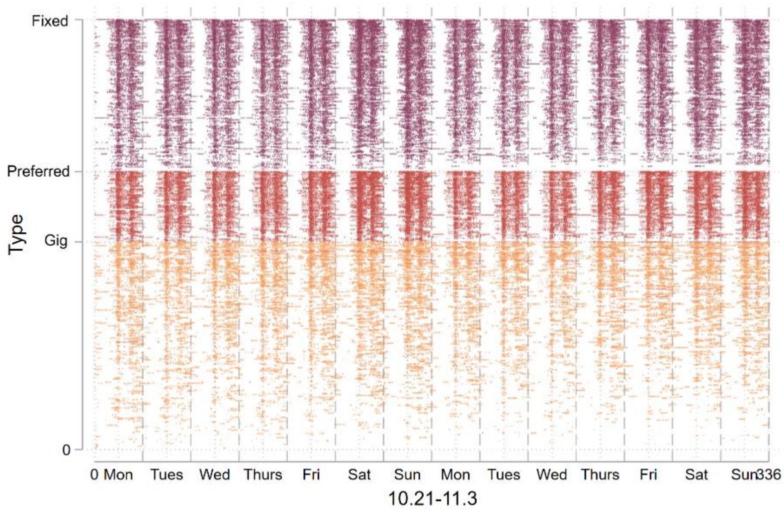

Figure 3: Net delivering time for 930 riders during the two-week off-season by type

Notes: The heat map visualizes active delivery periods (colored segments) vs. inactivity (blank gaps) over a 336-hour window from 10.21-11.3. Each horizontal line corresponds to a single rider, with colored segments denoting periods when that rider is actively fulfilling orders and blank gaps indicating inactivity. Along the vertical axis, riders are grouped into three categories, and within each group are sorted top-to-bottom by their total orders across the two weeks. Vertical gray dashed lines mark daily midnight, while lighter dashed lines mark noon.

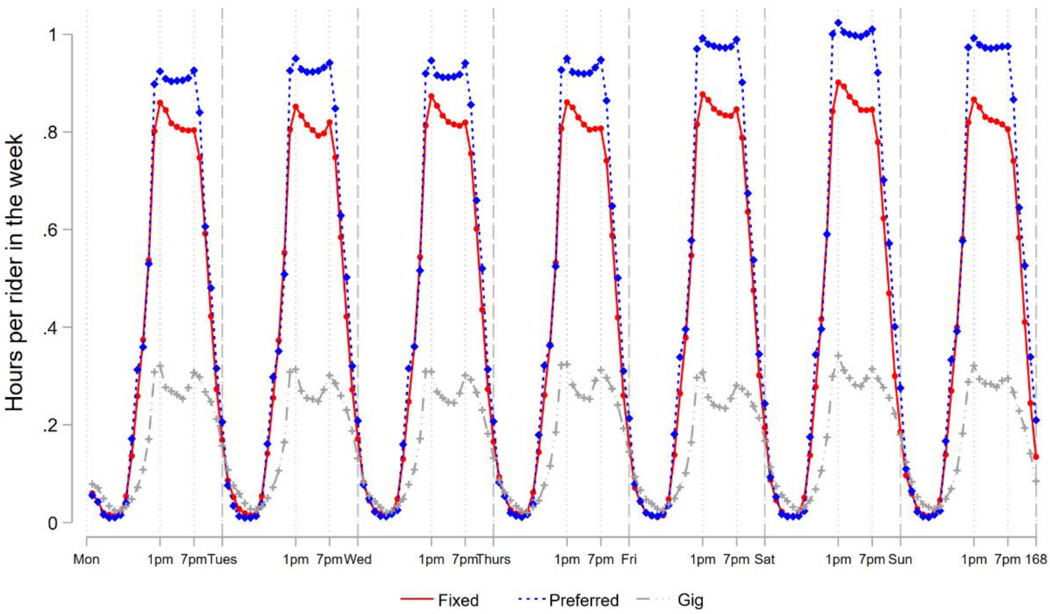

Figure 4: Delivery time allocation for different types of riders

Notes: This figure plots average active delivery time per rider by hour of the week, aligning corresponding hours across two consecutive weeks (e.g. Hour 1 covers both Oct 21 00:00–01:00 and Oct 28 00:00–01:00). On the horizontal axis, each point represents one of the 168 hours in a week; on the vertical axis, the value indicates the mean minutes spent actively delivering per rider in that hour (total delivery minutes divided by the number of riders in that category). A value near 60 means virtually all riders in the group worked the full hour.

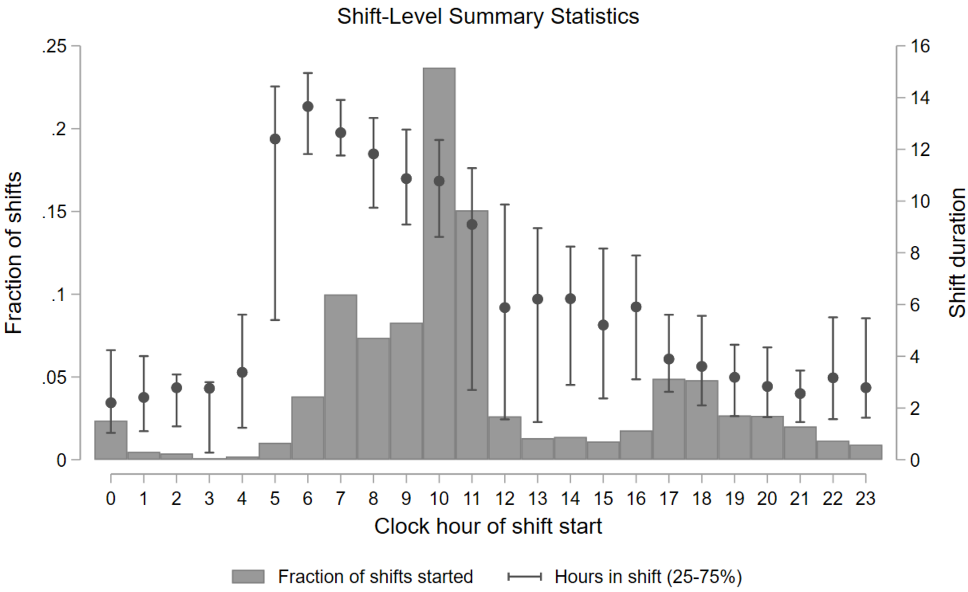

Figure 5: Shift-Level summary statistics

Notes: On the horizontal axis, each integer hour (0–23) marks the clock-hour at which a shift begins. The left vertical axis corresponds to the bar chart, which shows the proportion of all shifts that start in each hour; the right vertical axis corresponds to the overlaid lollipop plots, which indicate the 25th–75th percentile range of shift durations for shifts beginning at that hour, with the median marked by the dot.

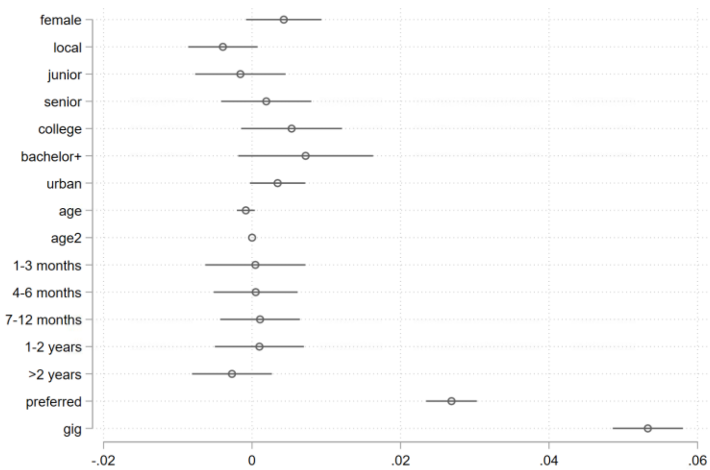

Figure 6: Balancing test

Notes: This figure plots the coefficients from regressing the occurrence of a negative shock on various rider-level characteristics. After controlling for contract types and temporal effects, shocks are shown to be approximately random (joint F-test p-value = 0.07).

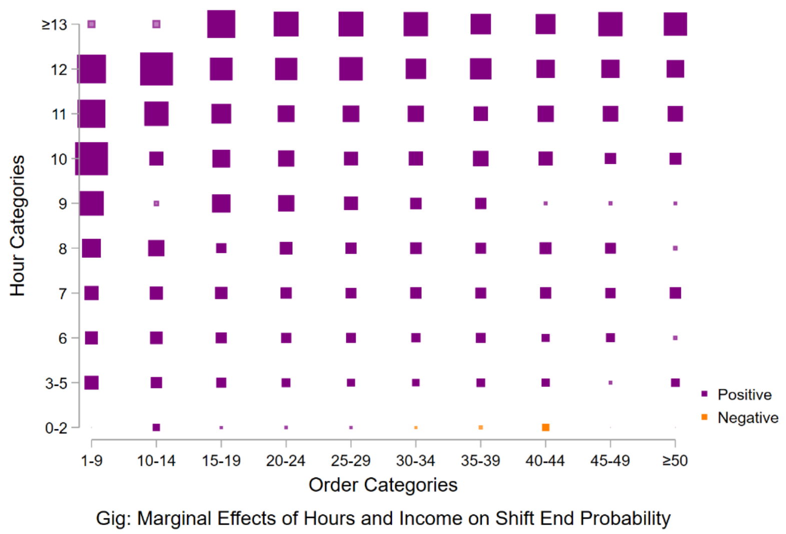

Figure 7a: Effects of cumulative orders and hours on stopping for flexible riders

Notes: We partition total hours worked into ten bins (0–2 h, 3–5 h, … ⩾13 h) and orders delivered into ten bins (1–9 orders, 10–14 orders, … ⩾50 orders). Each square's area is proportional to the absolute value of the estimated coefficient from our linear probability model (analogous to Farber's specifications), and every coefficient is negative, indicating that both longer hours and more orders tend to increase the probability of stopping (i.e. reduce the propensity to continue). Squares that are nearly invisible correspond to estimates that fail to reach conventional levels of statistical significance.

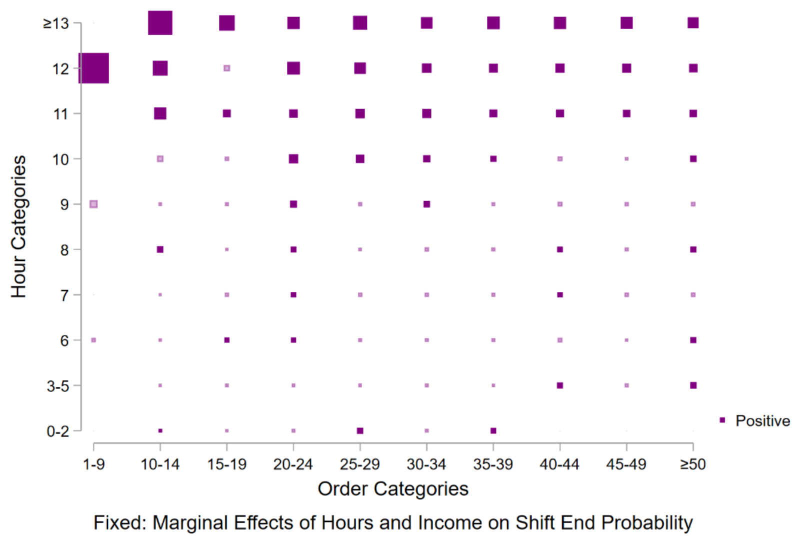

Figure 7b: Effects of cumulative orders and hours on stopping for fixed riders

Notes: We partition total hours worked into ten bins (0–2 h, 3–5 h, … ⩾13 h) and orders delivered into ten bins (1–9 orders, 10–14 orders, … ⩾50 orders). Each square's area is proportional to the absolute value of the estimated coefficient from our linear probability model (analogous to Farber's specifications), and every coefficient is negative, indicating that both longer hours and more orders tend to increase the probability of stopping (i.e. reduce the propensity to continue). Squares that are nearly invisible correspond to estimates that fail to reach conventional levels of statistical significance.

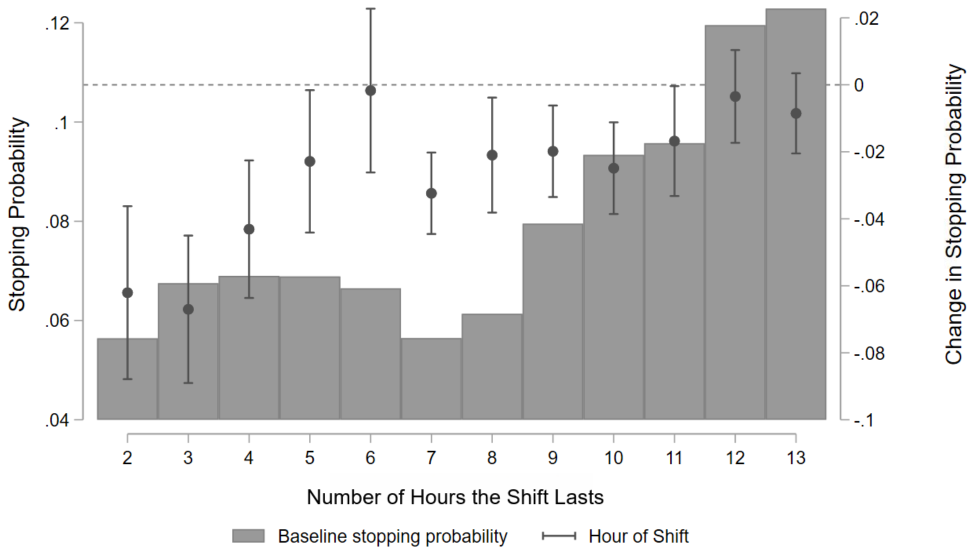

Figure 8: Heterogeneity by shift duration

Notes: The left y-axis indicates the baseline stopping probability, represented by gray bars, while the right y-axis measures the change in stopping probability due to a bad shock, shown as black dots with 95% confidence intervals. The x-axis denotes the total number of hours in a rider's shift.

Table 1: Summary Statistics

Panel A: Rider-Level Summary Statistics			
Statistic	Observations	Mean	SD
Female	911	0.166	0.372
Local	911	0.491	0.500
Urban	911	0.318	0.466
Age	911	35.138	8.745
Years of Education	911	7.921	3.841
Fulltime	911	0.434	0.496
Number of Months Worked	911	19.418	13.787
Panel B: Order-Level Summary Statistics			
Statistic	Observations	Mean	SD
Order Duration (min)	472045	26.080	12.687
Distance (m)	471861	2.141	1.390
Abnormal Report	472045	0.041	0.199
Bad shocks	472045	0.022	0.146
Low ratings	472045	0.001	0.028
Complaints	472045	0.021	0.142
Rider-Caused Cancelations	472045	0.001	0.025
Panel C: Shift-Level Summary Statistics			
Statistic	Observations	Mean	SD
Number of Orders	16112	29.298	20.505
Number of Total Hours	16112	8.450	4.711
Number of Working Hours	16112	5.289	3.092
Number of Bad shocks	16112	1.205	1.792
Number of Abnormal Events	16112	0.638	1.493

Notes: This table reports descriptive statistics at the rider, order, and shift levels. The sample consists of 917 riders and 472,045 orders from Ele.me administrative data in 2024. "Bad shocks" is a dummy variable indicating if an order received a negative rating, complaint, or was canceled by the customer due to dissatisfaction with the rider.

Table 2: Baseline regression results

Type	Full Sample		Fixed		Preferred		Gig	
Variables	$Quit_{it}$	$Rest_{it(T+1)}$	$Quit_{it}$	$Rest_{it(T+1)}$	$Quit_{it}$	$Rest_{it(T+1)}$	$Quit_{it}$	$Rest_{it(T+1)}$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Badshock	-0.0112***	-0.0256***	0.0029	-0.0122	-0.0047*	-0.0236***	-0.0204***	-0.0300***
	(0.0021)	(0.0029)	(0.0074)	(0.0084)	(0.0024)	(0.0040)	(0.0034)	(0.0045)
Controls	X	X	X	X	X	X	X	X
Rider FE	X	X	X	X	X	X	X	X
Observations	472,045	455,933	229,986	223,397	144,221	141,171	97,838	91,365
R-squared	0.0777	0.1784	0.0787	0.1708	0.0823	0.1898	0.0645	0.1848

Notes: Standard errors are clustered at the rider level and reported in parentheses. The dependent variable in odd columns is a dummy for quitting, in even columns, it is the rest time (in hours) in the following hour. All specifications include rider fixed effects, and controls for weather, hour-of-day, day-of-week, week, and nonlinear effects of cumulative hours and income.

* p < 0.1, ** p < 0.05, *** p < 0.01.

Table 3: Incorporate cubic polynomials of cumulative variables

Type	Full Sample		Fixed		Preferred		Gig	
Variables	$Quit_{it}$	$Rest_{it(T+1)}$	$Quit_{it}$	$Rest_{it(T+1)}$	$Quit_{it}$	$Rest_{it(T+1)}$	$Quit_{it}$	$Rest_{it(T+1)}$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Badshock	-0.0113***	-0.0232***	0.0028	-0.0100	-0.0044*	-0.0222***	-0.0206***	-0.0274***
	(0.0021)	(0.0028)	(0.0073)	(0.0085)	(0.0024)	(0.0039)	(0.0034)	(0.0042)
H_{it}	0.0095***	0.1584***	0.0027	0.1494***	0.0012	0.1251***	0.0202***	0.2146***
	(0.0019)	(0.0035)	(0.0028)	(0.0045)	(0.0034)	(0.0053)	(0.0038)	(0.0060)
Controls	X	X	X	X	X	X	X	X
Rider FE	X	X	X	X	X	X	X	X
Observations	472,045	455,933	229,986	223,397	144,221	141,171	97,838	91,365
R-squared	0.0786	0.2105	0.0796	0.1970	0.0822	0.2058	0.0655	0.2105

Notes: This table replicates the baseline model but replaces decile bins of cumulative variables with cubic polynomials of cumulative hours and orders. Standard errors are in parentheses.

* p < 0.1, ** p < 0.05, *** p < 0.01.

Table 4: Instrumental variable

Type	Full Sample		Fixed		Preferred		Gig	
Variables	$Quit_{it}$	$Rest_{it(T+1)}$	$Quit_{it}$	$Rest_{it(T+1)}$	$Quit_{it}$	$Rest_{it(T+1)}$	$Quit_{it}$	$Rest_{it(T+1)}$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Badshock	-0.4991***	0.4362**	-1.9736	4.3848**	-0.2353	0.7353	-0.2380***	-0.0958
	(0.1850)	(0.2077)	(1.5555)	(1.9856)	(0.4875)	(0.9437)	(0.0898)	(0.0713)
Controls	X	X	X	X	X	X	X	X
City FE	X	X	X	X	X	X	X	X
F-Statistics	99.744	99.744	7.326	7.326	4.254	4.254	134.080	134.080
Observations	163,049	158,473	95,044	92,493	51,131	50,047	16,870	15,928

Notes: The table presents 2SLS results. The "Badshock" variable is instrumented by customer-specific risk and shop-specific risk (historical low-rating and complaint propensities). F-statistics for the first stage are reported. All columns include rider fixed effects and the full set of controls used in Table 2.

* p < 0.1, ** p < 0.05, *** p < 0.01.

Table 5: The effects of other anticipated factors

Type	Full Sample		Fixed		Preferred		Gig	
Variables	$Quit_{it}$	$Rest_{it(T+1)}$	$Quit_{it}$	$Rest_{it(T+1)}$	$Quit_{it}$	$Rest_{it(T+1)}$	$Quit_{it}$	$Rest_{it(T+1)}$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$Report_{it(T-1)}$	0.0034***	-0.0374***	0.0008	-0.0442***	0.0018*	-0.0300***	0.0087***	-0.0366***
	(0.0009)	(0.0017)	(0.0013)	(0.0030)	(0.0011)	(0.0023)	(0.0021)	(0.0033)
Weather (Base: Normal Conditions)								
Level 1 Bad Weather	-0.0001	-0.0241***	0.0008	-0.0274***	0.0010	-0.0232***	-0.0065**	-0.0177***
	(0.0008)	(0.0021)	(0.0011)	(0.0029)	(0.0011)	(0.0039)	(0.0025)	(0.0046)
Level 2 Bad Weather	-0.0053***	-0.0592***	-0.0038	-0.0636***	-0.0014	-0.0576***	-0.0204***	-0.0527***
	(0.0017)	(0.0040)	(0.0024)	(0.0058)	(0.0020)	(0.0066)	(0.0064)	(0.0095)
Level 3 Bad Weather	-0.0069***	-0.0584***	-0.0058*	-0.0635***	-0.0024	-0.0494***	-0.0226***	-0.0604***
	(0.0023)	(0.0049)	(0.0032)	(0.0081)	(0.0024)	(0.0079)	(0.0082)	(0.0090)
Level 4 Bad Weather	-0.0012	-0.0854***	0.0026	-0.0981***	-0.0040	-0.0607***	-0.0105	-0.0868***
	(0.0048)	(0.0087)	(0.0061)	(0.0137)	(0.0074)	(0.0113)	(0.0122)	(0.0186)
Day of Week (Base: Monday)								
Tuesday	-0.0001	0.0019	-0.0004	-0.0022	0.0015	0.0045	-0.0024	0.0050
	(0.0007)	(0.0019)	(0.0008)	(0.0028)	(0.0010)	(0.0030)	(0.0025)	(0.0044)
Wednesday	-0.0005	0.0019	-0.0017*	-0.0001	0.0008	0.0039	-0.0005	0.0016
	(0.0007)	(0.0019)	(0.0008)	(0.0028)	(0.0010)	(0.0030)	(0.0028)	(0.0044)
Thursday	-0.0018**	-0.0026	-0.0021**	-0.0055*	0.0001	-0.0005	-0.0044	0.0000
	(0.0008)	(0.0020)	(0.0010)	(0.0030)	(0.0010)	(0.0031)	(0.0027)	(0.0042)
Friday	-0.0054***	-0.0125***	-0.0063***	-0.0122***	-0.0026**	-0.0163***	-0.0084***	-0.0098**
	(0.0008)	(0.0019)	(0.0010)	(0.0029)	(0.0010)	(0.0032)	(0.0028)	(0.0044)
Saturday	-0.0068***	-0.0226***	-0.0062***	-0.0224***	-0.0041***	-0.0305***	-0.0157***	-0.0151***
	(0.0008)	(0.0021)	(0.0010)	(0.0031)	(0.0011)	(0.0033)	(0.0028)	(0.0042)
Sunday	-0.0030***	-0.0276***	-0.0032***	-0.0293***	-0.0006	-0.0340***	-0.0089***	-0.0170***
	(0.0008)	(0.0022)	(0.0010)	(0.0032)	(0.0010)	(0.0036)	(0.0029)	(0.0045)
Controls	X	X	X	X	X	X	X	X
Rider Fixed Effect	X	X	X	X	X	X	X	X
Observations	472,045	455,933	229,986	223,397	144,221	141,171	97,838	91,365
R-squared	0.0777	0.1784	0.0787	0.1708	0.0823	0.1898	0.0645	0.1848

Notes: This table examines responses to predictable fluctuations. The baseline category for weather is “Normal”, and for the day of the week is “Monday”. Controls and rider fixed effects are included as in Table 2.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: The persistence of shock effects

Type	Full Sample		Fixed		Preferred		Gig	
Variables	$Quit_{it}$	$Rest_{it(T+1)}$	$Quit_{it}$	$Rest_{it(T+1)}$	$Quit_{it}$	$Rest_{it(T+1)}$	$Quit_{it}$	$Rest_{it(T+1)}$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Last 0-20 mins	-0.0041***	-0.0472***	0.0012	-0.0459***	0.0003	-0.0395***	-0.0107***	-0.0540***
	(0.0016)	(0.0026)	(0.0049)	(0.0068)	(0.0017)	(0.0033)	(0.0028)	(0.0044)
Last 20-40 mins	-0.0003	-0.0396***	-0.0112**	-0.0338***	-0.0029	-0.0331***	0.0069	-0.0510***
	(0.0022)	(0.0026)	(0.0051)	(0.0070)	(0.0024)	(0.0031)	(0.0049)	(0.0051)
Last 40-60 mins	-0.0041**	-0.0254***	-0.0104**	-0.0093	-0.0063***	-0.0240***	0.0011	-0.0346***
	(0.0021)	(0.0027)	(0.0042)	(0.0079)	(0.0023)	(0.0033)	(0.0045)	(0.0048)
Last 60-80 mins	-0.0002	-0.0094***	-0.0053	-0.0130	0.0019	-0.0137***	-0.0026	-0.0057
	(0.0023)	(0.0029)	(0.0049)	(0.0092)	(0.0029)	(0.0032)	(0.0044)	(0.0057)
Last 80-100 mins	-0.0003	0.0004	-0.0150***	0.0073	0.0009	-0.0044	0.0031	0.0000
	(0.0023)	(0.0034)	(0.0049)	(0.0086)	(0.0027)	(0.0045)	(0.0048)	(0.0066)
Last 100-120 mins	0.0027	0.0026	0.0058	-0.0084	0.0000	-0.0071	0.0063	0.0155*
	(0.0028)	(0.0040)	(0.0072)	(0.0102)	(0.0029)	(0.0046)	(0.0058)	(0.0084)
Controls	X	X	X	X	X	X	X	X
Rider FE	X	X	X	X	X	X	X	X
Observations	472,045	455,933	229,986	223,397	144,221	141,171	97,838	91,365
R-squared	0.0715	0.1558	0.0711	0.1531	0.0748	0.1798	0.0610	0.1317

Notes: This table estimates the impact of negative shocks partitioned by the time elapsed since the shock occurred

(0–20 mins, 20–40 mins, etc.). This tests for the dynamic adjustment of reference points.

* p < 0.1, ** p < 0.05, *** p < 0.01.

Table 7: The heterogeneity of time slots

Type	Full Sample		Fixed		Preferred		Gig	
Variables	$Quit_{it}$ (1)	$Rest_{it(T+1)}$ (2)	$Quit_{it}$ (3)	$Rest_{it(T+1)}$ (4)	$Quit_{it}$ (5)	$Rest_{it(T+1)}$ (6)	$Quit_{it}$ (7)	$Rest_{it(T+1)}$ (8)
Badshock	-0.0012	-0.0233***	0.0248**	-0.0105	0.0083**	-0.0073	-0.0104**	-0.0217***
	(0.0032)	(0.0042)	(0.0114)	(0.0132)	(0.0035)	(0.0050)	(0.0047)	(0.0061)
H_{it}	0.0089***	0.0319***	0.0083***	0.0309***	0.0076***	0.0237***	0.0092***	0.0446***
	(0.0005)	(0.0008)	(0.0010)	(0.0011)	(0.0012)	(0.0015)	(0.0008)	(0.0015)
Badshock	-0.0019***	-0.0006	-0.0036*	-0.0002	-0.0021***	-0.0028***	-0.0024**	-0.0021*
$\times H_{it}$	(0.0006)	(0.0007)	(0.0020)	(0.0018)	(0.0008)	(0.0008)	(0.0011)	(0.0013)
Controls	X	X	X	X	X	X	X	X
Rider FE	X	X	X	X	X	X	X	X
Observations	472,045	455,933	229,986	223,397	144,221	141,171	97,838	91,365
R-squared	0.0758	0.1823	0.0745	0.1745	0.0775	0.1891	0.0651	0.1965

Notes: This table includes interaction terms between the bad shock indicator and cumulative hours.

* p < 0.1, ** p < 0.05, *** p < 0.01.

Table 8a: Heterogeneity in rider characteristics (fixed riders)

Fixed	Scheme		Experience		Gender	
Variables	$Quit_{it}$ (1)	$Rest_{it(T+1)}$ (2)	$Quit_{it}$ (3)	$Rest_{it(T+1)}$ (4)	$Quit_{it}$ (5)	$Rest_{it(T+1)}$ (6)
Badshock (dummy)	-0.0001	-0.0170	0.0113	-0.0164	-0.0016	-0.0061
	(0.0094)	(0.0110)	(0.0117)	(0.0131)	(0.0075)	(0.0085)
Badshock×Fulltime	0.0066	0.0102				
	(0.0148)	(0.0171)				
Badshock×Experienced			-0.0162	0.0080		
			(0.0144)	(0.0169)		
Badshock×Female					0.0419*	-0.0582**
					(0.0215)	(0.0289)
Controls	X	X	X	X	X	X
Rider Fixed Effect	X	X	X	X	X	X
Observations	229,986	223,397	229,986	223,397	229,986	223,397
R-squared	0.0787	0.1708	0.0787	0.1708	0.0787	0.1708

Notes: "Full-time" status is based on survey self-reports (working exclusively for Ele.me). "Experienced" indicates riders with a tenure of more than one year on the platform. Columns (5) and (6) compare male and female riders within each contract category.

* p < 0.1, ** p < 0.05, *** p < 0.01.

Table 8b: Heterogeneity in rider characteristics (ordinary gig riders)

Gig	Scheme		Experience		Gender	
Variables	$Quit_{it}$	$Rest_{it(T+1)}$	$Quit_{it}$	$Rest_{it(T+1)}$	$Quit_{it}$	$Rest_{it(T+1)}$
	(1)	(2)	(3)	(4)	(5)	(6)
Badshock (dummy)	-0.0156***	-0.0251***	-0.0299***	-0.0406***	-0.0187***	-0.0313***
	(0.0038)	(0.0052)	(0.0050)	(0.0076)	(0.0036)	(0.0049)
Badshock×Fulltime	-0.0166**	-0.0170*				
	(0.0072)	(0.0100)				
Badshock×Experienced			0.0152**	0.0169*		
			(0.0066)	(0.0095)		
Badshock×Female					-0.0115	0.0086
					(0.0094)	(0.0125)
Controls	X	X	X	X	X	X
Rider Fixed Effect	X	X	X	X	X	X
Observations	97,838	91,365	97,838	91,365	97,838	91,365
R-squared	0.0645	0.1848	0.0645	0.1848	0.0645	0.1848

Notes: "Full-time" status is based on survey self-reports (working exclusively for Ele.me). "Experienced" indicates riders with a tenure of more than one year on the platform. Columns (5) and (6) compare male and female riders within each contract category.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 9: The effects on number of orders during the next hour

Type	Full Sample	Fixed	Preferred	Gig
$Orders_{it(T+1)}$	(1)	(2)	(3)	(4)
Badshock	0.1344***	-0.0664	0.2833***	0.0535
	(0.0379)	(0.1085)	(0.0621)	(0.0407)
Controls	X	X	X	X
Rider FE	X	X	X	X
Observations	393,301	196,127	125,820	71,354
R-squared	0.2912	0.2528	0.2824	0.2760

Notes: The dependent variable is the number of orders completed in the hour following order t.

This test confirms that the reduction in rest time translates into actual increased earnings, ruling out severe algorithmic penalties that might prevent recovery.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Appendix

Table A1: Order characteristics following negative shocks

Type	Full Sample		Fixed		Preferred		Gig	
Variables	$Fee_{it(T+1)}$ (1)	$Value_{it(T+1)}$ (2)	$Fee_{it(T+1)}$ (3)	$Value_{it(T+1)}$ (4)	$Fee_{it(T+1)}$ (5)	$Value_{it(T+1)}$ (6)	$Fee_{it(T+1)}$ (7)	$Value_{it(T+1)}$ (8)
Badshock	-0.0018	-0.0462	-0.0049	-1.1059	0.0115	0.4254	-0.0232	-0.4197
	(0.0280)	(0.3260)	(0.0464)	(0.9283)	(0.0147)	(0.3835)	(0.0597)	(0.5943)
Controls	X	X	X	X	X	X	X	X
Rider FE	X	X	X	X	X	X	X	X
Observations	393,299	393,299	196,127	196,127	125,820	125,820	71,352	71,352
R-squared	0.1998	0.0351	0.2229	0.0231	0.1896	0.1768	0.1479	0.1359

Notes: This table examines the impact of a negative shock on the delivery fee paid by the customer and the monetary value of products in the subsequent hour. These results are used to rule out lasting algorithmic penalties beyond the immediate financial fine.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A2: Within-Shift placebo test

Type	Full Sample		Fixed		Preferred		Gig	
Variables	$Quit_{it}$ (1)	$Rest_{it(T+1)}$ (2)	$Quit_{it}$ (3)	$Rest_{it(T+1)}$ (4)	$Quit_{it}$ (5)	$Rest_{it(T+1)}$ (6)	$Quit_{it}$ (7)	$Rest_{it(T+1)}$ (8)
Badshock	-0.0003	0.0037	0.0060	0.0021	-0.0029	0.0005	0.0006	0.0068*
	(0.0020)	(0.0025)	(0.0056)	(0.0082)	(0.0022)	(0.0034)	(0.0035)	(0.0038)
Controls	X	X	X	X	X	X	X	X
Rider FE	X	X	X	X	X	X	X	X
Observations	472,045	455,933	229,986	223,397	144,221	141,171	97,838	91,365
R-squared	0.0776	0.1782	0.0787	0.1707	0.0823	0.1896	0.0642	0.1843

Notes: To test for exogeneity, we randomly reassign the actual number of shocks experienced by each rider within a shift across their sequence of orders. The results show that these "pseudo-shocks" have no significant impact on labor supply, supporting the causal interpretation of our main results.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A3a: Sensitivity analysis using a six-hour interval to define shifts

Type	Full Sample		Fixed		Preferred		Gig	
Variables	$Quit_{it}$	$Rest_{it(T+1)}$	$Quit_{it}$	$Rest_{it(T+1)}$	$Quit_{it}$	$Rest_{it(T+1)}$	$Quit_{it}$	$Rest_{it(T+1)}$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Badshock	-0.0108***	-0.0255***	0.0009	-0.0115	-0.0044*	-0.0239***	-0.0189***	-0.0299***
	(0.0020)	(0.0030)	(0.0071)	(0.0084)	(0.0024)	(0.0040)	(0.0033)	(0.0047)
Controls	X	X	X	X	X	X	X	X
Rider FE	X	X	X	X	X	X	X	X
F-Statistics	472,045	456,508	229,986	223,574	144,221	141,242	97,838	91,692
Observations	0.0729	0.1794	0.0807	0.1699	0.0839	0.1903	0.0668	0.1889

Notes: This table test the sensitivity of our results to the definition of a "shift". While the baseline model uses a 5-hour gap between orders to define a new shift, this specification uses a 7-hour thresholds. The coefficients for "Badshock" remain stable across this definition.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A3b: Sensitivity analysis using a seven-hour interval to define shifts

Type	Full Sample		Fixed		Preferred		Gig	
Variables	$Quit_{it}$	$Rest_{it(T+1)}$	$Quit_{it}$	$Rest_{it(T+1)}$	$Quit_{it}$	$Rest_{it(T+1)}$	$Quit_{it}$	$Rest_{it(T+1)}$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Badshock	-0.0102***	-0.0254***	0.0004	-0.0115	-0.0042*	-0.0242***	-0.0175***	-0.0294***
	(0.0020)	(0.0030)	(0.0070)	(0.0084)	(0.0024)	(0.0040)	(0.0033)	(0.0047)
Controls	X	X	X	X	X	X	X	X
Rider FE	X	X	X	X	X	X	X	X
F-Statistics	472,045	456,777	229,986	223,637	144,221	141,273	97,838	91,867
Observations	0.0729	0.1793	0.0812	0.1695	0.0826	0.1889	0.0679	0.1900

Notes: This table test the sensitivity of our results to the definition of a "shift". While the baseline model uses a 5-hour gap between orders to define a new shift, this specification uses a 7-hour threshold. The coefficients for "Badshock" remain stable across this definition.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A4: Logit model estimation

Type	Full Sample	Fixed	Preferred	Gig
Quit$_{it}$	(1)	(2)	(3)	(4)
Badshock	-0.333***	0.161	-0.351**	-0.419***
	(0.0651)	(0.223)	(0.151)	(0.0715)
Controls	X	X	X	X
Rider FE	X	X	X	X
Observations	472,038	229,982	144,221	97,834

Notes: This table reports the marginal effects from a Logit model for the binary decision of ending a shift. This ensures that the findings from our linear probability model (LPM) are not driven by functional form assumptions. Standard errors are in parentheses.

* p < 0.1, ** p < 0.05, *** p < 0.01.

Table A5: Including more granular time fixed effects

Type	Full Sample		Fixed		Preferred		Gig	
Variables	*Quit*$_{it}$	*Rest*$_{it(T+1)}$	*Quit*$_{it}$	*Rest*$_{it(T+1)}$	*Quit*$_{it}$	*Rest*$_{it(T+1)}$	*Quit*$_{it}$	*Rest*$_{it(T+1)}$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Badshock	-0.0112***	-0.0248***	0.0026	-0.0113	-0.0046*	-0.0219***	-0.0205***	-0.0296***
	(0.0021)	(0.0029)	(0.0074)	(0.0084)	(0.0024)	(0.0039)	(0.0034)	(0.0045)
Controls	X	X	X	X	X	X	X	X
Rider FE	X	X	X	X	X	X	X	X
F-Statistics	472,045	455,933	229,986	223,397	144,221	141,171	97,838	91,365
Observations	0.0789	0.1821	0.0803	0.1748	0.0847	0.1970	0.0668	0.1884

Notes: This specification mirrors the strategy of Farber (2015) by adding a fully interacted “Hour×Weekday” fixed effect (168 dummies). This absorbs even finer-grained predictable fluctuations in market demand and supply.

* p < 0.1, ** p < 0.05, *** p < 0.01.

Table A6: Discrete stopping-choice model estimates grouped by order batches

Type	Full Sample		Fixed		Preferred		Gig	
Variables	$Quit_{it}$	$Rest_{it(T+1)}$	$Quit_{it}$	$Rest_{it(T+1)}$	$Quit_{it}$	$Rest_{it(T+1)}$	$Quit_{it}$	$Rest_{it(T+1)}$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Badshock	-0.0073***	-0.0407***	-0.0049	-0.0357***	-0.0035*	-0.0349***	-0.0127***	-0.0479***
	(0.0017)	(0.0027)	(0.0050)	(0.0068)	(0.0018)	(0.0034)	(0.0031)	(0.0043)
Controls	X	X	X	X	X	X	X	X
Rider FE	X	X	X	X	X	X	X	X
Observations	472,045	454,166	229,986	222,653	144,221	140,698	97,838	90,815
R-squared	0.0843	0.1793	0.0856	0.1711	0.0938	0.1909	0.0688	0.1863

Notes: If any order in a batch receives a negative shock, the whole batch is marked as treated. Quitting is defined only after the completion of the entire batch.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

References

Andersen, S., Brandon, A., Gneezy, U., & List, J. (2014). Toward an Understanding of Reference-Dependent Labor Supply: Theory and Evidence from a Field Experiment (No. W20695; p. w20695). National Bureau of Economic Research.

Becker, G. S. (1965). A Theory of the Allocation of Time. *The Economic Journal*, *75*(299), 493. https://doi.org/10.2307/2228949

Camerer, C., Babcock, L., Enstein, G. L., & Thaler, R. (n.d.). LABOR SUPPLY OF NEW YORK CITY CABDRIVERS: ONE DAY AT A TIME. *QUARTERLY JOURNAL OF ECONOMICS*.

Chou, Y.K. (2002). TESTING ALTERNATIVE MODELS OF LABOUR SUPPLY: EVIDENCE FROM TAXI DRIVERS IN SINGAPORE. *The Singapore Economic Review, 47*, 17-47.

Crawford, V. P., & Meng, J. (2011). New York City Cab Drivers' Labor Supply Revisited: Reference-Dependent Preferences with Rational-Expectations Targets for Hours and Income. *The American Economic Review*, *101*(5), 1912–1932.

Crawford, V. P., Jin, M., Meng, J., & Yao, L. (2025). Expectations-based reference-dependence and labor supply: Eliciting cabdrivers' expectations in the field. Journal of Economic Behavior & Organization, 239, 107259. https://doi.org/10.1016/j.jebo.2025.107259

Duong, H. L., Chu, J., & Yao, D. (2023). Taxi Drivers' Response to Cancellations and No-Shows: New Evidence for Reference-Dependent Preferences. *Management Science*, *69*(1), 179–199. https://doi.org/10.1287/mnsc.2022.4349

Farber, H. S. (2005). Is Tomorrow Another Day? The Labor Supply of New York City Cabdrivers. *Journal of Political Economy*, *113*(1), 46–82. https://doi.org/10.1086/426040

Farber, H. S. (2008). Reference-Dependent Preferences and Labor Supply: The Case of New York City Taxi Drivers. *American Economic Review*, *98*(3), 1069–1082. https://doi.org/10.1257/aer.98.3.1069

Farber, H. S. (2015). Why you Can't Find a Taxi in the Rain and Other Labor Su pply Lessons from Cab Drivers*. *The Quarterly Journal of Economics*, *130*(4), 1975–2026. https://doi.org/10.1093/qje/qjv026

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. Econometrica, 47(2), 263. https://doi.org/10.2307/1914185

Kőszegi, B., & Rabin, M. (2006). A Model of Reference-Dependent Preferences. *The Quarterly Journal of Economics*, *121*(4), 1133–1165.

Lucas, R. E., & Rapping, L. A. (1969). Real Wages, Employment, and Inflation. *Journal of Political Economy*, *77*(5), 721–754. https://doi.org/10.1086/259559

Ma, Y., Cao, S., Zhu, X., & Zhu, P. (2024). Weather shocks, dynamic pricing, and labor supply: Evidence from the Shanghai taxi market. *China Economic Quarterly*, *24*(5), 1567–1585. https://doi.org/10.13821/j.cnki.ceq.2024.05.12

Oettinger, G. S. (1999). An Empirical Analysis of the daily Labor supply of Stadium Vendors. *Journal of Political Economy*, *107*(2), 360–392. https://doi.org/10.1086/250063

Thakral, N., & Tô, L. T. (2021). Daily Labor Supply and Adaptive Reference Points. *American Economic Review*, *111*(8), 2417–2443. https://doi.org/10.1257/aer.20170768