Don’t Fake It If You Can’t Make It: Driver Misconduct in Last Mile Delivery

Srishti Arora
INSEAD, srishti.arora@insead.edu

Vivek Choudhary
Nanyang Technological University, vivek.choudhary@ntu.edu.sg

Pavel Kireyev
INSEAD, pavel.kireyev@insead.edu

In the last two decades, last-mile delivery (LMD) firms have seen immense growth fueled by e-commerce’s success, leading to faster and cheaper deliveries. This growth has brought forth new challenges including increased competition and heightened customer expectations. Operating on thin margins, LMD firms strive for successful first-time deliveries to avoid the financial and reputational cost from reattempts. Delivery Agents (DAs) are integral to LMD efficiency, influencing customer experience, delivery success, and productivity. However, most LMD performance enhancement research focuses on process, technology and incentives, presupposing that gig workers will conform to procedures and monitoring tools will function flawlessly. Nevertheless, in practice, DAs deviate from expected behaviors i.e., indulge in misconduct, negatively affecting delivery efficiency often resulting in returned parcels. One of the major misconducts are fake remarked deliveries, wherein DAs intentionally do not deliver the parcels and provide a fake reason for it. For instance, even without reaching a delivery address, a DA remarks ‘customer unavailable’ and record a delivery failure. In this study, we collaborated with a leading Indian LMD firm and, using instrumental variable regression, find that such misconduct leads to a spillover productivity loss. This effect reduces the next day’s successful deliveries by 1.60% and first-time-right deliveries by 1.86%. We discuss misconduct's correlation with several factors such as task complexity, offering actionable managerial insights. We offer novel insights into how undesirable worker behavior can be influenced by opportunistic circumstances. Our study is among the first to quantify misconduct's impact on LMD productivity, highlighting its impact on operational costs.

Keywords: Last mile Delivery; Platforms; Behavioral Operations; Misconduct; Productivity

1. Introduction

Global e-commerce is projected to grow 8% annually from 2020 to 2025 (Gläser et al. 2021), triggering a 36% increase in delivery vehicles across the world's 100 largest cities by 2030 (World Economic Forum 2020). The majority of e-commerce deliveries are managed by last-mile delivery (LMD) firms. These LMD firms encounter numerous challenges including competition, slim profit margins, and continuously increasing customer expectations, despite a lack of willingness to pay for these enhancements.¹ Present-day consumers expect faster deliveries but do not want to bear additional costs (Mangiaracina et al. 2019). To deliver, LMD firms employ contractual gig workers (Altenried 2019) often called delivery agents (DAs).

DAs are instrumental to LMD efficiency, and their conduct directly impacts both the company's reputation and profitability. Despite existing research focusing on consumer behavior and responses to alternative delivery solutions (Wang et al. 2020), the impact of DA misconduct - a pervasive issue in gig economy (Burbano and Chiles 2022) - is still largely understudied, thereby affecting our understanding of impact on delivery efficiency. This study aims to bridge this knowledge gap by focusing on misconduct, an aspect with substantial operational implications (Chan et al. 2021). We aim to quantify misconduct's effect on LMD performance and identify conditions that can exacerbate or mitigate this impact.

LMD, the most complex and expensive part of the supply chain (Seghezzi and Mangiaracina 2022), holds both practical and academic significance, drawing attention from fields such as economics, business, and sustainability (Comi and Savchenko 2021, Fatehi and Wagner 2022, Muralidharan et al. 2021). Despite its challenges, including high fragmentation (Winkenbach 2019) and limited observability,² LMD presents strategic value for businesses seeking competitive advantage (Harrington 2019, Lim et al. 2018). As a result, cost-effective strategies such as minimizing delivery misconduct are highly desirable. This is the focal point of our paper.

Existing research focuses on efficient process designs such as route optimization, ideal facility locations, innovative business models, and alternative delivery methods (Olsson et al. 2019) to enhance LMD performance. These efforts assume a precise adherence to procedures by gig workers and the infallible functionality of monitoring apps. Yet, the realities of the gig economy suggest that DAs often fail to adhere to established processes. Consequently, LMD firms identify misconduct as one of the primary reasons for inefficient delivery and the subsequent return of parcels to merchants. To the best of our knowledge, we are the first to analyze, quantify, and provide circumstances that exacerbate such behavior.

In collaboration with one of the largest LMD firms in India, we study DA misconduct, specifically “fake remarks”, a crucial contributor to parcels’ failed delivery and eventual return to merchants. For each unsuccessful attempt, a DA records a remark indicating failure reasons. DAs can be honest, stating ‘could not deliver,’ or dishonest, making a false remark like ‘the customer was not available,’ without even visiting the customer's address. To account for genuine failures, LMD firms allow three delivery attempts before returning a parcel to the e-commerce platform. These returns incur significant costs, affecting revenues, reverse logistics, and the company's reputation. However, this allowance unintentionally encourages fake remarks, which we frequently observe in our data. Our paper delves into the impact of such behavior on firm performance and seeks to understand the motivations driving DAs to engage in this misconduct.

In many contexts, firms cannot use penalties to deter such behavior as the DAs can easily switch to other platforms due to the contractual nature of the job as well as abundant job opportunities. Furthermore, the negative implications of increased monitoring of gig workers (Liang et al. 2022) may outweigh the benefits. This issue of fake remarks is increasingly acknowledged in the industry, and several startups³,⁴ are attempting to address it through technical solutions or manual verification methods. However, infrastructural limitations in India, the large volume of parcels, inaccuracies in postal codes, high GPS error rates, and unstructured addresses undermine the effectiveness of these solutions, thereby providing an easy loophole for DAs to evade accountability. But this issue isn't restricted to India; globally, there are similar complaints about missing parcels and incorrect delivery updates.⁵,⁶

We employ an instrumental variable regression approach on our archival data to identify the spillover effect of fake remarks on various outcomes such as overall delivery success and first-time-right success. Our results show that fake remarked deliveries have a cascading effect, decreasing overall successful deliveries by 1.6% on the following day. This effect is explained by effort reallocation to conceal misconduct. This misconduct spills over, causing a 1.86% decrease in first-time-right (FTR – successful deliveries in the first attempt) deliveries the next day, significantly affecting firm efficiency given the importance of FTR in supply chain efficiency (Rautela et al. 2021).⁷ Reattempts are costly, but do not hurt the DA as a major part of her compensation is fixed. This seemingly small percentage can cost millions of dollars annually. We explore how workers adjust their behavior to their peers and how opportunistic circumstances such as cash-on-delivery parcels can induce misconduct, exacerbating the impact of fake remarks. We also discuss how misconduct is associated with task complexity, loyalty, and professionalism.

---

³ https://www.clickpost.ai/blog/top-8-reasons-for-ndr-in-ecommerce-and-how-to-solve-them#title_11
⁴ https://www.shiprocket.in/blog/know-how-you-can-prevent-fake-delivery-attempts/
⁵ https://www.yugatech.com/business/why-your-lazada-orders-are-flagged-as-delivered-even-if-theyre-not/
to provide actionable insights. These findings suggest the need for strategic allocation of parcels, prioritizing FTR success, and recognizing situations favorable to misconduct when designing processes. Counterintuitively, the effect of fake remarks worsens when the DA is familiar with the area, suggesting DAs with misconduct history should be allocated to newer areas.

Our study contributes to the existing literature on LMD, people-centric operations, and unethical behavior, offering valuable insights for both academia and industry. Misconduct can inflate operating costs, increase carbon footprint, provoke customer churn, and tarnish the relationship with e-commerce partners. We underscore the need for further research into misconduct within delivery firms, emphasizing the strategic importance of the last mile in shaping a firm's competitive edge and the broader sustainability of the e-commerce ecosystem. By doing so, we contribute to the growing body of empirical behavioral operations research (Gallien et al. 2016, Simchi-Levi 2014, Tang 2016). We believe that our work offers insights into the profound impact of aberrant behaviors on workers’ productivity in last-mile logistics.

2. Last Mile Delivery Process

Our collaborator provides logistics for major e-commerce entities. As of June 2021, its expansive network includes over 2,000 delivery centers across 15,000+ postal codes in tier 1 and 2 cities. When a consumer places an order on the partner e-commerce website, the LMD firm collects the parcel from the designated warehouse, routing it via regional centers to the delivery center. The delivery center is the last stop for the parcel before it gets delivered to the customer by the DAs (i.e., the last-mile leg which is the focus of this paper). The LMD firm employs contractual DAs, who primarily use two-wheelers for deliveries.

Figure 1: Dispatch Lifecycle
Center managers, upon recruiting a DA, allocate them to an exclusive area (non-overlapping between DAs) within their center's coverage region, reflecting demand needs and the DA’s familiarity with the area. The intent is to balance parcel distribution amongst all the DAs by apportioning varied geographical territories. Such division of exclusive areas ensures a nearly equal workload across DAs. Once contracted, DAs must report to their delivery centers every morning when they pick up the parcels, and in the evening, when they return the undelivered parcels. The manager overseeing the center assigns each day's dispatch, which is a collection of parcels, to the DA. The allocation of dispatches takes no account of any backlogs; every delivery designated for a particular area on a given day is entrusted to the DA. On average, a mere 8.4% of the parcels are carryovers, with the majority being fresh parcels. To discourage intentional carryover, fresh parcels are supplied regardless of the presence of any carryovers. DAs have no control over the quantity of parcels, as these arrive exogenously based on customer demand in the area. This is independent of the work allocation and the behavior of the DAs.

As illustrated in Figure 1, a DA handles three parcel types daily – a) fresh and unattempted parcels to be attempted for the first time, b) fake remarked reattempt parcels, previously fake remarked by the same DA, and c) genuine reattempt parcels, undelivered previously by the same DA because of genuine reasons. An undelivered parcel is re-assigned to the same DA the following day. After the allocation, DAs have the autonomy to determine the parcel delivery sequence and route. Naturally, while many DAs prioritize proximity to the delivery center, others may consider parcel type (e.g., carryover first followed by fresh). Our DA interviews corroborate these strategies. Despite this autonomy, the firm requires that each parcel is attempted every day, and that any issues are reported transparently. If DAs are unable to attempt a parcel due to certain constraints, they can leave the remark “not attempted.”

Although “not attempted” remarks are permitted, they are seldom used due to potential for managerial reprimand and lack of incentive. According to our interviews with DAs, in practice, all parcels are expected to be attempted, regardless of the hours worked. Section 3.1 delves into this, but our analysis validates the infrequency of such entries. Therefore we do not consider “not attempted” in our analysis. However, some DAs leave parcels untouched without taking any action; we refer to such parcels as “unattempted”. Note that not every parcel delivery attempt culminates in success. Parcels from unsuccessful delivery attempts are returned by the DA to the delivery center at the end of the day for subsequent tries on the following days. After three unsuccessful attempts, parcels are returned to the merchant.

---

8 Meaning each DA gets 8-hours of workload every day.
9 “I deliver block by block based on distance from the delivery center” and “I deliver first in the nearest areas but attempt the carryover parcels before the fresh if the parcels are from the same area.”
10 “I have no choice but to deliver all parcels assigned to my area,” and “Regardless of the load, we have to attempt all the parcels.”
11 This is a common agreement and a norm between e-commerce firms and LMD firms.
To facilitate the delivery process, the firm provides a smartphone app to the DAs. The DAs can view each parcel’s attempt history in their app. Figure A.1a in the online appendix\textsuperscript{12} shows the screenshots of the DA app displaying parcels and delivery details. DAs’ can record delivery status and contact the consignees via the app. DAs are asked to call customers before attempting delivery. Calls can be made through the app, and the firm tracks their duration. When a DA attempts a delivery, they need to enter a remark indicating success or failure (Figure A.1b). For unsuccessful deliveries, they must select a reason from a pre-populated list of remarks within the app (Figure A.1c), such as “Consignee unavailable” and “Entry restricted area”.

According to the firm, the largest portion of the parcels returned to the merchants is due to fake remarks, a pervasive issue confronting all industry players.\textsuperscript{13} Consequently, the firm conducts a daily audit of failed deliveries and seller complaints of false attempts. Yet, due to the massive scale of operations, full auditing is impractical and limited to the center manager making random calls to consignees to verify failure reasons. Moreover, full auditing is avoided to contain costs and prevent tarnishing customer perceptions, acknowledging that transparency doesn't always ensure satisfaction (Bray 2020). Excessive auditing might even provoke a perception of misconduct.\textsuperscript{14} After years of refinement, the firm has developed an algorithm to identify remarks as “fake”, classifying them based on the DA’s location during the remark entry and call records to the consignee. Table A.1 illustrates the four conditions for identifying a fake remark. As fake remark audits are random, DAs cannot predict which deliveries may get audited, and the fake remark logic is not disclosed to them to avoid gaming. Only the daily identified fake remarks are communicated to the DAs, and weekly performance reports are generated for the DAs and their delivery center.

DAs receive a fixed salary with a minor variable component tied to successful deliveries made. This variable component is inconsequential as compared to their fixed pay, which is supported by DA interviews.\textsuperscript{15} The focal firm maintains this pay structure due to hiring challenges; initial attempts at a large

\textsuperscript{12} Note: All tables, figures, and sections having a prefix “A.” are available in online appendix (e-companion).

\textsuperscript{13} The advent of start-ups like Shiprocket and iThink Logistics, offering Non-Delivery Response Teams to handle fake remarks, underscores the scale of the problem.

\textsuperscript{14} Decrinis (2022) explains that enforcement systems, such as manual auditing, that rely on rewards and penalties, might aid ethical fading, as workers might feel controlled and thus resort to more misconduct under high behavioral invisibility. As we discuss in section 3, the choice to fake a delivery depends on the behavioral invisibility of the parcel. Bernstein (2012) echoes this, stating that more observability is sometimes detrimental to worker performance, as it induces strategic behavior due to perceived control. Liang et al. (2022) reported that the downside of increased monitoring in a gig economy makes it infeasible to keep a check on workers’ behavior, and Burbano and Chiles (2022) explained that the threat of monitoring reduces employees’ trust of employers.

\textsuperscript{15} “We get 7 Rs (0.08 USD) per parcel after the target. My incentive was just 184 Rs (2.27 USD) for this month. It is insignificant for me. Target is usually too high to earn any additional incentive (beyond my fixed salary)” and “Irrespective of the load we have to attempt all the parcels, so why think about incentive.”
variable component led to inconsistent compensation and driver dissatisfaction. Given the minimal impact of misconduct on incentives in our setup, we are able to distinctly isolate and understand the effect of misconduct on performance.

3. Literature and Hypotheses Development

Our work intersects with three research areas: last-mile delivery, people-centric operations, and unethical behavior. The first stream, studying LMD, falls under the broader umbrella of platforms, a topic gaining significant traction in operations management (Zhang et al. 2021). LMD is studied across disciplines, including economics, sustainability, and operations management. A multitude of management scholars explore how LMD parameters, like operational transparency (Bray 2020), delivery speed (Fisher et al. 2019), promised delivery date or delivery speed disclosure (Cui et al. 2020, Salari et al. 2022), and on-time delivery (Liu et al. 2021) can improve performance metrics, such as sales. Moreover, research on delivery worker behavior within LMD is growing and has identified the effects of factors such as geographical familiarity (Mao et al. 2019), and ratings and penalties (Xu et al. 2020) on LMD performance. A few papers address operational failures in LMD, such as the cost implications of failed delivery attempts (Lim et al. 2023), and the effects of order fulfillment delays on customers’ anxiety (Rao et al. 2011). Yet, to the best of our knowledge, no study has examined the impact of workers’ misconduct and strategic behavior on operational efficiency in LMD.

Additionally, ample research has been conducted to enhance LMD’s environmental sustainability, exploring topics like traffic congestion and emissions (Mucowska 2021). Proposed solutions include urban consolidation centers (van Heeswijk et al. 2020), customer pickup lockers (Lyu and Teo 2022), shared mobility (Qi et al. 2018), electric vehicles (Siragusa et al. 2022), and sharing environmental impact information with customers (Ignat and Chankov 2020). FTR failures, estimated between 10% and 50% (Voigt et al. 2022), not only increase operating costs (Hasija et al. 2020) and customer attrition, but also considerably increase emissions (Edwards et al. 2009). However, existing studies overlook the extra emissions from failed deliveries. While some research propose solutions like attended collection points (Song et al. 2009) and customer availability profiles in routing algorithms (Voigt et al. 2022), the role of DA misconduct in delivery failures remains unexplored.

---

16 As LMD firms compete for workers within India’s growing economy, the firm aims to mirror the stability of a fixed-salary job, offering monthly contracts over per-piece remuneration. Given the defined delivery areas but uncertain volume, some DAs might end up getting higher enumerations than others in piece rate compensation. This could spark competition for delivery areas and favoritism by the center managers. By minimizing variable pay, the firm prevents competition for delivery areas, pay disparity, and financial instability. This policy also encourages the attempt of all parcels, eliminating daily income targeting. Salaries are consistent across all centers but adjusted to city’s cost-of-living index. There was no change in incentive structure during our study, though the firm withheld incentive data due to confidentiality.
The second stream of literature our work speaks to is People Centric Operations, which is defined as “the study of how people affect the performance of operational processes” (Roels and Staats 2021). Most PCO studies focus on decision-making processes to improve performance. For instance, a few studies on employee strategic behavior examine how an increased workload’s impact on performance can be explained by a physician’s task completion preferences (KC et al. 2020) and how worker misconduct varies in the presence of counterproductive peers (Chan et al. 2021). However, workers’ strategic behavior and its implications on firm performance have been studied sparingly in the OM literature (Roels and Staats 2021). Our paper addresses these gaps by investigating the impact of strategic behavior on operational efficiency and demonstrating its adverse effect on firm’s performance.

Our work also intersects with a third stream of literature focusing on unethical or aberrant behavior. Aberrant behavior\(^\text{17}\) leading to decreased worker productivity has been conceptualized in various ways across economics and business literature (Eliyana and Sridadi 2020). The literature also provides theory on why workers indulge in misconduct. For instance, Pierce and Balasubramanian’s (2015) review suggests that social and professional interactions can increase misconduct within organizations, with social comparison serving as an additional catalyst. Following Kidwell and Bennett (1993), we connect the theory of unethical behavior to LMD, where worker behavior is crucial. They postulate several causes for misconduct: first, the focal worker might withhold effort if they believe others are doing so; second, they might limit effort when they perceive the task to be less observable to the supervisor; and third, the worker’s effort can be positively influenced by the perceived lack of alternative employment opportunities. Overall, the OM literature has insufficiently examined the impact of workers’ aberrant behaviors on their productivity, which we aim to contribute to in a LMD context.

3.1. Theoretical Framework and Hypotheses Development

The LMD setting is typically characterized by granular audits and a plethora of alternative employment opportunities, potentially fueling a tendency to withhold effort i.e., shirk. This paper focuses on an interesting and quantifiable instance of shirking: “fake remarks” input by the DAs.

A DA might input a fake remark for various reasons, such as unwillingness to travel to a certain delivery area or having to cover a considerable distance for just a handful of parcels. While DAs possess the autonomy to determine the sequence of their delivery attempts, they do not have the discretion to not attempt a parcel. We argue that at times, DAs may choose not to attempt certain parcels due to ethical fading, a phenomenon wherein individuals focus on one aspect (such as effort reduction) to the extent that they

\(^{17}\)Robinson and Bennett (1995) classify behaviors such as excessive breaks or work slow downs as “production deviance”, a subcategory of aberrant behaviors, whereas Kidwell and Bennett (1993) define “shirking” as employees deliberately withholding effort for self-interest and opportunism.
overlook the ethical implications (like company policy) of their decisions (Tenbrunsel and Messick 2004). While ethical fading has been extensively studied in fields like psychology (Tenbrunsel et al. 2019), economics (Puaschunder 2022), and business ethics (Gunia 2019, Rees et al. 2019), we employ this relevant theoretical construct to shed light on the research question at hand.

Ethical fading is triggered particularly in complex organizational settings (Decrnis 2022) such as LMD firms. One facet of this complexity is behavioral invisibility (closely related to private information of the agents that leads to information asymmetry), which is defined as the lack of visibility in the actions of the DAs. The firms cannot monitor every action of a DA (e.g., what the customer said to the DA over a phone call). Therefore, parcels that the DA perceives as less likely to generate customer complaints and more likely to be successfully delivered on a subsequent attempt provide the necessary behavioral invisibility, facilitating misconduct. When a DA decides not to attempt a parcel, they have three options: record it as “not attempted”, take no action (we term these as “unattempted” parcels), or input a “fake remark”.

Unattempted parcels represent discretionary non-deliveries. However, when a DA inputs a fake remark, it breaches the code of conduct and hence we (and the firm) categorize it as a misconduct. Organizational misbehavior encompasses actions that intentionally violate expected conduct (Vardi and Wiener 1996), extending beyond illegal or unethical actions to include deviation from process requirements (Gabbioneta et al. 2019) and misreporting (Amar et al. 2022). Hence a fake remark is essentially a misconduct.

Center managers conduct daily corrective and preventive action (CAPA) meeting, where DAs account for failed deliveries and face reprimanded for any misconduct detected during audits or through customer complaints. Conducted in front of the co-workers, these discussions induce a sense of fear and shame that our interviews with DAs confirmed. Unattempted parcels, visible to the center managers, guarantee reprimand, while fake remarks might not get caught, thereby avoiding immediate reproach. Hence, despite harsher penalties for fake remarks, their uncertain, delayed nature, and the aversion to immediate reputational loss might sway DAs towards this choice. This aligns with theories suggesting individuals prefer a higher probabilistic loss over a lower certain loss (Tversky and Kahneman 1992), and may act unethically for short-term personal gains (Gino 2015). Hence, DAs prefer entering a fake remark over leaving the parcel unattempted. Both fake remarks and unattempted parcels decrease same-day successful

---

18 This is part of the Quality Management System and is an integral part of ISO certification (link).
19 “At the end of the day, I have to provide explanations to the center manager for all the undelivered parcels in front of everyone. Usually, I work for 10-11 hours because the route is mostly long, and the number of parcels is high. I don’t get time to rest. We are expected to attempt all the parcels irrespective of the time spent on duty” and “We get scolded for pending items, so I have to attempt all.”
20 Loss means reputational loss in our context.
deliveries and create backlogs. With the need to conceal misconduct, DAs prioritize fake remarked parcels, strategically reallocating effort (Larkin and Pierce 2016).

Principal-agent theory is used in operations management to understand strategic interactions (Leider 2018). According to this theory, the agent’s (in this case, the DA’s) effort to deliver a parcel creates value for the principal (LMD firm), despite the effort being a disutility to the agent (Arrow 1985). Therefore, the DA aims to minimize effort and maximize incentives. Misconduct is a form of agency problem in gig work settings (Burbano and Chiles 2022), often influenced more by social norms and the saliency of misconduct than cost-benefit analysis (Gino et al. 2009). Fake remarks reduce the DAs effort, enhance flexibility while not affecting financial incentives due to a compensation structure involving fixed pay and minor additional pay for successful deliveries. This provides short-term benefits and encourages myopic DA behavior. Given the negligible impact of misconduct on incentives in our setting, we are able to cleanly identify misconduct’s effect on performance irrespective of the incentives.

In our context, we measure the productivity of a DA by the number of successful deliveries, a key productivity metric for any LMD operation. The DA’s strategic behavior in effort reallocation is primarily driven by a need to conceal misconduct. Therefore, the DA might make suboptimal decisions regarding her delivery schedule on the following day, allocating extra time and effort towards the fake remarked deliveries. Considering the suboptimal effort allocation, we predict that fake remarks will create a negative spillover effect and adversely affect future productivity. Formally stated:

**Hypothesis 1.** Fake remarks will have a negative impact on future productivity.

Our data does not capture many behavioral traits (e.g., the willingness to enter a fake remark) as well as private information of the DAs. Therefore, we conducted interviews with DAs to understand how their behavior is motivated by their experience and information not available to the firm.

The interview snippets suggest that a DA uses their past experiences with customers and the delivery area to strategically order their deliveries. This also provides grounds for strategically deciding which parcels to fake remark upon, which they will prioritize the following day. Therefore, the DA, who aims to avoid shame and reprimands, is inclined to deliver these parcels diligently. To improve the success of such deliveries, the DA is likely to put extra effort into delivering undetected fake remarked parcels, to conceal their misconduct. Ackerman et al. (2020) argue that individual assessment of a task’s disutility influences the degree of effort, shaped by factors like personality, interest, and the job’s associated

---

21 “We have an idea that some customers will not accept the parcel. I have two such parcels today as well – one from c197 and another from d116. These two customers never accept the parcels. But they order frequently. I have to go all the way to their floor, I get tired, and my time is wasted. They don’t even pick up the call” and “I know for which customer I can go directly to their address and for whom I need to call 5-10 minutes before I reach there.”
punishment/rewards. Similarly, KC et al. (2020) underscore the importance of understanding workers’ task completion preferences, particularly under stress or threat conditions, when they have the autonomy to manage their tasks. Given two parcels, one fake remarked and one fresh, the DA will reallocate effort to conceal mistakes, thus exerting more effort in the fake remarked parcels. Therefore, we posit that fake remarks will increase the success of fake remarked reattempt parcels on the subsequent day.

Furthermore, in their attempt to manage backlogs and conceal misconduct, DAs divert effort from fresh parcels, creating a negative spillover from fake remarked parcels. As Larkin and Pierce (2016) explain, workers divert effort from other tasks to cover up misconduct. The increase in fake remarks require more effort/time from DAs, which we anticipate will negatively impact FTR. We state these two hypotheses that predict the effect of fake remarks on productivity of fake remarked reattempt and FTR deliveries as follows:

**Hypothesis 2.** An increase in fake remarks will have a positive impact on future fake remarked reattempt success.

**Hypothesis 3.** An increase in fake remarks will have a negative impact on future first-time-right (FTR) delivery success.

“Opportunism is a behavior that is motivated by self-interest and takes advantage of relevant knowledge asymmetry to achieve own gains, regardless of the principles,” as defined by Luo and Meyer (2017). Williamson (1993) suggests that “economic agents be described as opportunistic, where this contemplates self-interest seeking with guile.”

**Figure 2: Schematic Theoretical Framework**

![Figure 2: Schematic Theoretical Framework](image)

*Note: Misconduct becomes favorable for actions that exhibit higher behavioral invisibility, i.e., lower probability of detecting misconducts, resulting in an increase in fake remarks. This increase in fake remarks leads to a rise in the number of fake-remarked deliveries on subsequent days, which in turn amplifies clearance pressure. The desire to conceal fake deliveries can lead to a diversion of effort from other deliveries, increasing productivity for fake remarking.***
remarked deliveries (H2) but decreasing productivity for other deliveries (H3). Ultimately, this disproportionate effort allocation can lead to an overall decrease in productivity (H1). The magnitude of the effect of fake remarks is moderated by opportunistic circumstances that increase behavioral invisibility (H4).

Multiple factors can facilitate opportunism. Ping (1993) and Schwartz and Hirschman (1972) correlate opportunism with a low job-switching cost, while Alchian and Demsetz (1972) suggest that information asymmetry or insufficient monitoring facilitates shirking. Therefore, we define opportunistic circumstances as those amplifying knowledge asymmetry and enabling opportunism. In our context, opportunistic circumstances are those where entering a fake remark is easier and the probability of getting caught is lower. This follows the argument of Gino (2015) – that the more opportunities to justify, the more likely an agent will engage in unethical behavior, and Decrnis (2022), who asserts that lower detection risk prompts more misconduct. Thus, we contend that circumstances and environments that increase knowledge asymmetry will facilitate opportunism.

Considering DAs as opportunistic agents, we argue that opportunistic circumstances will enhance their opportunism, which will exacerbate (negatively moderate) the impact of fake remarks on productivity losses. Consequently, we hypothesize:

**Hypothesis 4.** Opportunistic circumstances will negatively moderate the effect of fake remarks on future productivity.

We illustrate the theoretical framework and related hypotheses (H1 to H4) in Figure 2. It summarizes how behavioral invisibility provides grounds for misconduct, how such misconduct leads to strategic effort reallocation, and how this effort reallocation affects productivity. Finally, it illustrates how opportunistic circumstances may moderate this effect.

4. **Data and Model**

Collaborating with one of the largest LMD firms in India, we obtain our dataset that consists of granular observations for six months (Jun – Nov 2019) from ~900 DAs across nine different cities (both metros and non-metros). During this period, these DAs handled 3.4 million parcels across 71 delivery centers.

Our dataset includes four categories of information: i) **dispatch details** ii) **parcel traces** iii) **DA details**, and iv) **delivery center information. Dispatch data** provides a macro view, specifying the assigned DA for a dispatch, parcel counts within a dispatch, and the aggregate number of parcels canceled, delivered, pending, etc. **Parcel traces** offer a micro view, detailing the promised delivery date for each parcel, DA interactions with each consignee (call duration), whether the firm previously served the consignee’s address, DA location traces with timestamps for the entire day, and the remark entered by the DA for each parcel. The **DA details** include information such as joining date, work area, and shift timings, while the **delivery center** table provides the dispatch center location and related information.
In addition to the delivery details, we have obtained data from the DAs’ phone GPS traces featuring more than 135 million location traces. This is, to our knowledge, among the most extensive datasets in last-mile logistics research. This helped in calculating key behaviors of DAs such as idle time, time to first attempt, and so on. We also used it to validate fake remarks identification.

4.1. Definitions of key variables

We created a day-level panel by aggregating dispatch-level data, where each row corresponds to a DA working on a particular day. Table 1 presents the description of our variables including our dependent variables: *success*, *FTR*, *re_success*, *fr_success*, and *non_fr_success*; main treatment variable: *fake_remark*, and control variables. One should note that *success* is the sum of *FTR* and *re_success*, where *re_success* is the sum of *fr_success* and *non_fr_success*.

We also include several control variables that could potentially affect our dependent variables. Given that workload (KC et al. 2020) and worker availability (Xu et al. 2020) can affect worker behavior, we control for the total number of parcels assigned to the DA and total number of workers. Additionally, we account for variables such as parcel types – fresh, reattempts, or cash on delivery, the DA’s idle time,22 and the number of unattempted parcels, all of which serve as indicators of the DA’s behavior. In sum, our control variables include *parcels*, *da_count*, *first_attempt*, *unattempted*, *idle_time*, and *cod*.

---

22Idle time is defined as the time during which a DA does not maneuver on her vehicle between two deliveries. The firm allows up to 15 minutes time per delivery for reasons like walking between apartments in the same building, waiting at the door, and so on, after the DA has reached the location. In all other cases, idle time is calculated using GPS traces. To do so, the duration for which the DA’s vehicle moves slower than a threshold speed of 6 km/hr (Pons et al. 2019) is classified as idle time. The firm calculates idle time as the proportion of a rolling 5-minute time window in which the DA is moving at a speed < 6km/hr based on the distance traveled by the DA.

\[
idle_time = \begin{cases} 
1 - \left( \frac{\text{speed of FE in km/hr}}{6} \right), & \text{if speed < 6km/hr} \\
0, & \text{otherwise}
\end{cases}
\]
Table 1: Descriptive Statistics for Panel Data

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Min</th>
<th>Mean (SD)</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>success</td>
<td>number of parcels successfully delivered by the DA in a day, representing</td>
<td>6.00</td>
<td>38.94 (19.14)</td>
<td>129.00</td>
</tr>
<tr>
<td></td>
<td>the main dependent variable</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FTR</td>
<td>number of parcels successfully delivered on the first attempt (first-time-</td>
<td>0.00</td>
<td>36.67 (18.45)</td>
<td>127.00</td>
</tr>
<tr>
<td></td>
<td>right deliveries) by the DA in a day</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>re_success</td>
<td>number of parcels that were previously unsuccessful but successfully</td>
<td>0.00</td>
<td>2.23 (2.69)</td>
<td>40.00</td>
</tr>
<tr>
<td></td>
<td>delivered by the DA in a day, i.e., reattempts that were successfully</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>delivered</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fr_success</td>
<td>number of parcels that were previously fake marked but successfully</td>
<td>0.00</td>
<td>0.31 (0.88)</td>
<td>24.00</td>
</tr>
<tr>
<td></td>
<td>delivered by the DA in a day, i.e., the total number of fake marked</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>reattempt success</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>non_fr_success</td>
<td>number of parcels that were previously fake marked</td>
<td>0.00</td>
<td>1.92 (2.36)</td>
<td>36.00</td>
</tr>
<tr>
<td></td>
<td>but successfully delivered by the DA in a day, i.e., the total number of</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>non-fake marked reattempt successes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>genuine_failures</td>
<td>number of parcels not delivered due to genuine reasons</td>
<td>1.00</td>
<td>19.62 (12.70)</td>
<td>118</td>
</tr>
<tr>
<td>fake_remark</td>
<td>number of fake remarks entered by the DA in a day, the primary treatment</td>
<td>0.00</td>
<td>2.70 (4.00)</td>
<td>41.24</td>
</tr>
<tr>
<td></td>
<td>variable of interest</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>parcels</td>
<td>number of parcels allocated to the DA in a day</td>
<td>15.00</td>
<td>63.01 (24.50)</td>
<td>149.00</td>
</tr>
<tr>
<td>da_count</td>
<td>number of DAs working at the same local delivery center as the focal DA in</td>
<td>1.00</td>
<td>6.11 (2.75)</td>
<td>17.00</td>
</tr>
<tr>
<td></td>
<td>a day</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>idle_time</td>
<td>total idle time (minutes) in a day</td>
<td>0.00</td>
<td>1.75 (2.79)</td>
<td>34.00</td>
</tr>
<tr>
<td>idle_time</td>
<td>total idle time (minutes) in a day</td>
<td>0.00</td>
<td>38.4 (27.13)</td>
<td>228.83</td>
</tr>
<tr>
<td>idle_time</td>
<td>total idle time (minutes) in a day</td>
<td>0.00</td>
<td>28.37 (14.67)</td>
<td>113.00</td>
</tr>
</tbody>
</table>

Note: N=46,085

4.2 Summary statistics and correlations

Table 1 provides the descriptive statistics for the primary variables. We only consider DAs that have worked on more than 6 dispatches in the six months that we observe. The final panel is unbalanced and includes 852 DAs who worked on nearly 2.9 million parcels over this six-month period, yielding 46,085 DA-day observations. On average, a DA works 54 days during our data period, with a mean tenure of 232 days and 25 days off. They often work four-day stretches without breaks and handle about 63 parcels daily, including new and reattempt deliveries. Nearly two-thirds (38.94) of these parcels are delivered successfully, with 94% of these delivered on the first attempt. We provide a further breakdown of delivery attempt outcomes in Figure A.2. They work an 8-hour shift, essentially delivering a parcel every 7.6 minutes. Note that many addresses have multiple items, allowing for quicker deliveries. Moreover, the
distance between about 43% of parcels in a dispatch is less than 500 meters (which is an artefact of dedicated assigned areas for each DA).

We observe an average of 2.7 fake remarks per DA daily, although this number can occasionally reach as high as 41. During our study period, 98.3% of the DAs enter at least one fake remark, and a total of 1.77% of parcels are fake remarked. DAs operate in pre-defined work areas, receiving parcels specific to that area. On average 88% of the assigned parcels belong to previously visited sub-areas. Also, 45% of the parcels required the DA to collect cash on delivery. Pairwise correlations reported in Table A.2 reveal no issues of multi-collinearity in our data. In Figure A.3, we provide the distributions of all variables.

4.3. Econometric Model & Identification

4.3.1. Econometric Model

We identify the impact of fake remarked parcels on the next day’s productivity of the DA using a fixed effect linear regression model:

\[ DV_{it} = \beta_1 fake_{remark_{i(t-1)}} + \beta_2 controls_{it} + fe_t + date_t + c_i w_t + \epsilon_{it} \]  \hspace{1cm} (1)

In equation (1), index \( i \) represents a DA, and \( t \) denotes a day. \( DV_{it} \) represents our dependent variables: success, FTR, fr_success, and non_fr_success. Using the DV success, we estimate the impact of fake remarks on the next day’s delivery success, while the other three DVs provide the effect on specific components of success.

\( fake_{remark_{i(t-1)}} \), our treatment variable, is lagged by one day for each DA, allowing us to estimate the impact of the previous day’s fake remarks on today’s productivity. As described in section 4.1, we incorporate several controls that could affect the delivery success, specifically parcels, da_count, first_attempt, and cod. In addition to these contemporaneous variables, we control for lagged variables, lag.unattempted and lag.idle_time, that could impact the next day’s success. We also control for time-invariant and time-variant fixed effects, i.e., DA (\( fe_i \)), date (\( date_t \)), and center-week (\( c_i w_t \)) fixed effects.

4.3.2. Endogeneity

The decision to enter a fake remark is likely endogenous because of the archival nature of our data. Several threats to the identification of \( \beta_1 \) exist, including sample selection, unobserved heterogeneity, reverse causality, and omitted correlated variables (Antonakis et al. 2014). These are discussed below, along with our countermeasures.

Selection Bias – One may argue that there might be an endogenous sorting of DAs across delivery centers. Specifically, DAs more likely to enter fake remarks could choose centers in areas where completing a delivery is more challenging. However, this concern is mitigated due to the operating style of our partnered LMD firm. Delivery centers are strategically located within a city, employing DAs who are
familiar with the service area of that center. These delivery areas do not overlap, limiting the choices for DAs to select a delivery center. We control for DA fixed effects, which absorb time-invariant factors such as the center in which the DA is employed.

**Reverse Causality** – While productivity or success is dependent on fake remarks, success can likewise affect a DA’s actions. To mitigate this, we use a temporal lag, a common approach in management (Ghose and Han 2011) and economics (Bollinger et al. 2020) research. We study the impact of the previous day’s fake remarks on the next day’s productivity. Moreover, DAs are not privy to tomorrow’s parcel assignments, precluding them from entering fake remarks in anticipation of tomorrow’s success.

**Omitted Variable Bias** – Certain DA- and time-varying factors that correlate with both our DV and our treatment variable (*fake_remark*) might be unobservable, posing a risk of inducing omitted variable bias. For instance, “strategically selecting parcels for fake remarks based on behavioral invisibility” is an omitted variable that may impact both the DA’s productivity and the likelihood of entering a fake remark. Our interviews and data analysis, discussed in section 5, provide evidence of such strategic behavior.

Figure 3 illustrates this omitted strategic behavior, which we expect to positively affect future productivity and propensity to enter fake remarks. If this strategic behavior is not accounted for, its positive correlation with subsequent success will be absorbed in \( \beta_1 \), resulting in an underestimation of the effect of fake remarks on productivity. An alternative source of omitted variable bias may be that DAs tend to enter fake remarks under conditions that might reduce future success (e.g., future traffic congestion in an area). Following a similar reasoning, this might overestimate the impact of fake remarks on productivity. However, our analysis does not support this assertion, as we observe stronger evidence in favor of the strategic behavior mechanism depicted in Figure 3.

We address these biases using instrumental variable (IV) regression and employing a two-stage least square (2SLS) regression in the next sub-section. In section 7, we conduct several robustness tests and alternative specifications, such as using propensity score matching, a nonlinear model with a control function, or applying log transformations to our outcome variables, to ensure the consistency of our results.

---

23 “Customers of reputed e-commerce merchants like Amazon are easily available and accept easily. For them, I don’t feel the need to make a call. For parcels from other merchants, it is always better to call the customer beforehand.”

24 We use the 2SLS approach based on the recommendations of Angrist and Krueger (2001). They suggest that even if the underlying relationship is non-linear, linear estimation methods like 2SLS can still provide robust estimates for instrumental variable applications, reducing the risks associated with misspecification.
Unobserved Heterogeneity – DAs and delivery centers may be heterogeneous in many ways. To address this, we control for three sets of fixed effects. We use DA and date fixed effects to account for time-invariant heterogeneity. Following Bai (2009), we include center-week fixed effects to capture time-varying unobservables. We opt for week-level interactive fixed effects because weekly reports and audits are in place to analyze and course-correct center performance.

Serial Correlation – A DA’s fake remarks may exhibit serial correlation, or there could be serially correlated unobservables such as the DA’s effort, requiring control measures to accurately identify the impact. Using time variant fixed effects (center-week in our case) can somewhat alleviate this problem (Bollinger and Gillingham 2012), but a better method in dynamic panel models involves incorporating a lagged dependent variable (Angrist and Pischke 2008, Leszczensky and Wolbring 2022). However, this can induce bias (Nickell 1981). Therefore, we adopt Angrist and Pischke's (2008) estimation strategy, which utilizes the bracketing property of fixed effects and lagged dependent variables estimates. This approach is commonly used in empirical studies (Falk et al. 2018). The robustness of our results is verified using this bracketing property in Appendix A1.7.

4.3.3. Instrumental variables

Although, there is no established empirical test for endogeneity, we conduct the Wu-Hausman test and fail to reject the null hypothesis that the fake_remark variable is endogenous. We use the fake remarks entered by the co-workers of the focal DA as an instrument for the endogenous variable fake_remark. Norton et al. (2003) suggest that the peers or co-workers (people whom we identify with) can stimulate behavioral change. Misconduct and peer effects, while not extensively studied in OM, have been well-studied in various other research streams (Gino et al. 2009, Pierce and Snyder 2008). Complementing this literature, our interviews also suggest substantial peer interaction at the end of day. As a result, we draw on the peer effects literature to construct our instrument. Similarly, co-workers’ incentives (Sinchaisri et al. 2019) and earnings (Xu et al. 2020) have been used as valid instruments in the literature.

25 Given that 87% of DAs worked at only one center, the center is largely time-invariant, and hence, DA fixed effects take care of center-level heterogeneity.
We identify co-workers as DAs working in the same delivery center as our focal DA. Given DAs have different days off during the week, and may take occasional leaves, we establish the co-worker set for each DA for each day, including only those DAs who are working on that day. Let $C_t(i)$ be the set of co-workers of DA $i$ on day $t$, excluding the DA $i$. The number of co-workers of DA $i$ on day $t$ is denoted by $|C_t(i)|$. Then, we define the variable $\text{co}_fr\_avg_{it}$ (mean = 2.7, sd = 1.97) as the past seven-day moving average of the fake remarks per co-worker of our focal DA $i$ starting from day $t$, i.e.,

$$\text{co}_fr\_avg_{it} = \frac{1}{7} \sum_{k=t-6}^{t} \frac{1}{|C_k(i)|} \left( \sum_{j \in C_k(i)} \text{fake\_remark}_{jk} \right)$$ (2)

We provide robustness checks in section 7 using different windows for our moving average such as 1, 3, and 10 days, which yield consistent results.

As shown in Figure 4, to estimate equation (1) we use $\text{co}_fr\_avg_{i(t-2)}$ as an instrument for the focal DA’s fake remarks on day $t-1$ i.e., $\text{fake\_remark}_{i(t-1)}$. First, we estimate an OLS model with the endogenous variable $\text{fake\_remark}_{i(t-1)}$ (fake remarks of the focal DA on day $(t-1)$) as a dependent variable, employing the instrument $\text{co}_fr\_avg_{i(t-2)}$. From this, we obtain a predicted value of $\text{fake\_remark}_{i(t-1)}$ as $\text{fake\_remark}_{i(t-1)}$. Then, using this predicted value, we estimate the effect of fake remarks on DA productivity by employing different dependent variables, such as success and FTR.

Figure 4: Co-worker Fake Remarks as an Instrumental Variable

To ensure robust identification, we need to address potential bias inherent in peer effects settings. The main challenges in studying peer effects include reverse causality, correlated unobservables, and selection bias (also known as endogenous group formation) (refer to Bollinger and Gillingham 2012, Bramoullé et al. 2009, Manski 1993). As discussed in previous section, we address reverse causality by using lagged co-worker fake remarks (similar to Bollinger et al. 2020, Ghose and Han 2011), manage correlated unobservables with a rich set of fixed effects (similar to Hanushek et al. 2003, Oestreicher-Singer and Sundararajan 2012), and deem selection bias less pertinent as DAs cannot choose their center.
A valid instrument must satisfy three conditions for an unbiased estimate. The *exclusion restriction* requires the IV to be uncorrelated with the error term, and the *relevance condition* requires the IV to be correlated with the endogenous variable (Wooldridge 2002). The IV should influence the dependent variable only via the endogenous variable. Lastly, the instrument needs to be sufficiently strong to provide an unbiased estimate of the effect through instrumental variable regression (Stock and Yogo 2005). Our instrument satisfies these requirements, as confirmed by the first stage of 2SLS; it’s relevant (coefficient = -0.402, \(p<0.01\)) and strong (F-stat = 273.3).

The exclusion restriction stipulates that our IV influences the performance of the focal worker solely through its impact on the worker’s fake remarks. We argue that a co-worker’s behavior cannot directly influence the future successful deliveries of focal workers because these workers operate in distinct work areas and the arrival of parcels is random—driven by consumer demand rather than DA decisions (as discussed in section 2). DAs have pre-defined work areas that are assigned an ample workload, thereby preventing interference among the DAs. The distribution of parcels to individual DAs is not dictated by the combination of fresh and carryover parcels; rather, it is primarily determined by the demand specific to each area. While the center manager endeavors to distribute the load of fresh and carryover parcels evenly among the DAs on average, the pre-defined work areas for DAs limit the manager’s ability to exercise much discretion. The areas are designed with the expectation of similar workloads, enabling DAs to earn comparable salaries for similar efforts. Our interviews with DAs substantiate this claim. For a comprehensive analysis and description, we direct readers to Appendix A1.5. We also discuss the direction of the peer effect that we identified in Appendix A1.6. Lastly, to handle serial correlation, we undertake a robustness test with lagged dependent variable estimation, the results of which remain consistent and are discussed in Appendix A1.7.

5. **Results**

We present the estimation results of the impact of fake remarks on productivity.

5.1. **Effect of Misconduct**

We provide results of our estimation of equation (1) in Table 2. Column 1, shows the results using only the control variables. We observe that an increase in the parcel count assigned to the DA increases successful deliveries (0.439, \(p<0.01\)). However, the unattempted parcels from the previous day do not significantly affect today’s productivity. We also find that an extended duration to attempt the first parcel in the morning correlates with lower productivity (-0.020, \(p<0.01\)), while a restful or more idle previous day enhances productivity (0.006, \(p<0.01\)). Cash on delivery positively impacts successful deliveries (0.233, \(p<0.01\)). We will delve into cash on delivery parcels in section 6. Column 2 includes our treatment
variable \textit{lag.fake\_remark} and estimates the model using OLS. Our data suggests the previous day’s fake remarks do not affect today’s productivity. However, this finding suffers from the aforementioned biases.

<table>
<thead>
<tr>
<th>Table 2: Effect of Fake Remarks on Productivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{success}</td>
</tr>
<tr>
<td>\textit{lag.fake_remark}</td>
</tr>
<tr>
<td>\textit{parcels}</td>
</tr>
<tr>
<td>\textit{da_count}</td>
</tr>
<tr>
<td>\textit{first_attempt}</td>
</tr>
<tr>
<td>\textit{lag.unattempted}</td>
</tr>
<tr>
<td>\textit{lag.idle_time}</td>
</tr>
<tr>
<td>\textit{cod}</td>
</tr>
<tr>
<td>\textit{IV}</td>
</tr>
<tr>
<td>\textit{adj-R^2}</td>
</tr>
<tr>
<td>\textit{N}</td>
</tr>
</tbody>
</table>

\textit{Note: *** p<0.01, ** p<0.05, * p<0.1; Standard errors are in parentheses and clustered at the DA level. Estimated with da, date, and center-week fixed effects.}

In column 3, we use 2SLS to estimate equation (1), with the first stage results reported in Table A.7. Our IV, \textit{lag2.co\_fr\_avg}, is significant and negatively associated to \textit{lag.fake\_remark} (-0.402, p<0.01). The F-stat 273.3 suggests that our instrument is strong. This confirms the relevance condition of our instrument. The estimate suggests that an increase in average fake remarks by co-workers decreases the focal DA’s fake remarks. A plausible explanation is the strategic response of the focal DA, who might fear heightened scrutiny in case of excessive fake remarks at the center. This mirrors findings by Chan et al. (2021) in a misconduct context, where workers are less likely to engage in restaurant theft if their co-workers exhibit higher levels of such behavior on a given day, termed ‘strategic peer response’. Negative peer effects have also been identified in several studies across psychology and economics (Angrist 2014, Brady et al. 2017, Rauhut 2013, Schreiner and Bremer 2013) with examples from marathon racing (Emerson and Hill 2018) and labor market unethical behavior when under supervision or reputation risk (Pascual-Ezama et al. 2015).
In the second stage, our estimates suggest that \textit{lag.fake_remark} negatively affects \((-0.625, p<0.01)\) productivity,\footnote{Angrist and Kolesár (2021) explain that using instrumental variable to counteract omitted variable bias occasionally results in a sign and significance flip. Several studies in both economics (Coe et al. 2012, Friedberg 2001) and management literature (Gopalakrishnan et al. 2022, Song et al. 2020, Tan and Netessine 2014) provide examples of econometric analysis where the sign and significance of the endogenous variable morphs from OLS to 2SLS regression analysis. Friedberg (2001) further elucidates how the direction and magnitude of bias yields divergent estimates between OLS and 2SLS estimation. In our case, we anticipate a positive bias given the positive correlation between strategic behavior (i.e., omitted variable) and both the DA’s fake remarks and the DA’s success.} supporting our H1. A unit increase in fake remarks results in a 0.625 decrease in the next day’s successful deliveries. On average, a DA is assigned 63.01 deliveries, successfully delivering 38.94. A decrease of 0.625 equals a daily reduction of 1.6\% (0.625/38.94) successful deliveries. During the study, the company was delivering an average of 1.5 million parcels per day. Therefore, an additional fake remark could result in a spillover effect of 24,000 (0.016×1.5 million) fewer successful deliveries each subsequent day, totalling 8.8 million annually. This significant economic impact should be considered in addition to same-day failed deliveries caused by fake remarks.

After estimating the effect of fake remarks on next day successful deliveries, we want to test if fake remarks lead to a prolonged cascading effect on productivity. To do so, we estimate the effect of the previous days’ fake remarks (days t-1 to t-4) on today’s (day t) success. These results have been reported in Table A.8 and illustrated in Figure A.5. We find that today’s success is affected by fake remarks made over the past three days, with the effect size diminishing progressively until day (t-4) where it becomes insignificant. Aggregating this effect \((0.625+0.611+0.538 = 1.774)\), we deduce that success on day t is decreased by an average of 1.774 deliveries due to the previous days’ fake remarks. Given that the DA is assigned 63 parcels and manages to successfully deliver 39 of them, a decrease of 1.774 corresponds to an average daily decline of 4.5\% \((1.774 / 39)\) in successful deliveries.

In our case, the coefficient for fake remarks becoming significant and negative as we transition from OLS to 2SLS. Introducing an instrumental variable could potentially alter the direction and significance of other exogenous variables, especially if the endogenous variable correlates with other variables included in the model. Tan and Netessine (2014) have reported similar results. In our scenario, the correlation between \textit{lag.fake_remark} and \textit{lag.unattempted} is 0.22, which explains why we observe a change in the direction and significance of the lag.unattempted coefficient when we move from OLS to 2SLS.

5.2. Effect of Misconduct on Different Types of Deliveries

To test H2 and H3, we use data concerning parcels that were fake remarked earlier and reattempted today, in addition to the first-time-right parcels. We identify three distinct parcel categories (fresh parcels, fake remarked reattempts, and genuine reattempts), effectively decomposing \textit{success} into its three
components. We evaluate the effect of fake remarks on these specific delivery types by estimating equation (1) with each of the three outcomes serving as dependent variables. These results are presented in Table 3.

In column 1 of Table 3, we observe that \( fr_{success} \) has a positive coefficient of 0.120 (\( p<0.01 \)), indicating that parcels previously fake remarked by the DA are more likely to be successfully delivered the following day, thus supporting H2.

To test H3, we use \( FTR \) as our dependent variable in equation (1). In column 2, we estimate the effect of an increase in the previous day’s fake remarks on today’s fresh parcels. Using the same IV, we find that the coefficient is significant and negative (-0.682, \( p<0.01 \)), i.e., an increase in fake remarks decreases \( FTR \) deliveries by 1.86% (or 0.682/36.67) on the subsequent day. Thus, \( FTR \) deliveries are negatively impacted by the increase in fake remarks, corroborating H3.

Column 3 presents the estimate for equation (1) with \( non_{fr_{success}} \) as the dependent variable, which accounts for the delivery success of parcels genuinely failed earlier. The coefficient is significant at the 10% level, indicating that an increase in fake remarks detrimentally impacts the success of these parcels (-0.086, \( p=0.055 \)). This suggests DAs exert more effort to deliver fake remarked reattempt parcels, thereby negatively affecting their success for other parcels (fresh parcels and genuine reattempts). In summary, among the three parcel types, \( fr_{success} \) is positively affected, whereas \( FTR \) and \( non_{fr_{success}} \) suffer due to an increase in previous day’s fake remarks.

<table>
<thead>
<tr>
<th>Table 3: Effect of Fake Remarks on Different Delivery Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fake remarked success</td>
</tr>
<tr>
<td>( fr_{success} ) (1)</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>( lag_{fake_{remark}} )</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>( adj-R^2 )</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Note: *** \( p<0.01 \), ** \( p<0.05 \), * \( p<0.1 \); \( N=38,472 \); Standard errors are in parentheses and clustered at the DA level. Estimated with instrumental variable using 2SLS with all controls and da, date, and center-week fixed effects.

Two possible mechanisms can underlie these findings. First, the delivery center manager’s pressure to deliver the fake remarked parcels identified in the end-of-day audit might encourage DAs to prioritize these parcels. Second, DAs might anticipate which of today’s parcels are likely to be successfully delivered the subsequent day, thus strategically entering fake remarks for these parcels to alleviate today’s workload. We posit that both mechanisms are at play. However, we cannot distinctly examine an increase in managerial
oversight. Consequently, we focus on DA strategic behavior, supported by extant literature (discussed in section 3), data patterns (discussed below), and our interviews with DAs.27

DAs, based on their work area experience, discern which parcels have higher behavioral invisibility and tend to input fake remarks for such parcels. The greater the probability of success the following day, the lower the risk of discovery and penalty for this misconduct, and hence, a greater likelihood to input fake remarks. This explains the observed positive impact of fake remarks on the successful delivery of previously fake remarked parcels (fr_success) in Table 3.

However, as DAs tactically select fake remarked reattempts, their positive efforts on these parcels result in effort reallocation from fresh parcel deliveries, impacting overall productivity. Hence, the impact of fake remarks on FTR deliveries primarily determines a DA’s overall productivity loss. DAs are less concerned about the successes of genuine reattempts compared to FTR deliveries, but they are more cognizant of the reasons behind these parcels’ initial failure. With genuine reattempts, they can often coordinate with customers without additional effort (for instance, a customer may have requested next-day delivery during a genuine attempt). Thus, the information gained somewhat mitigates the negative effect of effort reallocation. This means that for genuine reattempts, there is a risk of wasted effort if the parcel remains undelivered compared to a fresh delivery. As a result, we observe a smaller decline in the success of genuine reattempts compared to FTR. DAs strategically prioritize fake remarked reattempts over genuine reattempts to conceal misconduct and often relegate FTR as their last priority. This approach helps in clearing backlog while reducing their likelihood of being caught for fake remarks.

In summary, our findings support our assertion that DAs can anticipate the chances of successful reattempts for fake remarked parcel. Table 3 demonstrates that fake remarks correlate with the DA’s knowledge of next-day success probability: fake remarked reattempt success increases while both FTR and genuine reattempt success decreases. Furthermore, as corroborative evidence, we note that the median time required to deliver a fake remarked reattempt parcel is 4.15 minutes, as compared to 5.25 minutes for genuine reattempts and 6.41 minutes for fresh parcels. These statistically distinct values affirm our hypothesis of fake remarks as strategic DA behavior and provide evidence of our mechanism at play.

5.3. **Causes of Misconduct and Impact on Platform**

We identified three potential predictor categories that could indicate a tendency to enter fake remarks: loyalty and professionalism, task attributes, and peer behavior. Our findings suggest that DAs with higher loyalty and professionalism – indicators such as longer tenure and less idle time – are associated with fewer

27 “There are areas where if the customer is not picking the call, the customer will not accept the parcel”, “I attempt the carryover parcels because if the customer was not available yesterday, he might not even accept today”, “Most of the DAs prioritize carryover parcels to get rid of the pending work” and “I call the next customer just before leaving the previous customer location that I will be coming to deliver your parcel.”
fake remarks. Vaughan (1999) associates task characteristics to misconduct, leading us to explore this connection. We observe that greater task complexity and workload variability corresponds with an increase in fake remarks. We investigate the influence of peer behavior on the likelihood of fake remarks. We find that, the average number of fake remarks made by co-workers and the proportion of co-workers making fake remarks negatively correlates with the focal DA’s fake remarks. This corroborates our mechanism that DAs strategically adjust their behavior to avoid detection, mindful of their peers’ conduct. Note that while these are not causal analyses, their associations still hold valuable insights for management. Detailed analyses can be found in the Tables A.9, A.10, and A.11.

Next, we investigate how fake remark behavior affects platform performance in the long-term. We label a DA as a “faker” if her fraction of fake remarked parcels exceeds the median. As reported in Table A.12, our finding suggests that despite working the same hours (mandated by the center) and covering the same distance, on average these “fakers” idle more and exhibit lower productivity.

In Table A.13, we conduct a sub-sample analysis that indicates that fakers make on average 4.2 fewer successful deliveries than non-fakers, equating to a 10% decrease in productivity (the DA successfully delivers 39 parcels). Thus, DAs with more fake remarks tend to complete fewer deliveries in the long run on the platform. Furthermore, fakers’ total work hours are not as productive as those of non-fakers.

Though fake remarked reattempts are not costly in terms of operating expenses, given there is no real attempt made, the number of attempts is inversely related to delivery success. Consequently, there’s a higher likelihood that a reattempted parcel may end up being returned to the merchant. This creates an added reverse logistics cost for the firm. Moreover, beyond causing customer dissatisfaction, these fake remarks damage the relationship between the LMD firm and its e-commerce partner, incurring further costs. We also observe a slight negative impact of these fake remarks on the success of genuine reattempts on the subsequent day. Genuine reattempts, in turn, carry significant costs in terms of fuel, labor, and time.

6. Effects of Opportunistic Environments

In addition to the main effect, we also examined "opportunistic circumstances" that could exacerbate the effect of fake remarks. These are situations where making fake remarks is easier and detection risk is low. Although precise measurement of opportunism is challenging, we identified indicators in the parcel delivery context, supported by literature (Bryson and Forth 2007, Mao et al. 2019) and practice.\(^{28}\) We specifically analyzed variables visited_before and cod for their interaction effects with fake remarks on DA productivity. We mean-centered cod, visited_before, lag.fake_remark, and lag2.co_fr_avg to mitigate multicollinearity and ascertain the average effect (Afshartous and Preston 2011, Coulton and Chow 1992).

\(^{28}\) https://tedium.co/2020/06/12/cod-cash-on-delivery-history/
Each parcel has a consignee address that belongs to a postal code. Our variable visited_before serves as a proxy for the DA’s familiarity with the delivery area, representing the total number of postal codes in the dispatch the DA has visited up until the previous day. When a DA becomes familiar with the delivery area, it becomes easier for her to shirk, justify lower productivity with plausible excuses, and allocate more effort in concealing prior fake remarks. First stage estimates reveal that the propensity to enter fake remarks has a positive association with visited_before (0.008, p<0.05) suggesting that DAs are more likely to enter fake remarks for parcels in a familiar area.

### Table 4: Effect of Familiarity on Productivity

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>lag.fake_remark_demean</td>
<td>-0.510***</td>
<td>0.092</td>
</tr>
<tr>
<td></td>
<td>(0.100)</td>
<td>(0.147)</td>
</tr>
<tr>
<td>lag.fake_remark_demean × visited_before_demean</td>
<td>-0.028***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td></td>
</tr>
<tr>
<td>visited_before_demean</td>
<td>0.866***</td>
<td>0.853***</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.009)</td>
</tr>
<tr>
<td>adj-R²</td>
<td>0.927</td>
<td>0.921</td>
</tr>
</tbody>
</table>

Note: *** p<0.01, ** p<0.05, * p<0.1; N=38,472; Standard errors are in parentheses and clustered at the DA level. Estimated with instrumental variable using 2SLS with all controls and da, date, and center-week fixed effects.

In Table 4, we report the second stage results. In column 2, we find that visited_before negatively moderates the effect of fake remarks (coeff. of lag.fake_remark_demean × visited_before_demean is -0.028, p<0.01). This suggests that the effect of fake remarks exacerbates as the DA becomes more acquainted with a delivery area, supporting H4. Contrary to literature suggesting that area familiarity increases delivery efficiency (Mao et al. 2019), our results indicate that if the DA has visited an address before, location familiarity can decrease success (in the presence of fake remarks).

We assess the robustness of our results using COD deliveries in Appendix A2.3. These results support our H4 that productivity loss due to aberrant behavior exacerbates in opportunistic circumstances.

### 7. Robustness Tests

We have conducted several robustness tests to ensure that our results are consistent. To test the robustness of our IV approach, we use matching as an alternate identification strategy. Matching is a

---

29 Considering that India’s nearly 19,100 postal codes cover approximately 2,973,190 square km, each postal code spans an average of 155 square km (radius of 7 km).

30 We do not use visited_before as a control in our main model due to its high correlation with parcels (0.88). Table A.14 provides the first stage of the model with visited_before as an additional control.
common identification tool in the management literature both as a main model (Lin et al. 2021, Terwiesch et al. 2020) and a robustness check (Bobroske et al. 2022, Choudhary et al. 2021, Wowak et al. 2021). We use two different matching methods - nearest neighbor (NN) matching (Abadie and Imbens 2011), and coarsened exact matching (Iacus et al. 2012), detailed in Table A.17, showing robust findings.

We conduct 2SLS with logarithmic transformations, a common approach in literature (Cachon et al. 2019, Goh et al. 2011), as a robustness check so that we can ensure that our results are not driven by the skewness of the dependent variable. As our DVs are count variables, in addition to 2SLS, we also implement the control function (CF) approach (Wooldridge 2015) using fixed effects OLS in the first stage and negative binomial regression in the second. The CF approach is often used as a robustness test for endogenous explanatory variable models (Guajardo et al. 2012, Kamalahmadi et al. 2021). We then bootstrap standard errors as suggested by Wooldridge31 and present the estimates in Figure A.6. The results remain robust with both the logarithmic transformation and the CF as shown in Table A.18.

Identifying all fake remarks with 100% accuracy, even ex-post, is impossible without auditing every delivery, a prohibitively expensive task. This is a common issue in most fraud detection applications (e.g., credit card fraud).32 Approximately 44% of fake remarks fall under the most conservative condition in Table A.1, where the DA has not reached the location nor called the customer before giving a remark. To ensure the robustness of our results, we perform a sub-sample analysis with deliveries classified as fake remarks solely based on the first condition, deeming all others as genuine failures. Table A.19 validates the directional robustness of our results.

To ascertain the robustness of our main result, we vary the length of the time window (1, 3, and 10 instead of 7) over which we calculate the peer effect (co-workers’ moving average fake remarks described in equation 2). In Table A.20, we report the results. Our results are qualitatively the same in terms of size and direction. The first stage estimates for each IV are produced in Table A.21. All of our alternative instruments remain relevant and strong.

Next, we use all the instruments together and conduct the Sargan test, which is frequently used as an overidentified exclusion restriction test in the literature (Moon et al. 2022, Wu et al. 2020). We fail to reject the null (stat =1.28851, p = 0.73) that all instruments are valid (Table A.22).

One could postulate that our results are influenced by the promised delivery date (pdd) since a delivery date far in the future might be more prone to misconduct. To mitigate this concern, we calculate the

---

32 Microsoft’s “Dynamics 365 Fraud Protection” application only decreased false positives by 0.36% as denominated by transaction count (Nanduri et al. 2020).
difference between the pdd and the current date in terms of days, and then compute the average (\textit{diff\_from\_pdd\_mean}). As demonstrated in column 1 of Table A.23, this deviation from the pdd does not have a correlation with the number of successful deliveries or instances of fake remarks. Subsequently, we examine whether the number of delivery attempts influences the frequency of fake remarks. We add the average number of attempts for all parcels in a dispatch (\textit{num\_attempt\_mean}). We find that an increase in attempts correlates with a decrease in successful deliveries (column 1 of Table A.23), and an increase in fake remarks (column 2 of Table A.23). We discuss this in detail in Appendix A3.6.

Attrition is a valid concern that can lead to selection bias (Elwert and Segarra 2022). We run the main model with a sub-sample of DAs who did not quit during our study period, i.e., DAs whose end date was post 30-Nov-2019. As shown in Table A.24, the results are robust, as the coefficient of \textit{fake\_remark} is \(-0.649\) (p-value = 0.06). Other tests for attrition have been provided in the Appendix A3.7.

We conducted two additional robustness tests (the tables have been omitted for brevity). Firstly, to ensure the effect is not driven by DA experience, we compare the effect for new and experienced DAs and find no significant difference. Lastly, we conducted analysis with an alternative dependent variable i.e., \textit{percent\_success} \(= \frac{\text{success}}{\text{parcels}}\), to address concerns regarding whether or not the effect could be workload-driven. Our results continue to hold.

\section{Discussion and Concluding Remarks}

Our study underscores the impact of employees’ misconduct on operational performance. We provide evidence of how misconduct can lead to workers’ effort reallocation, thereby creating a cascading effect on productivity. When workers strategically shirk on any given assignment, it also impacts their productivity on new assignments, due to the time and effort involved in concealing their misconduct. We collaborated with an Indian LMD firm, analyzing one of the largest datasets to date in LMD research using instrumental variable regression. We established the robustness of our results using methods such as matching, control functions, and many more. Assuming a nominal commission for each successful delivery, our results imply million-dollar annual losses for the LMD firm due to productivity spillovers resulting from fake remarks. Put plainly, these fake remarks could increase operating costs, carbon footprint, customer churn, and tarnish the relationship with e-commerce partners. Further, we establish the association of misconduct with task complexity, loyalty, and professionalism. The consequences of aberrant behavior in LMD extend beyond the immediate productivity loss discussed in the extant literature.

Although our investigation pertains to parcel delivery, the findings can be generalized to other settings such as grocery and courier mail delivery, where all the deliveries are assigned to the driver at the start of the day. The methodology we used to identify fake remarks errs on the side of caution. Thus, we believe that our estimates of the losses are conservative compared to the actual damage.
Our research recommends that LMD firms implement mitigation strategies against misconduct, which include addressing opportunistic circumstances, reducing task complexity, encouraging loyalty and professionalism, and monitoring employee behavior in light of peer actions. Ibanez et al. (2018) examined how employee discretion in task completion sequence impacts productivity in healthcare. A similar study in LMD might consider whether a DA’s autonomy in parcel delivery sequence affects their productivity. Understanding how DA productivity varies with parcels originating from different e-commerce platforms and how a DA’s perception of the consignee influences their behavior could be beneficial (Altman et al. 2021). We hope that our study will motivate future research into individual behavioral misconducts in service operations.

References


