# Making Reference-Dependent Preferences: Evidence from Door-to-Door Sales

Samuel Dodini*
Norwegian School of Economics

June 2023

**Abstract**

People often set goals that they fail to meet. One intuitive way of addressing this is to evaluate one's progress in smaller components. Theory suggests this should reveal reference-dependent preferences and loss aversion. This paper uses novel data from a sales company to test for reference-dependent daily labor supply as a commitment device to offset present bias for achieving longer-run goals. I show that daily labor supply shifts downward at a worker's expectations. Daily expectations are selected by workers based on long-run objectives around the bonuses paid by the firm at the end of the sales season. After surpassing their bonus threshold, workers reduce their hours from what was a consistent labor supply in their prior personal equilibrium, consistent with the bonus being the impetus behind daily reference dependence. An online real-effort experiment further supports the idea that short-run reference dependence can be "made" through a firm's compensation scheme. The experiment implies that this leads to greater firm profitability.

**Keywords:** reference dependence, loss aversion, non-linear compensation, goals

**JEL Codes:** D9, J22, J33, M52

*Email: samuel.dodini@nhh.no. I am grateful to Michael Lovenheim, Maria Fitzpatrick, Evan Riehl, Ted O'Donaghue, Alex Rees-Jones, Seth Sanders, Sam Hirshman, Bertil Tungodden, and Devin Pope for helpful comments, guidance, and support with the experimental design. Thank you to Sebastian Fest for support in implementing the online experiment. I am also thankful to seminar participants at the Cornell Behavioral Economics Workshop, Cornell Labor Economics Workshop, and Cornell Labor Work in Progress Seminar for helpful comments and discussion. This project was partially funded by the Research Council of Norway through its Centers of Excellence Scheme, FAIR project no. 262675.

# 1 Introduction

People often use goals to plan and manage their motivation. However, they often fail to reach their goals due to motivational problems stemming from present bias (Dellavigna, 2009). One possible approach to addressing these issues is to subdivide a goal into components and evaluate one's performance in a smaller window or "narrow bracket" such as a single session of the task or a short time frame. In other words, people may set a short-run goal. In the language of Prospect Theory (Kahneman and Tversky, 1979), these goals then act as reference points.

The motivational power of narrow goal setting relies upon people behaving in ways consistent with reference-dependent preferences (Koch and Nafziger, 2016; 2020; Imas et al., 2017). That is, a person's utility will depend not only on her absolute performance but also on where her performance stands relative to some mental target—in this case, the goal. Under loss aversion, when the person is operating below that mental target, the negative comparison to her target induces lower utility. However, her marginal utility is higher than if she were operating above her target. All else equal, this higher marginal utility leads her to exert more effort until she reaches the target. This means that in creating reference points, behavioral agents can use self-imposed, psychological costs in the short term to overcome self-control problems in the long term.

While the idea of using short-term goals as reference points seems intuitive, there is limited field evidence of what this looks like in practice, the factors influencing the formation of narrow-bracketed reference points in the real world, and how these interact with people's long-run objectives—the possible "whys" of reference dependence. The empirical literature on taxi drivers (e.g. Camerer et al. (1997); Crawford and Meng (2011); Thakral and Tô (2021)) that has come to define much of our understanding provides no evidence concerning this. The prior literature examining goals as reference points thus far has been limited to analyzing distortions in the distributions of final performance around a target in settings like marathon running (e.g. Allen et al. (2017); Markle et al. (2018)) or in firms (e.g. (Freeman et al., 2019; Kuhn and Yu, 2021; Cai et al., 2022)). These studies have not explored the person's important day-to-day choices in pursuit of the larger goal. In other words, we know little about how people might use short-run goals to achieve broader targets.

To address this gap, this paper investigates reference dependence and goal setting in a new context: door-to-door sales. By analyzing high-frequency data from a company that employs fixed-term, commission-based sales contractors and conducting an online real-effort task experiment, this study makes two main contributions to the literature. First, I establish the baseline observation that workers do exhibit reference-dependent labor supply in a novel setting. I test for this on two margins: the extensive margin (the choice to stop working for the day) as well as the intensive or "exertion" margin (effort conditional on working). Prior studies have almost exclusively analyzed the choice of when to stop working, but my data allow me to also test other aspects of labor supply. I provide clear evidence that door-to-door sales workers exhibit loss aversion around expectations in their extensive margin labor supply choices. However, they do not appear to do the same in their exertion conditional on continuing to work.

Second, I provide evidence that this loss aversion around daily expectations interacts with and is induced by the firm through its non-linear bonus scheme, which is realized only at the end of the sales season (the long run). From a theoretical perspective, we expect a relationship between daily expectations and end-of-season bonuses based on Kőszegi and Rabin (2006) (the KR model): when a worker plans for what she perceives to be the optimal path forward based on expectations about the future, the planned choice becomes her reference. In other words, workers choose a feasible long-run objective (or goal) and then subdivide it into short-run objectives and expectations. Kőszegi and Rabin (2006) call this "personal equilibrium" and posit that the firm may have a substantial influence on a worker's expectations, though empirical support for this claim is sparse in the literature. Koch and Nafziger (2016; 2020) theorize that this behavior is designed to combat present bias in multi-period tasks. This stands in contrast to the standard model, which assumes rationality in intertemporal utility affecting long-run performance (i.e., no self-control problems and, therefore, no need for short-run goals) and no utility responses to perceived reference points in the short run.

The prior empirical literature has not considered the interaction between clearly defined long-run objectives and daily targets in reference dependence, primarily because of a nebulous definition of the "long run" in other contexts or a lack of granular data on the short run. This is the first field study of which I am aware to examine both behaviors in a unified way using real-world data. I provide evidence that workers are using daily reference dependence to overcome self-control problems and describe how the firm can affect this dynamic. My real effort task helps to confirm this relationship in a controlled environment with randomly assigned compensation conditions.

I first motivate my analysis with a discussion of the Koch and Nafziger (2020) model of narrow goal-setting in which present-biased agents intentionally induce loss aversion around period-specific targets as a way to combat present bias. I then discuss the key empirical questions related to this setting. These include examining whether workers behave in a way consistent with reference-dependent preferences with loss aversion, whether this behavior is driven by the firm's bonus scheme or other behavioral factors, and if the relationship is likely causal—or in other words if the firm's incentive structure can induce or "make" reference-dependent preferences. I then use my sales data and online experiment to examine these key questions.

In the sales data, I use a detailed panel of observations in half-hour increments with each seller's location, cumulative service contracts generated (which I call "sales" throughout the paper), pitches presented to a prospective customer, and the probability of stopping work for the day (the extensive labor supply margin). My measure of intensive margin effort (or alternatively phrased "exertion") is pitches per half hour. I define the reference point as a sample proxy of expectations: each seller's own average daily number of sales for all past workdays in the season, which I show is highly correlated with revealed long-run objectives. I find significant evidence of reference dependence with loss aversion in stopping behavior. Upon reaching their expectations-based reference point, the probability a worker stops for the day increases significantly by a factor of 2.8–4.1 times relative to below the reference point. This suggests that losses loom larger than gains by a factor of approximately three to four. On the "exertion" margin, the change is approximately 1.5–3.9 times, though the effect on pitches is quantitatively small. The choice of when to stop working is the key margin at which reference-dependent daily labor supply operates.

I then use a panel of each sales worker's daily performance outcomes and work hours to examine the relationship between a worker's sales, work hours, and the firm's lump-sum bonuses paid at the end of the season. The commitment device hypothesis of reference dependence suggests that the firm's contract structure incentivizes the worker to optimize around a long-run goal at a bonus threshold and workers then distribute that goal into daily goals.

I show three pieces of evidence in support of this hypothesis. First, I show that workers in this setting are forward-looking, as evidenced by the fact that their work hours do not significantly respond to changes in their realized commission rates. Rather, they set plans for their work hours based on their predicted commission rates. Second, I show that upon reaching their relevant bonus threshold, workers significantly reduce their work hours, pulling them out of their prior personal equilibrium. If habituation or status quo anchoring were the source of the daily reference dependence I document, we would expect to see no such change. In the standard rational agent model, we would not expect evidence of daily reference dependence, and a one-time bonus payment that does not reduce the commission of the next sale should not lead to a significant reduction in labor supply if agents are rationally forward looking.

Third, I show that the distributions of performance are subject to significant "bunching" around bonus thresholds, but where such bunching emerges early in the sales season. This is particularly visible when examining the distribution of performance among those targeting the same bonus threshold. The distribution of cumulative performance subsequently narrows from the top of the distribution at the end of the sales season because those above the bonus reduce their labor supply. This dynamic leads to a significant upward shift in the variance of daily work hours during the final two weeks of the sales season and as a worker nears her final tally. These observations provide strong evidence that the firm, through its compensation scheme, can shape the choice of long-run objectives, and, through goal setting in personal equilibrium, short-run expectations.

To show that the bonus schedule in the sales data is causally affecting the formation of daily reference points, I conduct an online real-effort task experiment. Performance is measured across four rounds, but performance is only payoff relevant at the end of the final round. The online task confirms the hypothesized relationship between "long-run" bonuses and the use of short-run references as a commitment device. Relative to those offered a piece rate, participants in my experiment that were offered a lump-sum bonus exhibit bunching in the distribution of their performance, both at the end of the experiment when the bonus is realized and during early rounds before the bonus is realized. The bunching is most pronounced at the average they would need to reach in each round to achieve the bonus by the end of the last round. Many participants set targets above this level as a hedge against final-round fatigue risk. In the bonus treatment with a higher bonus threshold, bunching in round-specific performance happened higher in the distribution, meaning that round-specific goals were manipulable via the final bonus threshold. In the final round, those that surpassed the bonus threshold drastically reduced their effort similar to the final weeks in the sales setting. Furthermore, through this targeting behavior, the bonus condition in the experiment significantly increased "profits" from the task by reducing labor costs relative to worker productivity.

The prior literature that touches on reference-dependent labor supply or task effort is divided into two types. The first addresses the core empirical question of whether short-run reference dependence exists. The second addresses how reference dependence around major targets is manifest in the distribution of final performance around a target.

In the first strain, several papers have found a negative relationship between daily wages and hours worked, downward shifts in labor supply at particular earnings levels, or negative labor supply responses to large tips among taxi and rideshare drivers.[1] The most recent of these, Thakral and Tô (2021), provide evidence not only of reference-dependent labor supply, but that reference points adjust over the course of the day. This is relevant for my setting because sellers may update their reference points upon gaining new information about their skills. However, the taxi cab literature has not empirically explored the purpose of having reference points at all—adaptive or fixed. Even though the earliest taxi cab studies (e.g. Camerer et al. (1997)) hypothesized that income targeting may help drivers address self-control problems, none of these studies has empirically explored this dynamic.[2] In a recent experiment in Kenya, Dupas et al. (2020) show that a person's stated income needs and expectations for earnings (rather than just total income) act as reference points. The authors suggest such targeting motivates workers to perform their physically demanding jobs. This directly relates to the "commitment device" function of reference points I test in my analysis, though intertemporal dynamics are missing from that analysis.

Previous studies testing reference dependence are quite narrowly focused on routine and manual tasks like taxi driving or physical labor. This paper provides new evidence in a novel work

[1] See Camerer et al. (1997); Chou (2002); Crawford and Meng (2011); Farber (2015); Morgul and Ozbay (2015); Agarwal et al. (2015); Martin (2017); He et al. (2018); Schmidt (2018). However, the literature is far from settled. A competing set of studies of drivers finds a positive relationship between daily wages and hours worked and concludes that the standard model performs better than prospect theory (Farber, 2005; 2008; 2015; Sheldon, 2016). Other analyses that find evidence supporting the standard model examine day laborers in Malawi (Goldberg, 2016), stadium vendors (Oettinger, 1999), fishermen in Florida and India (Stafford, 2015; Giné et al., 2016), and markets in India (Andersen et al., 2014).

[2] Other papers that find evidence of reference-dependent labor supply analyze the behavior of bike messengers (Fehr and Goette, 2007; Goette et al., 2004) and fishermen in Hawaii (Nguyen and Leung, 2013).

context that uses adaptive cognitive and social skills in a developed country. Understanding this distinction is crucial if workers in manual occupations differ significantly in their attributes (e.g. discount rates, cognitive capacity, risk preferences, etc.) from those who select into primarily social or cognitive occupations or who have the education to enter these occupations.[3] In addition, the "lumpy" nature of income in my sales context and the lottery-like nature of success at each door that decouples immediate income from effort makes this setting quasi-experimental and ideal for the study of loss aversion. Daily income in my sales context can be very high, so daily labor supply decisions are more financially consequential than in the past literature.[4]

The second strain of the literature focuses on the distributions of final outcomes around a single or ending target or reference. In a firm-worker setting, Kuhn and Yu (2021) examine the effects of kinks in a commission schedule on team performance and find these act as symbolic rewards or targets, leading to bunching in the distribution of performance. Cai et al. (2022) examine kinks in the compensation schedule at a Chinese manufacturing firm to estimate labor supply elasticities and firm cost savings. Their analysis does not consider behavioral factors such as loss aversion. A similar dynamic appears in Freeman et al. (2019) in relation to shifting the bonus threshold at a Chinese insurance company, which led to significant increases in worker output and firm profits. Beyond knowing that these incentive schemes increased total worker output, little is known about how or why these approaches were effective. The underlying day-to-day behavioral dynamics have immense implications. For example, if a new compensation scheme induced workers to set short-run goals with loss aversion, this represents, from the firm's perspective, a low-cost *psychological* incentive rather than a high-cost *monetary* incentive. Finding ways to influence worker expectations may be a more cost-effective method for motivation compared to an increase in wages.

The most relevant studies that deal with explicit long-term goals are recent papers examining the behavior of marathon runners. These demonstrate significant distortions around round numbers or a runner's stated time goal along with survey evidence of a discontinuous change in satisfaction upon reaching the reference (Allen et al., 2017; Markle et al., 2018). These studies provide compelling evidence that stated goals are important for final outcomes and runners exhibit loss aversion. Marathon running is an extreme athletic event that requires effort on a minute-by-minute basis during the final race and frequent training and planning on before it. Like many other arenas in which substantial, sustained effort is required (like the firm-worker studies above), marathon final times can be considered the "long run" outcome that is the result of a dynamic process. Despite this, previous studies have not focused on how the long-run target interacts with short-run choices

[3]See, for example, Cadena and Keys (2015); Bellemare and Shearer (2010); Patnaik et al. (2020); Fouarge et al. (2014); Warner and Pleeter (2001).

[4]As an example, encountering one extra resident willing to purchase pest control services leads to an increase in income of \$100–\$250. The average amount earned in an entire shift for a taxi driver is \$270, so an extra sale or two by a seller is worth the same amount but takes roughly the same amount of time as 1-2 taxi trips (16–32 minutes) (Thakral and Tô, 2021).

or preferences. The setting of this study along with granular data on short-run behavior allows me to relate short-run loss aversion to long-run goals in a new and unique way.[5]

My analysis of long-run objectives and short-run expectations is the first field study of which I am aware to examine the conditions around which workers select expectations and goals in personal equilibrium and subsequently exhibit loss aversion around those expectations in their daily efforts as a commitment device—the "why" of daily reference points. I show that the firm can play a role in that choice and that workers are responsive to firms' broad incentives even in their daily activities. Self-imposed, short-run reference points appear malleable when long-run incentives change. My online experiment confirms the causal interpretation of these dynamics by specifically linking an increase in reference-dependent behavior to non-linear compensation schemes.

Finally, I contribute to the literature on "insider econometrics" and the use of non-linear incentives and bonuses in the workplace and their effects (Ichniowski and Shaw, 2003). Recent work shows that bonus payments for reaching a performance threshold increase worker effort and productivity (Freeman et al., 2019; Graff-Zivin et al., 2019; Cai et al., 2022; Kuhn and Yu, 2021). I demonstrate through both the sales and the experimental settings that one way in which these incentives may enhance productivity is by shaping long-term goals and expectations, which are then translated into short-term, internalized performance benchmarks. In the online experiment, I document significant cost savings via the use of these non-linear incentives relative to piece rates when ex ante mean performance is the same. From the firm's perspective, it is costly to directly monitor or incentivize high effort and productivity among workers on a daily basis. Longer-run non-linear incentives appear to be a cost-effective means of motivating effort via narrow brackets.

The question of reference-dependent labor supply is central to our understanding of the power of incentives to induce effort. This is particularly applicable to managerial practice. Reference dependence makes it easier to motivate a worker if she perceives herself to be in a "loss" domain, but the opposite is true of the "gain" domain. If firms can use non-linear payments or other tools to shape worker expectations, then the firm can anchor the worker's reference-dependent labor supply at a higher level, leading to higher worker effort and significant cost savings.

# 2 Door-to-Door Sales Context

The door-to-door sales industry constitutes a sizable portion of the "direct sales" industry, which generates approximately $35 billion in revenue each year in the United States.[6] Workers within the direct sales industry are presented with high-powered incentives, including high commission rates that rise with performance and the use of bonuses. These are also common in a

[5]One study that does approach this dynamic is an analysis of professional golfers (Pope and Schweitzer, 2011) wherein the hole-specific target of par acts as a reference. However, because the larger goal of winning the tournament is difficult to predict and depends on opponents' performance, par may act as the only well-defined marker of success, especially when hole difficulty varies significantly.

[6]See statistics from the Direct Sales Association (link) (Accessed November 1, 2020).

variety of sales occupations.

A large number of firms that engage in door-to-door sales are located in the Mountain West region of the United States and employ thousands of college-age workers each summer to sell their products and services.[7] These include solar panels, pest control services, knives, and home security systems. General industry practice is relatively homogeneous across these products. Recruited sellers meet with managers, listen to an explanation of the work and earnings potential, and sign independent contractor agreements that stipulate the commission structure under which they will sell and their assigned city. The work itself is unpleasantly hot in the summer and often entails distasteful interactions with local residents. To entice skilled sellers to join their teams under these conditions, most companies will advertise that sellers make an average of $40,000 during the late April to late August sales season selling six days a week. There is a high level of competition between companies seeking to land top talent, and there is an extremely wide variance in sales skills among recruits, leading to a large variance in income. The company whose data I analyze, which I will call "PestCo," operates within these norms.

A "sale" at PestCo is recorded when a resident signs a contract for pest control services that lasts 12–18 months for services given quarterly. The contract is recorded electronically. Within pest control sellers at PestCo, the timing of sales can vary widely. On average, sellers generate one sale for every 20 pitches they present, but exactly which of those 20 pitches will result in a sale and at what time each sale will occur is highly uncertain. Any single pitch could result in a sale, so each knock on a house door is akin to entering a type of lottery. Hitting one's expected number of sales early in the shift comes as a meaningful surprise. Because the value of each sale to the seller is large, the stakes for each sales pitch are high.

PestCo, like nearly all door-to-door sales companies, pays large commissions in the range of 18–40% on the value of the service contracts they generate. A typical sale can result in an income to the seller between $100 and $250 depending on the value of the service contract signed by the customer and the seller's commission rate. Importantly, commission rates are increasing in cumulative sales performance and increase discretely in increments of 50 sales. The final commission percentage for each sale is calculated at the *end* of the sales season. The result is a set of discrete bonuses in 50-sale intervals (with a small marginal increase in per-sale commissions past the threshold). Sellers are paid an up-front portion of their commissions ($75 per sale) during two-week pay periods, similar to a regular paycheck. The balance of commission payments is calculated at the end of the season based on final performance and paid out thereafter.

[7]One core reason for locating in this region is the large supply of young college students (usually age 20-25) who have recently returned from 2-year or 18-month proselytizing missions for The Church of Jesus Christ of Latter-day Saints, which is headquartered in Salt Lake City, and whose members are the majority in the state of Utah. These proselytizing missions, in a purely practical sense, use skills very similar to a sales job: approaching strangers and striking up a conversation, connecting quickly, moving conversations toward a specific goal, and winsomely absorbing rejection. Recruiters understand this dynamic and seek to capitalize on this skill density.

Figure 1 characterizes the total income a seller earns at the end of the summer season depending on their total sales at an assumed contract value of \$500. A seller who produces 149 sales receives a commission of 25% on all sales at the end of the season, while a seller who generates 151 sales receives a 27% commission for all sales at the end of the summer. This results in a lump-sum bonus of approximately \$1,500 for crossing the 150-sale threshold (plus approximately \$10 more per sale above it, so the kink is trivial compared to the bonus). In addition to this de facto bonus from the commission change, the seller receives a flat "rent bonus" of about \$2,000 that covers the seller's apartment rental costs for the summer.[8] The average first-year seller yields between 100 and 150–175 sales, while experienced sellers generate 150 to 300 on average. The highest ability sellers generate over 350 sales for incomes in the \$60,000–\$80,000 range. If a worker expects to end in a particular 50-sale interval, the operating incentive is a linear piece rate with a bonus.

The skill requirements of the job make this setting unique in the literature. Sellers must be able to strike up a conversation with a stranger, understand and respond to objections, communicate the value of the product, and adapt their strategy on the fly as more information about the customer is revealed. Each of these tasks is cognitively demanding, and any interaction requires strong interpersonal communication skills. The prior literature has generally considered occupations in which income is a smooth function of hours worked, and deviations from average income are relatively small. For example, a standard deviation in wages for a taxi driver is only about 10% of the mean (Thakral and Tô, 2021). At PestCo, a standard deviation in the daily number of sales is 100% of the mean, and the effective daily wage can double in as little as 30–60 minutes. Income is also accrued in discrete units, creating more salient opportunities for earnings references than in the past literature.[9] The "lumpy" nature of income in this context, therefore, is an advantage over existing studies because each door interaction is quasi-experimental.

Another unique feature of this setting is that outside considerations that might influence the formation of medium- and long-term earnings targets in other settings are absent from this setting. Most sellers are below the age of 25 and have not formed financial commitments that require set payments that might influence the formation of salient "income needs" as examined in a prior study (Dupas et al., 2020). The largest determinant of a person's income needs, the cost of housing, is paid for by PestCo unless the seller fails to secure 150 sales. These needs are, therefore, baked into the performance schedule. Income needs over the short-run (e.g. the week) or over the medium-run (e.g. the month) are not fully operable because the vast majority of compensation for their work is received by workers at the *end* of the season.

Finally, there is an important information innovation to note regarding these sellers. Through

[8] At 250 sales, sellers qualify for the company vacation: an all-expenses-paid trip that includes airfare, hotel accommodations, food, and excursions.

[9] E.g. it is much easier to count contracts sold than total income earned net of tips while driving, even when the tips are "large" (approximately \$30) as in Schmidt (2018).

the company's centralized website and mobile application, sellers can view their entire performance history. All workers are aware of their normal performance, including their cumulative sales and average output each day. The availability of this information makes references related to one's own performance highly salient. Through its website and mobile app, PestCo tracks every sale and house "knock" recorded by each seller. This forms the basis of my analysis dataset. See Appendix C for more details on industry practice and contracts.

# 3 Data

My analysis datasets come from the comprehensive sales and seller tracking databases from PestCo for 2018–2019. The company uses a common sales tracking app that documents every door at which a seller records interacting with a resident and the location and timestamp of those interactions. Sellers also mark a house on their tracking software when the resident is not home or if the customer requested not to be contacted again. PestCo separately tracks the date and time each service contract is signed, the location of each customer, and the seller who generated the sale. Together, these two systems give a comprehensive view of the activities of each seller every day they are knocking on doors and selling in their work area.

Using the raw sales and knocking data, I construct two panels of individual seller performance. First, I build a daily panel of each seller's sales, work hours (defined as the time between the first knock/sale and the last knock/sale), cumulative sales over the season, and cumulative average daily sales as a measure of "recent expectations." Following the past literature (Crawford and Meng, 2011), I calculate a proxy for each seller's recent expectations by examining each seller's average past daily sales during the season. The selling week runs Monday through Saturday, and because residents are home at higher rates on Fridays and Saturdays and seller experiences differ by day of the week, I calculate each seller's average daily sales specific to each day of the week from all past days in the same sales season. These expectations can update and evolve over the course of a season, though the measure is remarkably stable after the first 2-3 weeks.[10]

In my second dataset, I construct a panel of each seller's pitches presented to a prospective customer, daily cumulative sales, and stopping probability each half hour of their shift. This interval of observation is the same as that in the recent taxi literature (Thakral and Tô, 2021). For each seller in each half hour, I create a measure of their current distance to their daily expectations: their number of cumulative sales so far that day minus their average sales for that day of the week. For values less than zero, a seller has not yet achieved her expectations and is therefore in a "loss" domain, while values greater than or equal to zero indicate a seller is in a "gain" domain. In this dataset, I define "starting" a shift as the half hour of the day in which a seller records her first knock of the day, and I define "stopping" as the half hour of the shift when the last knock of the day was recorded. In all, my half-hourly panel contains approximately 459,000 observations for 512 sellers

[10]Using various definitions of recent expectations such as sales in the prior five weeks yields similar results.

across 180 days in 2018-2019 covering the late-April to mid-August season.

I supplement these panels with daily weather data from the National Oceanic and Atmospheric Association (NOAA) National Climate Data Center (Menne et al., 2012). I include as controls daily total precipitation, high temperature, and low temperature for the weather station nearest to the ZIP code in which each seller is working. These factors may be important because door-to-door sales is an almost exclusively outdoor job. During these summer months, heavy rain and humid heat greatly increase the marginal cost of effort, and heat can have negative effects on cognitive ability and learning (Park et al., 2020). Alternatively, these factors might keep people inside their homes if the outdoor conditions are inhospitable, so the relationship between sales and these conditions is ambiguous ex ante.

One theoretical concern in this context is that sellers might be differentially sorted by managers into neighborhoods that are "easier" or "harder" to sell in. I include in my analysis controls for the characteristics of each person's work area. I use ZIP code data from the American Community Survey's 5-year summary files for 2013-2017 to serve as controls. I include variables that are likely to affect demand for pest control services or the ability to pay for them.[11] However, there is essentially no evidence of sorting behaviors correlated with seller performance. Additionally, managers emphasize that making assignments to work areas based on perceived skill or other seller attributes is costly to them as managers and generates unclear returns, which significantly undermines the business case for it.

Summary statistics for my two panels are in Table A1. Across all half-hour periods, the average number of sales is 0.16 based on 2.28 pitches. The average number of sales per day across all sellers is approximately two based on 6.9 hours per day, though there is substantial variation. Sellers work in relatively high-income areas. The median household income in their sales areas is \$86,000, and nearly 20% of residents in the average ZIP code have incomes between \$100,000 and \$150,000. Seller work areas are mostly single-family homes (mean of 80%), are predominantly Non-Hispanic white (mean of 80%), are relatively highly educated (mean of 45% Bachelor's degree or more), and have stable populations.

From the half-hourly panel, Figure A1 shows the distribution of start and stop characteristics for each working day. Panel A shows that most sellers start their shift with their first knocks and sales between 1:00 PM and 2:30 PM, though there is substantial variation in start times. Some start as early as 10:00 AM, while others begin working in the late afternoon or early evening. After

[11] These variables are median household income, rates of unemployment and poverty, the share of home values in specific ranges, total housing units, the share of units that are owner-occupied or are single-family homes, the share of households with a married couple, the share of adults with a Bachelor's degree or more, and the share of the population that has not moved in the past year. These jointly explain less than 3% of the variation in sales both between and within sellers. Appendix Table A2 provides details of a regression of daily sales on weather and ZIP code characteristics. Only three coefficients are statistically significant at the 10% level. Estimates excluding these controls are nearly identical but slightly less precise. See Appendix C for additional background.

starting their shift, the majority of sellers stop working between the sixth and eighth hours, though a large share stop working for the day before their sixth hour of work.

From an incentive standpoint, if there is a positive autocorrelation in sales each day—that is, if success now is predictive of success later in the day—then a worker having success right now faces lower marginal costs of effort in the coming hours. This will work against the downward shift in labor supply predicted by Prospect Theory. To test for this, I residualize sales each half hour by regressing sales each half hour on fixed effects for seller, day of the week, week of the season, and year as well as controls for actively knocking on doors, weather, and ZIP code characteristics. I then calculate the autocorrelation in these residuals between half-hour periods and present the results in Panel A of Figure A2. The results suggest that there is high autocorrelation in residualized sales for just under one hour, or that success now is predictive of success at least for the next half hour. If sellers understand this, they have an incentive to continue to work. Panel B of Figure A2 shows that average seller performance increases as the day progresses, particularly after 5:30 PM when residents return home from work (for what the company calls "peak knocking hours") and sellers begin following up with contacts from earlier in the day or week that request a callback. This is not due to a change in the composition of workers, but because workers have more opportunities to make meaningful contact with residents. The marginal cost of sales falls later in the day. In short, despite these two features "stacking the deck" against reference dependence in terms of incentives, I still detect meaningful evidence using my formal tests, which I describe in Section 6.1.

This context and the availability of comprehensive data provide a unique opportunity to test for the existence of reference-dependent labor supply, examine how short-run and long-run goals interact, and study how short-run goals and reference dependence may be shaped by the firm.

# 4 Conceptual Framework

## 4.1 Reference Dependence vs Standard Model for Labor Supply

One key insight of Prospect Theory is that losses loom larger than similarly sized gains (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992). The implication for labor supply is that workers supply greater effort while in a "loss domain" (before achieving a target) relative to what they supply in a "gain domain" (after achieving a target). This leads to a discontinuous change in marginal utility after surpassing some reference point, with the marginal utility of income falling significantly by some factor $1/\lambda$, where $\lambda$ is the parameter of loss aversion. This induces a discontinuous change in labor supply. Importantly, no such discontinuity is predicted by the standard model. Appendix D provides more general background on linear gain-loss utility in labor supply.

## 4.2 Goals and Daily References

As important as the parameter of loss aversion ($\lambda$) is to the model of reference dependence, equally important is the definition (or location) of the reference point itself. In an essential theoretical paper, Kőszegi and Rabin (2006) (the KR model) theorize that "recent expectations" act as important references. But how do people form short-term expectations? The KR model proposes that these expectations are determined in what they call "personal equilibrium," that is, by behaviors that are optimal given the worker's expectations about the future. Put another way, a worker can make a plan around what she perceives to be the optimal path forward, and when the final choice is made in real time, the planned path becomes her reference point. This path is her personal equilibrium.

This theoretical result has important implications. The first is that if wage increases are anticipated or predictable, a worker will respond by working more hours and can make a plan ahead of time, similar to the standard model. In the context of the bonuses paid in door-to-door sales, this means that sellers make their initial daily labor supply choices based on what they determine to be optimal given what they expect to be their most feasible bonus. If workers obtain new information about their abilities, they can quickly adjust their future goals to a new bonus and then adjust their daily reference points. This creates a feedback loop between future expectations and recent experience wherein a simple measure of average past performance integrates both pieces of information. The second key implication is that workers exhibit gain-loss utility over outcomes that deviate from expectations. After setting her plan for the path ahead, the seller responds each day to whether or not her performance is below or above what she expects for the day. Significantly, negative comparison utility and higher marginal utility while working below daily expectations induce more effort to reach expectations.

Building on the concept of personal equilibrium, multiple papers propose that goals and the expectations they generate are rational if the worker has a problem with self-control as a result of present-biased preferences (see Shefrin and Thaler (1992); Camerer et al. (1997), with Koch and Nafziger (2016; 2020) presenting the most formal recent treatment). Given that present bias has been documented in a variety of contexts such as exercise goals (DellaVigna and Malmendier, 2006), education (Ariely and Wertenbroch, 2002), credit markets (Meier and Sprenger, 2010), and savings (Ashraf et al., 2006), it is simple to extend the concept to labor markets.[12]

To combat present bias, as a commitment strategy, a worker will subdivide (or "bracket") her broad or long-run objectives into narrow evaluation periods or tasks with the *intention* of inducing loss aversion if she is below her reference point. Not achieving daily performance expectations then induces a sense of loss, and the worker will increase her effort toward those expectations to avoid it. This idea is mentioned as a possible explanation for observed daily income targeting in

[12]For a comprehensive discussion of present bias, see (Dellavigna, 2009).

the first empirical analyses of taxi driver behavior (e.g. Camerer et al. (1997)), though the analysis does not explore it in detail.[13] Dupas et al. (2020) similarly invoke this explanation, though the analysis considers stated income needs rather than goals per se. This is important because "income needs" can apply to any context in which budgetary targets like rent/mortgage payments may act as medium- or long-run targets. This means that exogenous manipulation of targets (e.g. by a firm) is not a necessary condition for reference-dependence with goal-setting.[14]

It is worth explicitly exploring this dynamic of goal-setting under the framework in Koch and Nafziger (2020). Like in my sales setting, suppose workers perform the same task each day (in time $t \in [1, T]$) with effort level $e_t$ that incurs costs $c(e)$ that are convex. Then suppose there is a total benefit $b$ at the end of a long-run evaluation period that is a function of total effort, and effort is deterministic over utility outcomes. If a worker is a quasi-hyperbolic discounter (Laibson, 1997), then there are $t$ versions of the worker, one for each day, with utility $U_t = u_t + \beta[\sum_{\tau=t+1}^{T+1} u_\tau]$ and instantaneous utility $u_t$ and a present-bias factor of $\beta$. Instantaneous utility is $u_t = -c(e_t)$, and final period utility $u_{T+1} = \sum_{t=1}^{T} b(e_t)$. Ex ante, a period zero self sets marginal costs and benefits equal such that $\beta = 1$ and $b'(e_0^*) = c'(e_0^*)$. This would be the equilibrium effort under the standard model of some chosen long-run outcome or "broad" bracket.

Now suppose each daily self after period 0 discounts future benefits by $\beta < 1$. Equilibrium effort with present-biased preferences would be $\beta b'(e_0^*) = c'(e_0^*)$. A worker that set out to perform at $e_0^*$ to achieve total benefit $b(e_0^*)$ in time 0 has an incentive in time $t$ to substitute effort from today to tomorrow. The prospect of substituting effort across days (because total outcomes are fungible across days) may lead to suboptimal effort in time $t$ under the ex ante assumption that the worker may increase effort in $t + 1$.

But suppose self 0 sets a narrow bracket through a daily goal to bind the incentives for self $t$ in the future through additional comparison utility penalties, i.e. for $e_t < g_t$, $\hat{\beta}(g_t - e_t)$. For a sophisticated individual that correctly predicts $\beta$ and calibrates $\hat{\beta}$, personal equilibrium suggests that $g_t$ should be the same as the optimal effort that period 0 self would choose given their beliefs about future effort, or in other words, that $\hat{e}_{t,0} = g_t$. When tasks are repeated daily, $g_t = [b(e_0^*)]/T$. Self $t$ then provides effort $g_t$ each day, thus solving the self-control problem. For partially naive individuals who underestimate their present bias ($\beta$), Koch and Nafziger (2020) suggest that daily goals are still achievable as long as target effort does not exceed maximum feasible daily effort as

[13]"Daily targets can also serve a second purpose: like many mental accounts, they help mitigate self-control problems." (Camerer et al., 1997), pp. 426. The commitment device mechanism is also a possible reason why people that correctly predict loss aversion nevertheless prefer loss-framed contracts (Imas et al., 2017).

[14]In the context of, for example, taxi drivers, setting a daily target and exhibiting loss aversion each day can be a method of ensuring that monthly payment obligations can be successfully managed, particularly if the work imposes disamenities (e.g. if it is boring, physically demanding, physically risky, etc.). In addition, round numbers may serve as reference points in several settings (e.g. in marathon running (Allen et al., 2017; Markle et al., 2018)).

expected ex ante, while broad bracketing through longer-run (e.g. weekly) goals is more likely to fail.[15]

In many occupations (like sales), effort costs may fluctuate through a day-specific cost function ($c_t(e_t)$), that is, the time and effort cost of achieving the same objective (e.g. sales). The standard model predicts a worker will provide more effort on "good days" where the marginal costs of effort are low and less on "bad days" where the marginal effort costs are high, i.e. when $c_t'(e_t)$ is high. That is, when the marginal benefits are consistent from day to day, higher marginal costs will lead to lower equilibrium effort. Effort, therefore, will fluctuate from day to day. When there is present bias, a worker has further incentive on "bad days" to implement effort substitution because of the expectation of future "good days" to make up for it.[16]

Combating this effort substitution is the key incentive introduced by narrow bracketing. In the case of daily goal-setting intentionally inducing gain-loss utility, because the marginal utility of income is higher in the loss domain, workers have the incentive to exert more effort on "high-cost" days to achieve a minimum performance. On "low-cost" days, they surpass their target more easily, but the marginal utility of additional income falls, so the worker has the incentive to reduce their labor supply upon surpassing it.

For the worker, there are two downsides to narrow bracketing relative to broad bracketing. The first is that narrow bracketing induces negative comparisons and lower experienced utility while in the loss domain. The second is that she may suboptimally reduce her effort on low-cost days. Given these features, if a worker who (ex ante) set out to achieve $b(e_0^*)$ has achieved it using narrow bracketing, the worker then has no incentive to continue inducing negative comparisons each period because the larger target has been (or will imminently be) reached. The worker then may reduce her period-specific effort in the final period(s).[17]

What does this framework imply for my sales setting? Here, a descriptive example is helpful. When PestCo sets a bonus at 200 sales, the bonus directly affects a forward-looking worker who knows her ability on the job could reasonably yield her at least 190 sales. The bonus can generally lead to three separate responses in the above framework. First, in the case of both the standard model and the KR model, she may raise her objective for total sales at the end of the season from 190 to 200 (a new $b(e_0^*)$) because she believes it is attainable and the \$2,000 bonus is worth the extra effort (an incentive effect).

Second, if she is a present-biased worker that does not engage in narrow bracketing of her target, then she may engage in effort substitution. This will lead, over the course of the sales season, to lower total effort each period than required to achieve the 200 sales she set out to

[15]Koch and Nafziger (2020) show theoretically that individuals with daily goals provide daily/total effort greater than they would without gain-loss utility but still less than their period 0 self would prefer if $\beta$ is sufficiently low.

[16]This is an implication of Proposition 2 in Koch and Nafziger (2016). Under various assumptions about the probability of failure, narrow bracketing is strictly optimal.

[17]This represents the combination of hypotheses H3a and H3b in (Koch and Nafziger, 2020).

accomplish before the season started. In this case, there should not be a structural break in labor supply around expectations because there is no mental bracket. She will tend to work more on "low cost" (high selling) days than "high cost" (low selling) days to try to make up for this shortfall, but ex post realized performance is likely to fall below 200 if her present bias is severe.

Third, if she engages in narrow-bracketing in personal equilibrium, knowing she needs 200 sales over 100 days, she can easily form an expectation for each day's performance: just over two sales per day (a narrow bracket). She then works each day with these two sales per day in mind, which satisfies the personal equilibrium condition. If she views each day in a separate mental account, being below her two sales creates a sense of loss, so she will work harder or extra hours to get the remaining sales. If she does achieve her two sales, she can then quit for the day and feel satisfied with her performance. Achieving her two sales then keeps her on track to hit her goal of 200 by the end of the season. This keeps the variance in her performance each day lower than without narrow bracketing despite fluctuations in effort costs each day.

An important prerequisite for a personal equilibrium amenable to goal-setting is that a worker must have a realistic, forward-looking view of what she can plausibly accomplish, i.e. not full naivete about $\beta$, and effort costs. In my sales setting, a sign that sellers are forward-looking would be that their daily labor supply does not substantially change as their cumulative performance (and therefore realized commission rate) increases because they have already optimized for their long-run expectations. While this assumption is reasonable, it is not certain; evidence from other contexts indicates that myopia affects the optimality of decision-making in areas like pension planning (Mitchell, 1988), health behaviors (Cawley and Ruhm, 2011), and take-up of financial aid (Bettinger et al., 2012). I explain my test for forward-looking expectations in the next section.

The prior empirical literature on reference dependence has been unable to examine personal equilibrium as a long-run goal because the work settings analyzed to date do not provide a clear endpoint at which a worker evaluates any goals she may have. The "long run" is too nebulous. On the contrary, my sales setting provides a clear end date. A second reason the prior field literature has been unable to examine long-run goals is that the occupations under study provide no external menu of goal choices and are measured in settings in which other factors such as income needs may form the most salient (or only) form of medium- to long-run targets, which remain unobserved to the researcher.[18] PestCo, through its use of lump-sum bonuses, provides external incentives for workers to set their sights upon specific long-run outcomes. These bonuses increase the salience of particular points to act as targets. Meanwhile, the sales setting is a fixed-term job that is conducted far away from "home" and is paid mostly at the end of the season among young workers without

[18] Bénabou and Tirole (2004) propose that self-reputation is what animates the use of daily targets. However, they do not consider the use of longer-run financial targets or needs in their models. In a variety of economic interactions, reputational considerations are intertwined with economic incentives such as promotions, bonuses, or the option value of future job prospects.

major fixed-schedule financial obligations whose housing costs are baked into the bonus schedule. This limits the scope for outside income needs to dictate specific points in the earnings distribution as targets. Thus, this empirical setting provides a unique opportunity to study these questions empirically.

## 4.3 Key Empirical Questions and Predictions

Based on the above theory, there are three empirical questions to consider in this setting that I address: first, do workers in this sales setting behave in a way consistent with reference-dependent preferences with loss aversion? Second, is this related to long-run goals around the bonus scheme or the result of another behavioral process such as habituation or status quo bias? Third, is this relationship causal? or in other words, can an incentive structure induce the formation of or "make" reference-dependent preferences through long-run targets?

The first key empirical question in this framework is whether or not workers actually have reference-dependent preferences in their daily labor supply choices. This question has been addressed in prior studies, primarily from the perspective of manual and routine occupations such as taxi drivers, delivery workers, fishers, and day laborers. This has not been explored in more cognitively adaptive and social occupations. My sales setting is unique because knocking on each door has a strong element of chance, like entering a small lottery, and the income arrives in countable, discrete units (sales). Evidence of this behavior would establish a baseline behavior inconsistent with the standard labor supply model. It also narrows the conversation to then focus on the more novel question of how this behavior relates to the structure of compensation.

How would we know these sellers exhibit this type of reference dependence? Three features would provide strong evidence of this phenomenon. First, within each day, there would be a distinct break in labor supply choices (e.g. stopping time or exertion) upon reaching the day's sales expectations in the form of a kink and/or discontinuity. This is in contrast to having no such break in the standard model. Second, all else constant, the relationship between performance and work hours should be more pronounced for days that fall at or below expectations compared to days that were above expectations. This is because workers would put in more hours on below-expectations days; if they find themselves above expectations, they lower their efforts, which weakens the relationship between the two. This is in contrast to the standard model, which predicts the opposite because low effort cost days (unexpectedly successful) make up for high effort cost days (unexpectedly unsuccessful) and performance across days is fungible. Third, the distribution of cumulative performance around expectations across workers should be narrow and subject to heaping. This is because the distribution is compressed by the change in marginal utility across workers, with upward pressure on the distribution coming from the bottom.

The second key empirical question is whether or not the long-run target generated by the bonus scheme is the reason behind the narrow bracketing and loss aversion in daily performance. How would we know that daily performance targets in this setting are related to the bonus schedule rather than being the product of another behavioral process like habituation? If loss aversion around expectations is the result of habituation, anchoring to the status quo, or other explanations like planning heuristics that are independent of the bonus schedule, then labor supply within each worker's personal equilibrium should persist throughout the sales season regardless of one's distance to the final threshold. This is because their "recent expectations" have not changed. Importantly, this means that the variance of daily effort across sellers should remain constant, even as the end of the contract or long-run target nears or is achieved.

However, if workers have established their daily targets *as a commitment device* to achieving the long-run objective, then those that have surpassed their long-run target will reduce total effort and change their labor supply significantly, taking themselves out of their prior personal equilibrium. This is because their expectations about the future have changed—in particular, the returns to additional work when the bonus has been achieved. Mean effort for workers just past the long-run target should fall substantially. This results in a wider variance in period-specific effort across workers at the end of the contract period, while the variance in cumulative performance should narrow during the final period(s).[19]

The third key empirical question is distinct but related to the second: can the incentive structure presented to a worker *causally* induce her to set short-run goals and exhibit reference dependence in her daily or period-specific labor supply choices? This is distinct from the second empirical question because my sales setting does not have exogenous variation in the types of structures presented to sellers. While the menu of bonuses is exogenous to worker preferences and abilities, there is no condition in which a worker does not face a bonus at all. To leverage exogenous variation in incentive structures and test whether compensation structure causally influences the making of short-run reference-dependent preferences, I created an online real-effort task experiment, which I describe below. Put briefly, in comparison to a simple piece race, an incentive structure such as a bonus makes salient particular points in the performance distribution, which then causes workers to set period-specific targets in a multi-period task. In my experiment, participants in the bonus conditions would have period-specific performance distributions that are, relative to a piece rate structure, more in line with the first empirical question above. These distributions would be narrower and bunched around the target. However, in the last period, those that surpassed the target should lower their effort, the variance of performance in the final period should widen, and the distribution of total cumulative performance should narrow.

To summarize, here are the empirical predictions I consider in my analysis:

(I) **If there is reference-dependence in daily targets:**

[19] In the standard model, to a first approximation, rational agents have already optimized their labor supply around the present value of their predicted compensation (including the bonus), so the timing of a one-time bonus should not lead to significant changes in labor supply (Kahneman and Thaler, 1991).

There will be a kink and/or discontinuity in labor supply (e.g. stopping behavior or exerted effort) upon surpassing expectations. No such shift is predicted by the standard model.

(II) As a result of (I), the relationship between performance and effort (e.g. work hours) should be stronger for days that fall at or below expectations compared to days that were above expectations. The opposite is true in the standard model.

(III) As a result of (I) and (II), the distribution of performance around expectations across workers should be narrow and subject to bunching.

(IV) **If the bonus scheme shapes the behavior rather than other behavioral mechanisms:**
Those that have surpassed their long-run target will deviate from their prior personal equilibrium and reduce their effort. The variance of effort in the final period(s) should increase, while the variance of cumulative performance should decrease. Under other determinants of personal equilibrium, no such deviation is expected.

(V) **If compensation structures causally shape narrow bracketing and personal equilibrium:**
In comparison to a simple piece race, a bonus for total performance causes workers to set period-specific targets in a multi-period task. The distributions of performance for each period should then match Prediction (III): narrower than the piece rate and bunched around the period-specific target. In the last period, those that surpassed the target should lower their effort, the variance of final period performance should widen, and the distribution of total performance should narrow, as in Prediction (IV).

Below, I describe my experimental design and its relation to the above predictions.

# 5 Experimental Design

To test the mechanisms underlying my analysis of sales data, I conduct an online real-effort task experiment on the Prolific platform. The task is a simple button-pushing task in which participants are asked to alternate pushing the "a" and "b" buttons on their computer keyboard, following closely the procedure in DellaVigna and Pope (2017). A successful sequence of "a" and "b" results in 1 point. Participants were asked to perform the task for a total of ten minutes in four rounds lasting two minutes and thirty seconds for each round with a break of ten seconds between rounds.

Each participant was paid a flat $3 payment to participate. I then randomly presented participants with one of three incentive conditions:

1. A bonus of $1 for achieving 2,000 points (2,000 bonus condition)
2. A piece rate of $0.05 per 100 points (the piece rate condition)
3. A bonus of $1 for achieving 2,400 points (2,400 bonus condition)

Importantly, the payment rates were calibrated based on the distributions of performance in DellaVigna and Pope (2017) to have equal predicted mean performance over ten minutes, meaning the expected payoff for a performance of 2,000 points is exactly equal in the bonus condition and the piece rate condition. The core difference between the two is that the bonus makes the 2,000-point target salient and payoff-relevant for the end of the 10-minute task period. This also proxies the bonus environment in my sales setting.

Under the Kőszegi and Rabin (2006) model, a priori anticipated or known changes in compensation and expectations will lead to greater effort provision, predicting a difference in performance between the bonuses set at 2,000 and 2,400 points. The Koch and Nafziger (2020) model predicts that if participants are induced to set a target by the bonus offer, optimality suggests they will set narrow brackets for themselves with their target being some point at or above the average number of points they would need to achieve to reach their target by the end of the task. They would exhibit reference-dependent preferences around this target. This is designed as a commitment device in the face of present bias in what is a tedious task.

This conforms to the discussion above with Prediction (V) under reference dependence causally induced by non-linear incentives. First, those in the piece rate condition will behave more consistently with the standard model and will exhibit greater variance in performance than those in the bonus condition with the same expected payout at 2,000 points. Second, loss aversion in the bonus condition will lead to a bunched distribution of performance in every period until the bonus is surpassed. Third, once the bonus is surpassed, the purpose of the narrow bracketing is no longer relevant, so these workers will reduce their effort and performance below their former targets. The variance of performance in the final period in the bonus condition will increase, with greater density at the bottom of the distribution. Fourth, the distribution of total performance will narrow in the final period. Finally, consistent with the KR model's treatment of anticipated changes in expectations and compensation, the bonus condition at 2,400 points will lead to more average effort than either the piece rate or the bonus at 2,000 points by setting expectations and targets at a higher level, but the pattern of effort over periods will follow the 2,000-point bonus condition.

I also included in the experiment questions after the task about strategies they may have used. I also asked whether they enjoyed the task and whether they felt stress during the task. These questions allow respondents to state their internal thought processes about their observed performance each round as well as a proxy of experienced utility and disutility.

# 6 Empirical Strategy

## 6.1 Tests of Reference-Dependent Labor Supply

For my empirical analysis, I first use my half-hourly panel to test for the presence of reference dependence in daily labor supply choices consistent with Prediction (I). As outcomes, I focus on stopping work for the day, a measure common to the past literature, as well as pitches presented in the next half hour, a measure of effort "exertion" conditional on continuing to work.[20]

[20]In Appendix A, I also examine the probability of recording any knocks during the next half hour. This measures if workers are more likely to take a break as a result of their position relative to expectations. These results closely

In the experimental ideal, the amount of income earned as of any particular hour of the day or the daily wage would be randomly assigned, after which each seller would make her labor supply choices. My empirical approach approximates this experiment by separating out conditions correlated with effort costs and the number of sales a seller has generated to that point. The underlying assumption is that conditional on my various fixed effects and controls, the exact number of sales a seller has at a particular point in the day is as good as random. Given the context in Section 2 and the full set of controls and fixed effects I present, this assumption is reasonable. The sales setting presents a unique opportunity to study this behavior because whether or not a sale occurs depends strongly on who answers the door when a seller knocks—similar to a small lottery.

I first estimate a non-parametric model of labor supply with respect to each seller's distance from their sales target to trace out patterns without imposing a functional form. Following the past literature (Crawford and Meng, 2011), I define expectations and targets in all my models as the average daily sales from all past workdays in the season specific to each day of the week (i.e. a specific mean for Mondays, Tuesdays, etc). I note here that PestCo runs various competitive tournaments during the sales season of three different types. Because these significantly change the incentives faced by the sellers and may shift the worker's target, I separately analyze behavior during non-tournament days and present those results in my tables and figures.[21]

For seller $i$ during half hour of the shift $t$ and half hour of the day $h$ on day of the week $d$ in week of the season $w$ in year $a$, I estimate the following model:

$$\begin{aligned} y_{ithdwa} =& \beta_0 + \sum_{e=-k, e\neq 0}^{k} \beta_e * \boldsymbol{I}_e\{sales_{ithdwa} - \overline{Sales}_{idwa} = e\} \\ &+ \alpha X_z + \sigma W_{dwa} + \mu_{it} + \eta_h + \nu_d + \omega_w + \tau_a + \varepsilon_{ithdwa} \end{aligned} \tag{1}$$

Here, $y$ is the probability of stopping work for the day and the number of pitches presented to a resident in the next half hour. The expression $\{sales_{ithdwa} - \overline{Sales}_{idwa} = e\}$ represents the seller's current distance to expectations as of a particular half hour of the day; in other words, her current cumulative sales that day ($sales_{ithdwa}$) minus the worker's average daily sales specific to the day of the week ($\overline{Sales}_{idwa}$), which is a proxy for recent expectations. $\boldsymbol{I_e}$ is a dummy variable assigned to each distance value. The coefficients of interest, $\beta_e$, capture non-parametric effects of being $e$ distance from one's expectations target. Distance values below zero are characterized as being "losses" and values above zero are "gains." Because sales are discrete values, these coefficients

mirror the knocks-based exertion margin. See Appendix Figure A4.

[21] I also estimate my parametric models using a pooled sample across all tournament and non-tournament periods and interact my coefficients of interest with indicators for tournament periods. These estimates are in Appendix Table A4. I report the non-tournament coefficients in my figures (see Section 7.1.1). See Appendix C for more on these tournaments.

include values rounded to the nearest integer, with the (0,1.5) interval being included in $\beta_1$.[22] Under reference dependence, beginning with $\beta_1$ there will be an upward change in stopping probability or a downward change in effort as the distance from expectations increases.

The various fixed effects ($\mu_{it}$, $\eta_h$, $\nu_d$, $\omega_w$, $\tau_a$) are for seller by half hour of the shift, half hour of the day, day of the week, week of the season, and year, respectively. Importantly, $\mu_{it}$ captures a seller-specific baseline hazard over the shift. That this factor is omitted by the prior literature is noted by Thakral and Tô (2021). They include a driver-specific hazard in their estimates of taxi driver behavior and conclude this is vital for unbiased estimates of labor supply responses to daily earnings. I incorporate this methodological improvement into my estimates. The $X$ vector is the set of ZIP code characteristics from the ACS, and $W$ is the set of weather controls each day from NOAA (discussed in Section 3).

In my main models of interest, I fit parametric estimates that impose a functional form to match the non-parametric estimates in Equation 1. This equation is a regression kink and discontinuity design with linear splines on each side of the target.

$$\begin{aligned} y_{ithdwa} =& \beta_0 + \beta_1\{sales_{ithdwa} - \overline{Sales}_{idwa}\} \\ & + \beta_2\{sales_{ithdwa} - \overline{Sales}_{idwa}\} * \boldsymbol{I}_{sales \geq \overline{Sales}} \\ & + \beta_3 * \boldsymbol{I}_{sales \geq \overline{Sales}} \\ & + \alpha X_z + \sigma W_{dwa} + \mu_{it} + \eta_h + \nu_d + \omega_w + \tau_a + \varepsilon_{ithdwa} \end{aligned} \tag{2}$$

This approach allows the slope of the relationship between labor supply and distance to one's reference to differ in the gain and loss domains. $\boldsymbol{I}_{sales \geq \overline{Sales}}$ is a dummy for if current sales are above average sales, or in other words, for reaching expectations and entering the gain domain. The coefficient $\beta_1$ defines the slope of the relationship between one's current distance to average sales and labor supply in the loss domain. The $\beta_2$ coefficient captures the change in slope upon crossing the reference and entering the gain domain. Finally, $\beta_3$ captures any discontinuous level shift in stopping probability or effort from reaching the reference. The fixed effects and controls are all the same as in Equation 1. In a standard framework, there should be no sudden change in the slope and no discrete level shift upon reaching the reference. Under reference dependence with loss aversion, we should expect to see an upward change in the slope of stopping probability. In other words, $\beta_2$ will be significantly positive in the stopping model. The coefficient $\beta_3$, while not predicted by simple loss aversion, represents a discrete penalty for "losing," or for falling short of

[22] Other studies examining reference dependence discretize earnings into ranges. The "correct" size of the earnings range has been the topic of some disagreement (Farber, 2015; Martin, 2017; Thakral and Tô, 2021). In sales, earnings are already discrete, so I do not have to impose any binning structure. Because the common support in the distance to expectations gets very thin outside the [-4,4] interval, I plot that interval in my figures. I report the full set of distance dummy coefficients corresponding to my figures in Appendix Table A3.

expectations, which suggests reference dependence.[23] If $\beta_2$ and/or $\beta_3$ are significant and positive in the stopping model, this represents strong evidence of reference-dependent labor supply.

As a secondary battery of tests, I perform two regression exercises to look for other indications of reference dependence. The first is a regression of each seller's total sales at the end of the season on their average sales in the first two weeks of the season as well as the first five weeks of the season. A high R-squared indicates that initial daily sales outcomes and labor supply choices have high predictive power for total cumulative sales. Though not conclusive, this indicates a high degree of persistence in performance.

The second and more important regression is a test of Prediction (II). I use a panel of worker-day observations to estimate a model of hours worked each day on the number of sales that day interacted with an indicator for if the day's total sales were higher or lower than expectations (average daily sales specific to each day of the week. I estimate:

$$\begin{aligned} y_{idwa} =& \beta_0 + \beta_1 Sales_{idwa} * \boldsymbol{I}^{+}_{Expectations} + \beta_2 Sales_{idwa} * \boldsymbol{I}^{-}_{Expectations} \\ &+ \alpha X_{zdwa} + \sigma W_{zdwa} + \mu_i + \nu_d + \omega_w + \tau_a + \varepsilon_{idwa} \end{aligned} \tag{3}$$

I include fixed effects for seller ($\mu_i$), day of the week ($\nu_d$), week of the season ($\omega_w$), and year ($\tau_a$). The outcome is hours worked that day, while $Sales$ is the total number of service contracts the seller sold that day. The indicators $\boldsymbol{I}^{+}_{Expectations}$ and $\boldsymbol{I}^{-}_{Expectations}$ are dummy variables for if the total sales that day were above expectations or below. In the standard model, because workers will increase their hours when the cost of effort is low, $\beta_1$ will be more strongly positive than $\beta_2$. In other words, the relationship between work hours and sales will be stronger when total sales for the day are above average (Dellavigna, 2009). The opposite is true under reference dependence.[24]

The combination of all these pieces of information supports the hypothesis that these door-to-door sales workers exhibit reference-dependent preferences with loss aversion in personal equilibrium as they make their labor supply choices each day.

## 6.2 Tests of Goal-Setting and the Bonus Schedule

I next use my daily panel of sales and labor supply to examine goal setting by sellers. A prerequisite for setting long-run goals and personal equilibrium is that sellers are forward-looking. To test for this, I estimate how sellers adjust their labor supply as their cumulative sales increase throughout the season. Sellers only know their final earnings per sale at the *end* of the sales season after their total number of sales and total revenue are calculated. If sellers have realistic, forward-looking expectations for what they can achieve, perceived changes in their wages that come with entering a new 50-sale performance interval should not change their daily labor supply because

[23]Estimating with second-order polynomials results in small and statistically insignificant coefficients on the squared term for both outcomes.

[24]That contact and sales rates increase at the end of the work day additionally "stacks the deck" against reference dependence in this model because sales late in the day are less costly to achieve.

they have already optimized over their chosen long-run outcome. According to the KR model, those with higher daily expectations for their commission rates should work more hours than those with lower ex ante daily expectations for their commission rates. Conversely, myopic agents would respond to an increase in their realized commission rate, and this myopia would be inconsistent with the KR model's personal equilibrium condition.

According to Prediction (IV), sellers that have previously worked with reference-dependent utility each day specifically in the pursuit of their long-run targets should shift out of their personal equilibrium once they have surpassed or will imminently surpass their long-run targets. This would result in a significant reduction in labor supply for those at or near their target.

To test both of these dynamics, I use my daily panel to regress hours worked per day on indicators for 10-sale intervals of current cumulative sales interacted with indicators for 50-sale bins of total sales at the end of the summer. I estimate the following equation for seller $i$ on day of the week $d$ in week of the season $w$ in year $a$ working ZIP code $z$:

$$\begin{aligned} y_{idwa} =& \beta_0 + \sum_{e=[0,10)}^{[320,330)} \sum_{f=[100,125)}^{[300,325)} \beta_{ef} \boldsymbol{I}_e * \boldsymbol{I}_f \\ &+ Efficiency_{idwa} + \alpha X_{zdwa} + \sigma W_{zdwa} + \mu_i + \nu_d + \omega_w + \tau_a + \varepsilon_{idwa} \end{aligned} \tag{4}$$

The outcome variable $y$ is the number of hours worked per day. The indicators $\boldsymbol{I_e}$ and $\boldsymbol{I_f}$ are indicator variables for currently working in interval $e$ and for having total season sales in interval $f$. In this specification, $\beta_{ef}$ captures the non-parametric effects of being in interval $e$ for an individual whose total sales for the season were in interval $f$. These coefficients trace the labor supply path of those who ended with a similar total number of sales. The $X$ vector controls for the characteristics of the ZIP code. The $W$ vector controls for weather in the worker's ZIP code. The $Efficiency$ variable is a time-varying measure of each seller's average sales per hour for all past workdays that season, which proxies for sales ability and may evolve as the season progresses. Changes in this measure capture learning effects over the season, which shifts the expected marginal earnings of an additional period of work. I include fixed effects for seller ($\mu_i$), day of the week ($\nu_d$), week of the season ($\omega_w$), and year ($\tau_a$). These fixed effects ensure that the $\beta$ coefficients characterize within-seller choices holding constant other characteristics of the sales season, fatigue, or learning effects. If the $\beta_e$ coefficients are constant within different types of sellers $f$ as they cross intermediate 50-sale intervals, then it does not appear that sellers are responsive to a change in their realized wage. Similarly, if reference dependence evolves due to habituation or status quo bias, then these coefficients should be consistent regardless of one's distance to a bonus threshold.

The KR model predicts that the coefficients on all intervals in $e$ should be consistently larger as their expected total sales—therefore, expected commissions—increases in $f$. Importantly, if sellers are engaged in imposing loss aversion upon themselves simply to reach a bonus threshold, reaching the bonus should lead to a shift in labor supply out of their prior personal equilibrium. Specifically, the coefficients for $\beta_e$ will be much smaller after crossing the worker's final bonus threshold. This would result in a significant drop in hours worked. For example, a worker that finished with 150-175 sales (just beyond the 150-sale bonus threshold) would work fewer hours in the intervals just at or after the bonus threshold. Equation 4 captures this dynamic for each bonus threshold from 100 to 300 sales.

Next, to visualize how workers respond to bonus incentives with the predictions discussed above, I present kernel density estimates of cumulative sales at the end of the season and as well as throughout the season for all workers. I present densities at two-week intervals to draw the evolution of sales over time. I also perform the same analysis for subgroups in particular total performance bins from the end of the season to trace how the densities within groups progress (relevant to predictions (II) and (III) ). As the focal example, I present these for those whose total sales at the end of the season were between 175 sales and 225 sales, putting them around the bonus threshold of 200 sales. If workers with the same apparent goal at the end of the season have a narrow and/or bunched distribution of performance, this further suggests that workers are anchoring to their goals and exhibiting more effort while below their daily expectations, which compresses the distribution upwards. Subsequently, if this behavior is done in the pursuit of surpassing their long-run targets, we would expect the distribution of cumulative performance to narrow as workers pass or will shortly pass their bonus.

Similarly, the variance of worker efforts each day should be relatively constant throughout the sales season until the bonus threshold nears or has been passed. I examine this within-day effort variance by looking at the standard deviation of worker daily hours across two dimensions: over time and over distance to each worker's final sales tally. To remove any composition effects that might drive this variance on any particular day, I regress each seller's daily work hours on a set of seller and day-of-the-week fixed effects. This removes volatility attributable to day-of-the-week effects and worker composition effects, i.e., who shows up to work that day. I present the standard deviation of the distribution of the subsequent residuals by elapsed time in each seller's season and by distance to the seller's final total for the summer.[25] In both instances, the variance of the sellers' daily effort should be consistent until the very end of the contract, either in the final days or in the final sales. A sharp rise in the variance of labor supply signals a departure from personal equilibrium and a break from prior expectations consistent with the bonus being the impetus behind the initial daily loss aversion.

## 6.3 "Making" Reference-Dependent Preferences

Given the nature of the work and the fact that all sellers are aware of and accept the employment agreement knowing about the bonus schedule ex ante, one could imagine that some unobserved

[25] When examining distance to one's total tally, I also include a fixed effect for week of the season to distinguish progress toward the goal from time effects.

selection effect might lead these sellers to endogenously adopt references on a knife-edge case in which their references coincidentally fit the bonus schedule. While this is unlikely, because no sellers operate in a work environment free from the bonus schedule, we may not be able to fully attribute a causal effect of the compensation structure on workers' reference-dependent preferences.

To test Prediction (V), show that the relationship I document is causal, and demonstrate that reference-dependent preferences can be induced or "made," I examine the data from my online real-effort task experiment. I analyze the distribution of performance in each round across treatments of my experimental setting with kernel densities. In comparison to the piece rate condition, bonus conditions will have lower variance even in early rounds (prediction (III)). If participants experience loss aversion as a commitment device around the round-specific goals they invoke, there will be bunching at or above the average performance needed to achieve the goal at the end of the final round, i.e. heaping on the right side of this threshold. In addition, enjoyment of the task and feeling stress during the task will relate more strongly to performance at particular points in the distribution of performance even in early rounds. If this behavior is done in service of achieving the bonus, then worker effort in the final round in the bonus condition should fall once the worker is confident he/she has surpassed the threshold, the variance of effort that round should rise, and the variance in total cumulative performance should fall. Importantly, the only difference in the distributions between the piece rate and the bonus conditions is the random assignment of the compensation structure, so differences in behavior are causally attributable to this assignment.

# 7 Results

## 7.1 Do Workers Exhibit Reference-Dependent Labor Supply?

I now present the results from my half-hourly panel estimates of labor supply responses to daily references. Figure 2 shows each of the coefficients from the non-parametric estimates from Equation 1 as well as the linear estimates from Equation 2. Panel A indicates that as sellers approach their target from the loss region, the probability that they stop working for the day is relatively flat at a slope of 0.0021. After surpassing their target number of expected sales, there is a clear upward kink in the probability of stopping work. The slope of the relationship between cumulative sales and stopping probability in the gain region for expectations-based targets is 0.0058, or 2.8 times that in the loss region.

Panel B of Figure 2 shows the same estimates for pitches during the next half hour, which is a measure of effort conditional on continuing to work. In contrast to the results for stopping probability, there is a relatively smooth relationship between exertion and sales each day. There is a minimal change in this measure at the reference. The size of the decline is small in percentage terms: each sale reduces effort conditional on continuing to work by approximately 1% from a baseline mean of 2.38 pitches at the reference. These results suggest that reference dependence in exertion is negligible, but that there is a steady decline in effort as sales increase. Reference dependence is most apparent at the extensive margin. In other words, if sellers stay on the job after reaching their expectations, their exerted effort is similar.[26]

Next, I use my estimates to calculate the parameter of loss aversion, $\lambda$. My setting requires an approach to measuring loss aversion that is not dependent on the measurement scale of the output units (sales). One such approach is advocated by Köbberling and Wakker (2005). Their measure focuses on the difference in the slopes of the utility function in the gain domain and the loss domain. Because my empirical model partials out all covariates correlated with effort costs and because the timing of sales is conditionally random, the only difference between the gain and loss domains is the difference in the marginal benefit, so the ratio of slopes for each outcome measures $U'(0)_{\uparrow}/U'(0)_{\downarrow}$. The ratio of slopes in the stopping model is $0.0058/0.0021$, or 2.8, and the slopes at the "exertion" margin have a ratio of 1.5. Using a bootstrap with 250 replications for inference on the ratio of these slopes in the stopping model yields 95% confidence intervals for my estimates of loss aversion of 1.2 to 4.4 (1.5 to 5.3 for the bias-corrected interval). My estimate of 2.8 in my baseline models is the most conservative of my stopping model estimates, and my non-parametric estimates imply an even larger ratio. Other specifications, which I detail in Section 7.1.1 yield estimates as high as 4.1 or 5 for the stopping model and 3.9 at the margin of effort conditional on continuing. In their review, Gächter et al. (2007) find loss aversion coefficients of approximately 1.4 to 4.8 across various measurements, with an average value of 2.6. A coefficient of loss aversion in my results of 2.8 is, therefore, quite consistent with the prior literature.

One notable feature of the KR model is that personal equilibrium is established immediately following the formation of expectations, even when the final payoff is far away. If reference dependence with loss aversion around long-run goals is operating in daily labor supply choices, then there should be evidence of this phenomenon early in the sales season when performance relative to the bonus threshold is not immediately payoff relevant. To investigate this, I separately estimate my models during the May, June, and July months. In May, opportunities for effort substitution are more plentiful, and these opportunities fall during June and July. Appendix Figure A6 presents these estimates. The evidence is consistent with my overall results, even in May and June. The upward shift in stopping behavior after surpassing the expectations-based reference may be most pronounced early on, consistent with goal-setting behavior as a commitment device.

The results of my two regression exercises using my daily panel are in Table 3. In Panel A, the R-squared for the regression of total sales at the end of the season on average daily sales in weeks 1–2 is 0.752, meaning that average daily performance in the first two weeks explains over three-quarters of the variation in total cumulative sales at the end of the season. Expanding this

[26] In two of my later specifications, there is a downward kink and a discontinuity in effort. In Appendix A, I also examine the probability of actively knocking, meaning recording any knocks during the next half hour. These results in Appendix Figure A4 closely mirror the exertion margin. At a mean active knocking share of 80%, the slope estimates are quantitatively small and not economically meaningful. Sellers do not appear more likely to take breaks during their work as a function of their position relative to expectations.

period to the first five weeks, the R-squared is 0.872, explaining almost 90% of the variation in total sales. There is little unexplained variation in total season sales after conditioning on the first two to five weeks, and there is high congruence between sales outcomes in the first two weeks and behaviors the rest of the season. In Panel B, the relationship between daily sales and hours worked is stronger on days that fell *below* expectations compared to days that exceeded expectations. This runs counter to the predictions of the standard model that workers will work more hours on days that have high wage returns (prediction (II)).

### 7.1.1 Robustness and Alternative Specifications

Rather than separately estimating stopping behaviors for non-tournament periods, my first alternative specification pools together all tournament/non-tournament periods and interacts each of my key measures of distance to daily expectations with indicators for what kind of tournament/non-tournament is operating. This allows the effect of crossing the reference to differ based on period type. The results of this specification are in Figure A7. The result for non-tournament periods is a more pronounced upward kink in stopping probability and the emergence of a downward kink in pitches per half hour. The slope in the gain domain is 4.1 times that in the loss domain for stopping probability, meaning loss aversion in this model is higher than in my baseline model. At the exertion margin, the slope in the gain domain is 3.9 times that in the loss domain. This specification confirms the results of my baseline model and provides even stronger evidence for reference dependence than in my baseline model.

My estimates impose a linear structure with a cutoff at each seller's cumulative average sales. This choice is in line with the KR model of reference dependence around recent expectations. As a robustness check against incorrect specifications of the cutoff at zero, I estimate my models again using non-linear least squares. To incorporate my fixed effects and controls into my specification, I first residualize the probability of stopping with my full battery of fixed effects and controls and use the residuals in my non-linear least squares estimate. Rather than imposing slope and intercept coefficients at zero, I allow the cutoff itself to be a parameter of the model. The results are in Table 1. The non-linear least squares estimates confirm that there is, indeed, a structural break at the worker's average cumulative sales and a strong upward tilt in stopping probability. The exact cutoff in the non-linear least squares estimate is 0.11, approximately one-tenth of a sale from my measure of the seller's expectations, which may be consistent with the "buffer" idea against later fatigue, which is explicitly invoked by respondents in my online experiment.[27] The ratio of slopes in the stopping model is 5.1, meaning that my baseline estimates may be quite conservative. For pitches per half hour, even though the estimates show a statistically significant kink downward and

[27] This is consistent with Kőszegi and Rabin (2007), who build on their concept of personal equilibrium in their theory of "preferred personal equilibrium." Here, the anticipation of risk leads to a strong tendency toward planning and the purchase of insurance. This performance buffer may, therefore, be a form of small-scale insurance.

that the ratio of the slopes across the reference is 3.3, the magnitude is small in percentage terms; each sale past the reference leads to a 1.8% decline in effort conditional on continuing to work, and the results are more sensitive to specification.

As an additional test, I estimate my baseline model but include exertion effort on the right-hand side: cumulative pitches that day as a measure of total exerted effort. If a worker is exerting a high level of effort on the job and becomes fatigued, the fatigue could be affecting her willingness to continue working or to exert effort in the next half hour. Table 2 presents these estimates for my parametric models. The results are nearly identical to my baseline model. The results for stopping behavior imply that my baseline model adequately controls for effort differences at the intensive margin that may have generated differences in sales. At the exertion margin, the negative slope in the loss domain is not as steep as my baseline model. Upon entering the gain domain, there is essentially no change in the slope from the loss domain, indicating that the decline in pitches across the reference is smooth.

Finally, I create an alternative measure of each worker's daily reference and also estimate my models with the full tournament/non-tournament interaction. I construct what I call a "goal-based" reference by examining the first 2 weeks of the worker's performance. I project their average daily sales from this period to the end of the season and then round to the nearest bonus threshold. If workers are projected to be within 15 sales of a bonus, I round up to the bonus, but if they are less than 35 sales over a bonus, I round down. I base this on the pattern of bunching from the kernel density estimates in the next section (7.2). I then allocate the average daily sales the worker would need to achieve this nearest bonus. These "goal-based" references are highly correlated with my proxy for recent expectations (0.82), consistent with a worker's rational expectations matching her likely goals. In Appendix Figure A8, I show that the use of this reference point is consistent with my baseline results.

## 7.2 Is the Bonus Schedule a Source of Reference Dependence?

Here, I present evidence for Predictions (IV) and (III). I begin by showing evidence that sellers are forward-looking and that they significantly shift their labor supply out of personal equilibrium upon nearing or passing their relevant bonus threshold.

Estimates from Equation 4 are summarized in Figure 3, which shows the predicted hours worked per day over 10-sales increments of cumulative sales from this model. Each line shows the labor supply trajectory of different bins of total sales at the end of the season. Sellers do not appear to be myopic. After what appears to be an initial adjustment period, those whose sales totaled over 300 quickly began working approximately 8 hours per day, while those with fewer than 200 total sales worked approximately 7 hours per day consistently over the season. This difference across expected commission rates is consistent with the KR model's prediction that *expected* increases in wages would increase labor supply. Notably, within tiers of total sales, there is very little variation in the predicted hours worked each day over current cumulative sales, and labor supply does not significantly shift upon receiving a commission raise by crossing into a new 50-sale interval. These results show that workers do not change their work hours regardless of how much of a commission increase they have secured, suggesting a singular focus on long-run performance expectations.

These patterns shift significantly once the relevant bonus threshold has been reached. Even when conditioning on length of time into the season, weather patterns, efficiency gains, and group composition, after passing the bonus, sellers drastically reduce their work hours by 1.5 to 2.5 hours per day, or by 20-30%.

I next present descriptive evidence on the distribution of sales throughout the sales season, which is relevant to predictions (IV) and (III). Figure 4 shows the results of kernel density estimates for total sales at the end of the season. Around each 50-sale bonus threshold, there is significant bunching, particularly at 150 and 250 sales when the bonuses include rent payments and the company vacation. This indicates that the bonuses are salient for the sellers. Figure 5 breaks down the density of total cumulative sales for each seller in two-week increments over the season.[28] Unevenness in the estimated density graphs is apparent beginning in week four and becomes clearer in weeks 8–10, which is just over the halfway point in the season. Notably, bunching groups that form early persist further up the sales distribution over time.

An even starker pattern emerges when examining groups of workers with a similar total performance at the end of the season. Figure 6 presents the kernel density estimates of cumulative sales over the same two-week intervals as Figure 5, but I limit this to those whose total sales at the end of the season were between 175 sales and 225 sales, or those around the bonus at 200 sales. In week 4, the distribution is tightly centered, after which bunching emerges in the distribution. This persists until approximately week 12, at which time the growth rate of the top of the distribution starts to slow, which compresses the distribution from the top. These figures confirm that sellers are particularly cognizant of and responsive to these lump-sum bonuses.[29] Visual evidence of upward pressure from the left tail of the distribution, especially in the early to middle stages of the season, is consistent with predictions for workers with reference dependence. Individual workers perform in narrow ranges around their expectations every day rather than experiencing peaks and troughs by substituting effort and sales performance across days.

I show more formally in Figure 7 that the standard deviation of residualized seller labor supply remains relatively flat over the course of time and distance to the seller's total until the very end of the contract. In Panel A, while the variance of hours decreases a small amount during the first month as new sellers learn about their capability in the field, there is very little change from day

[28] If a seller left relatively early in the season, their sales are included in the total as of the date they left and hold the same value as the weeks progress, so the relatively high density below 100 includes those who only worked a portion of the season.

[29] A similar pattern is visible for those who finish the season around the 150-sale bonus threshold, as seen in Appendix Figure A5.

30 until approximately day 85. Panel B shows that the variance of seller labor each day does not systematically vary until the goal or final tally is within approximately 50 sales.

These aggregate patterns may mask composition effects; for example, those that achieved 100 sales in total are not represented in Panel B when the distance to the final tally is more than 100 sales away. To investigate this further, Panels C and D show the same phenomenon as Panels A and B but are separated by groups of total sales (0-100, 100-200, 200-300). Panel C shows that the rise in the variance of seller labor supply is most concentrated among sellers with over 200 sales during the season, though there is a rise in the variance among those below 100 approximately halfway through the sales season. These workers are more experienced, on average, and so are likely to have a better sense of when or how they may reduce their efforts around their bonus threshold. Alternatively, they may have a better understanding of when to accelerate their efforts in the final weeks to ensure that they achieve the bonus. Both dynamics increase the variance of effort. Panel D shows that all these groups exhibit an increase in the variance of their labor supply in the final 25-50 sales.

Taken together, the results of each of these exercises in Sections 7.1 and 7.2 show that these sellers are 1) able to predict their own performance very early in the sales season; 2) aware of and responsive to the bonus schedule; 3) setting goals around bonus thresholds in the schedule; 4) distributing their long-run goals into daily expectations; and 5) shifting out of their personal equilibrium upon reaching or surpassing the bonuses. That all five of these conditions hold empirically is consistent with the conditions of the KR model's personal equilibrium definitions for reference dependence and the use of narrow goals in pursuit of long-run objectives (Shefrin and Thaler, 1992; Camerer et al., 1997; Koch and Nafziger, 2016; 2020).

## 7.3 Can Reference-Dependent Preferences be "Made?"

The previous analysis provides evidence of this goal-setting and loss aversion dynamic. The remaining question from Prediction (V) is whether or not this dynamic holds causally, or in other words, did the firm's compensation structure *cause* new reference dependence? This relates to a larger empirical question: can reference-dependent preferences be "made?"

I now present the results from my experiment from each round in Figure 8. Here, I primarily focus on the comparison between the bonus payment at 2,000 points and the piece rate condition, which has shown comparable mean performance in prior studies (DellaVigna and Pope, 2017). Figure A9 shows remarkably similar dynamics when comparing the piece rate to the bonus at 2,400 points. Panel A of Figure 8 shows the distribution of performance in the first round comparing the piece rate treatment and the bonus condition at 2,000 points. Panel B shows the same for the subsequent rounds. Three notable patterns emerge: first, despite having the same payoff at 2,000 points, the distribution of performance is consistently narrower in the bonus condition than in the piece rate condition, not only in the distribution of end performance but for all of the first three rounds (prediction (III)). Second, the distributions in the bonus condition exhibit heaping to the right of the piece rate condition in every round, but most especially by the end of the 10-minute period as evidenced by narrower spikes (prediction (IV)). This heaping is especially clear in Panel C, which presents the differences in the densities of round 1 performance between the two conditions.

Third, the left tail in the density of round-specific performance increases in mass substantially in the final round and does so only in the bonus condition. On the other hand, there is more kurtosis or peakedness in the distribution at 500 and 700 points during this round, consistent with some participants increasing their effort. The result is a steep increase in the variance of effort during the final round in the bonus condition, whereas the variance does not exhibit the same behavior in the piece rate scheme. These patterns are clear in Figure A10. The variance in the piece rate condition is much higher during the first three rounds, and the changes in the variance across rounds are nearly perfectly parallel until the final round. During the final round, the standard deviation of performance rises by nearly 35% in the bonus condition.[30]

The panels of Figures 8 and A9 demonstrate three additional results. First, there is (more) heaping in the bonus conditions than in the piece rate condition.[31] Second, there appear to be two heaping points at 500 points and 600 points and this heaping is more pronounced in the bonus conditions. In Figure A9, the heaping points are higher when the bonus is set at 2,400 compared to 2,000 points (closer to 700), meaning that round-specific sub-goals are responsive to the location of the end goal. When respondents were asked an open-ended question about their strategy, many in the 2,000-point treatment responded they targeted 600 points for each round as a buffer against fatigue or surprises in later rounds in order to avoid missing the 2,000-point threshold. Similarly, those in the 2,400-point bonus treatment stated 700 was their round-specific target.[32]

Third, the heaping in the distribution in the 2,000-point treatment shifts effort from just below 500 (600) points to just above 500 (600-640) points relative to the piece rate condition. In the 2,400-point treatment, this heaping is much more pronounced above 600 points This is important because round 1 is at the very beginning of the task when performance is not directly payoff relevant in the bonus condition.[33] These differences across conditions highlight that making a single point salient with a bonus establishes expectations not only for total performance but also

[30]Similarly, in the sales data, sellers with greater than 100 total sales saw the standard deviation of daily work hours increase by approximately 15-30% during the final quarter of the sales season (see Panel C of Figure 7).

[31]In the piece rate condition, participants may be anchoring to round numbers, which phenomenon has been observed in other contexts (e.g. in marathon running Allen et al. (2017); Markle et al. (2018)). That there is more heaping in the bonus conditions is important because the piece rate represents a counterfactual that takes into account any tendency to bunch at round numbers.

[32]This, again, is consistent with Kőszegi and Rabin (2007) in their discussion of Preferred Personal Equilibrium.

[33]Payoffs above 2,000 points are also *higher* in the piece rate condition, so more bunching at 600 (above what would be necessary to achieve 2,000 total points if that performance continued) in early rounds in the bonus condition is notable.

for each *round*.

Finally, Panel D of Figure 8 shows the total performance across the three experimental conditions. These densities affirm the predictions discussed above. In particular, the variance of total performance is lower in the bonus conditions than in the piece rate condition. The distribution of performance is also higher with the bonus at 2,400 than the bonus at 2,000, consistent with the KR model. While there appear to be two heaping points in the 2,000 bonus condition (at 2,000 and 2,400), there is one significant heaping point at the 2,400 bonus treatment. Despite cumulative performance of 2,000 points not being relevant at all to payoffs in the 2,400-point bonus condition, there is still substantial distributional heaping at 2,000 total points. In both cases, there is significant heaping at bonus condition thresholds and even at lower points in the distribution. Non-linear payments create or make salient personal targets, even if those targets are not immediately relevant for final payoffs.

These results have significant implications for firm costs. In the experiment, the average bonus payouts for the piece rate condition were \$1.15 per worker compared to \$0.80 for the bonus condition at 2,000 points, representing a statistically significant reduction in per-person costs of 31%. Meanwhile, the average total output for the 2,000-point bonus condition was only 1.58% lower and not statistically different from the piece rate condition. In the 2,400 point bonus, total payouts were only \$0.62 per worker, while output was slightly higher than the piece rate. One of the core reasons for these differences may be attributable to the bonus condition leading to loss aversion in earlier rounds.

One strength of this online experiment is the ability to directly elicit measures of enjoyment and disutility (proxied by feeling stress during the task). This would help separate out actual reference dependence from planning heuristics or other non-utility-based models of behavior. The patterns of responses across the distribution of performance are informative, particularly because reference dependence with loss aversion implies lower total utility just below the reference point. Figure A11 reveals that Round 1 performance relative to benchmarks of 500 and 600 points much more strongly predicts enjoyment of the task and stress under the bonus condition (at 2,000) relative to the piece rate condition (Panels A and B). This relationship holds even when nonparametrically controlling for total performance at the end of the task (Panels C and D). Conditional on cumulative performance, Round 1 performance is not payoff relevant in the bonus condition, yet there is still a substantial gap in the distributions of enjoyment and stress reported at the *end* of the task at 500 and 600 points. This lack of enjoyment and increase in stress below the reference point is consistent with loss aversion for these targets despite their lack of payoff relevance.[34] A planning heuristic or other non-utility-based models of this behavior would not generate this pattern of enjoyment or stress around these cutoffs when conditioning on performance.

[34]Imperfect ability to predict final performance at the end of round 1 may be a reason for the enjoyment gap persisting despite narrowly missing 500.

Finally, after the end of the task period, I asked each participant an open-ended question: "Did you have any particular strategy when performing the task across these rounds?" Most participants shared information about their hand placement or other physical movements. However, 35 respondents explicitly stated unprompted that they had an internal target of 500 per round, 29 of which were presented with the bonus conditions. An additional 14 bonus participants pinpointed 600 as their target compared to only two in the piece rate, meaning that bonus condition participants were nearly five times more likely to articulate this type of targeted goal-setting as their primary, salient focus across rounds.[35]

# 8 Discussion and Conclusion

Using novel, comprehensive data from a door-to-door sales company and an online experiment, this paper provides evidence of reference-dependent preferences in daily labor supply in a new setting. Door-to-door sales workers exhibit behaviors around expectations-based references in personal equilibrium consistent with loss aversion in their labor supply choices. I provide new evidence of this behavior when examining the choice of when to quit working for the day as well as the choice of how much effort to exert conditional on continuing to work. I find that the extensive margin choice (when to stop working) is the margin at which reference dependence is most operative. Exertion conditional on continuing to work is not strongly responsive to surpassing expectations in some specifications.

My analysis also provides new information about *why* workers behave this way. Reference dependence with loss aversion puts upward pressure on labor supply in the loss domain. This behavior coincides strongly with the schedule of bonuses offered by the firm, implying that the firm shapes the choice of expectations for these workers. Loss aversion around daily goals acts as a commitment device, keeping these workers engaged in work when they might otherwise shirk today assuming they can catch up tomorrow. However, upon nearing and surpassing their long-run target—the bonus threshold—these sales workers shift out of their personal equilibrium, consistent with the idea that the bonus was the catalyst behind the original daily targets.

The online experiment further supports this dynamic by showing that by simply having a bonus payment that makes a particular long-run target salient, workers respond by imposing upon themselves short-run targets that serve as references. They behave in a way consistent with reference dependence even when exact performance around expectations is not payoff relevant early on.

Taken all together, the results from my sales data and the online experiment suggest that setting narrow goals and exhibiting loss aversion around them serves an important function as a method of individual accountability (Koch and Nafziger, 2016; Hsiaw, 2018; Koch and Nafziger, 2020). Goals instill a sense of loss for not meeting a narrow target. Narrow bracketing of goals there-

[35] Some examples of responses include: "I ... was trying to get over 600 in the first rounds since I knew my fingers would be tired by the last round." "I made it a goal to get to 500 on each." "[I went] as fast as I possibly could ... to go over the 500 mark per round for the first three rounds."

fore frequently induces effort by keeping workers in a loss domain at the start of each day. This is particularly important if workers are present-biased. Given the unpleasantness of door-to-door sales, self-control problems resulting from present bias may be nearly universal in the occupation. These results also reveal a key mechanism behind the effects of non-linear incentives in the workplace: the establishment of expectations, both for the short and long run. The online experiment confirms that compensation schemes used by firms can "make" or reinforce reference-dependent preferences.

This paper contributes to the literature on reference dependence by providing evidence of this behavior in a new context distinct from the literature on the behavior of taxi drivers (e.g. Camerer et al. (1997); Crawford and Meng (2011); Thakral and Tô (2021)). I innovate in this area by considering why workers might exhibit reference dependence and the influence of the employer on these references. I also contribute to the literature on goal orientation around internal and external long-run targets demonstrated by the recent literature on marathon runners (e.g. Allen et al. (2017); Markle et al. (2018), and firm workers (e.g. Freeman et al. (2019); Kuhn and Yu (2021); Cai et al. (2022), respectively. The fixed endpoints for both the sales and the online experiment allow me to link granular data for the short run to clear evaluation criteria for the long run, which has been missing in the prior literature. My online experiment provides significant supporting evidence for a causal interpretation of the sales data.

I demonstrate that one area in which non-linear compensation schemes and other interventions affect workers is through the setting of expectations in the first place. I show that firms, through lump-sum bonus payments, influence the formation of long-run expectations for a worker in line with the concept of "personal equilibrium" detailed in Kőszegi and Rabin (2006). Subdivided goal setting of this nature suggests that reference dependence may be a rational response to self-control problems (Koch and Nafziger, 2016; 2020).

These results are of general interest across many contexts. Sales as an industry is a large market globally, and these types of incentives—non-linear bonuses and piece rates—are common features of a wide variety of sales occupations. The behaviors of door-to-door sellers, therefore, can easily be generalized to other sales and marketing occupations. Other industries and labor markets use these types of incentives. Piece rates are common in many occupations in which outcomes can be finely measured, from fruit picking (Graff-Zivin et al., 2019) to investment commissions for financial managers. The use of formal and informal bonuses at performance targets is ubiquitous, from the highest-paid CEOs to children selling coupon books to raise money for their sports or performing arts programs. That these incentives are widely used across occupations and contexts indicates that a diverse set of actors acknowledge the power of these incentives for motivating people.

There are a few caveats and limitations to this study. I cannot view the particular terms in each individual independent contractor agreement. Any deviations from the normal sales contract would introduce noise into my analysis. Finally, the degree to which the experience of these workers is generalizable depends on how one views the labor market experiences of relatively well-educated, college-age adults. That the incentives in this market are broadly used provides some evidence of the generalizability of my results.

My results have important implications for how workers optimize their labor supply and how firms and policymakers affect worker labor choices. Because workers are more motivated in the loss domain and less motivated in the gain domain by additional income, the effectiveness of a wage increase depends on the worker's reference and how expected the change may be. My results suggest that the firm, rather than trying to motivate *around* a reference, can influence the *positioning* of the reference itself. These results have significant implications for firm efficiency as the online experiment shows that the bonus conditions produced significant cost savings relative to worker output. Inducing narrow goals among workers is, from the firm's perspective, a low-cost *psychological* incentive rather than a high-cost *monetary* incentive. This may partially explain the use of these non-linear incentives across many workplaces and industries.

This analysis poses new questions for future research. Workers in my setting are recruited knowing the compensation structure beforehand though not necessarily their exact skills at performing the task. Future research may explore how dynamic changes in a reference (through changing bonuses or pay schemes during the contract period) influence the formation or changing of personal equilibria for workers and the dynamics of that change. Future work may also examine how information and the reliability of information influence the formation of personal equilibrium or the role firms can play in setting references in the short run in addition to the long run.

# References


**Agarwal, Sumit, Mi Diao, Jessica Pan, and Tien Foo Sing.** 2015. "Are Singaporean Cabdrivers Target Earners?" *Available at SSRN 2338476*.

**Allen, Eric J, Patricia M Dechow, Devin G Pope, and George Wu.** 2017. "Reference-Dependent Preferences: Evidence from Marathon Runners." *Management Science*, 63(6): 1657–1672.

**Andersen, Steffen, Alec Brandon, Uri Gneezy, and John A List.** 2014. "Toward an Understanding of Reference-Dependent Labor Supply: Theory and Evidence from a Field Experiment."Technical report, NBER.

**Ariely, Dan, and Klaus Wertenbroch.** 2002. "Procrastination, Deadlines, and Performance: Self-Control by Pre-commitment." *Psychological Science*, 13(3): 219–224.

**Ashraf, Nava, Dean Karlan, and Wesley Yin.** 2006. "Tying Odysseus to the Mast: Evidence from a Commitment Savings Product in the Philippines." *The Quarterly Journal of Economics*, 121(2): 635–672.

**Bellemare, Charles, and Bruce Shearer.** 2010. "Sorting, Incentives and Risk Preferences: Evidence from a Field Experiment." *Economics Letters*, 108(3): 345–348.

**Bénabou, Roland, and Jean Tirole.** 2004. "Willpower and Personal Rules." *Journal of Political Economy*, 112(4): 848–886.

**Bettinger, Eric P, Bridget Terry Long, Philip Oreopoulos, and Lisa Sanbonmatsu.** 2012. "The Role of Application Assistance and Information in College Decisions: Results from the H&R Block FAFSA Experiment." *The Quarterly Journal of Economics*, 127(3): 1205–1242.

**Cadena, Brian C, and Benjamin J Keys.** 2015. "Human Capital and the Lifetime Costs of Impatience." *American Economic Journal: Economic Policy*, 7(3): 126–53.

**Cai, Xiqian, Wei Jiang, Hong Song, and Huihua Xie.** 2022. "Pay for Performance Schemes and Manufacturing

Worker Productivity: Evidence from a Kinked Design in China." *Journal of Development Economics*, 156, p. 102840.

**Camerer, Colin, Linda Babcock, George Loewenstein, and Richard Thaler.** 1997. "Labor Supply of New York City Cabdrivers: One Day at a Time." *The Quarterly Journal of Economics*, 112(2): 407–441.

**Cawley, John, and Christopher J Ruhm.** 2011. "The Economics of Risky Health Behaviors." In *Handbook of Health Economics*. 2: Elsevier, 95–199.

**Chou, Yuan K.** 2002. "Testing Alternative Models of Labour Supply: Evidence from Taxi Drivers in Singapore." *The Singapore Economic Review*, 47(01): 17–47.

**Correia, Sergio.** 2016. "Linear Models with High-Dimensional Fixed Effects: An Efficient and Feasible Estimator."Technical report, Working Paper.

**Crawford, Vincent P., and Juanjuan Meng.** 2011. "New York City Cab Drivers' Labor Supply Revisited: Reference-Dependent Preferences with Rational-Expectations Targets for Hours and Income." *American Economic Review*, 101(5): 1912–1932, DOI: http://dx.doi.org/10.1257/aer.101.5.1912.

**Dellavigna, Stefano.** 2009. "Psychology and Economics: Evidence from the Field." *Journal of Economic Literature*, 47(2): 315–372, DOI: http://dx.doi.org/10.1257/jel.47.2.315.

**DellaVigna, Stefano, and Ulrike Malmendier.** 2006. "Paying Not to Go to the Gym." *American Economic Review*, 96(3): 694–719.

**DellaVigna, Stefano, and Devin Pope.** 2017. "What Motivates Effort? Evidence and Expert Forecasts." *The Review of Economic Studies*, 85(2): 1029–1069.

**Dupas, Pascaline, Jonathan Robinson, and Santiago Saavedra.** 2020. "The Daily Grind: Cash Needs and Labor Supply." *Journal of Economic Behavior & Organization*, 177 399–414.

**Farber, Henry S.** 2005. "Is Tomorrow Another Day? The Labor Supply of New York City Cabdrivers." *Journal of Political Economy*, 113(1): 46–82.

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

**Farber, Henry S.** 2015. "Why You Can't Find a Taxi in the Rain and Other Labor Supply Lessons from Cab Drivers." *The Quarterly Journal of Economics*, 130(4): 1975–2026, DOI: http://dx.doi.org/10.1093/qje/qjv026.

**Fehr, Ernst, and Lorenz Goette.** 2007. "Do Workers Work More if Wages Are High? Evidence from a Randomized Field Experiment." *American Economic Review*, 97(1): 298–317.

**Fouarge, Didier, Ben Kriechel, and Thomas Dohmen.** 2014. "Occupational Sorting of School Graduates: The Role of Economic Preferences." *Journal of Economic Behavior & Organization*, 106 335–351.

**Freeman, Richard B, Wei Huang, and Teng Li.** 2019. "Non-linear Incentives and Worker Productivity and Earnings: Evidence from a Quasi-experiment." *NBER Working Paper*(25507): , URL: http://www.nber.org/papers/w25507.

**Gächter, Simon, Eric J Johnson, and Andreas Herrmann.** 2007. "Individual-Level Loss Aversion in Riskless and Risky Choices."

**Giné, Xavier, Monica Martinez-Bravo, and Marian Vidal-Fernández.** 2016. "Are Labor Supply Decisions Consistent with Neoclassical Preferences? Evidence from Indian Boat Owners." URL: https://elibrary.worldbank.org/doi/abs/10.1596/1813-9450-7820, DOI: http://dx.doi.org/10.1596/1813-9450-7820.

**Goette, Lorenz, David Huffman, and Ernst Fehr.** 2004. "Loss Aversion and Labor Supply." *Journal of the European Economic Association*, 2(2-3): 216–228.

**Goldberg, Jessica.** 2016. "Kwacha Gonna Do? Experimental Evidence About Labor Supply in Rural Malawi." *American Economic Journal: Applied Economics*, 8(1): 129–49.

**Graff-Zivin, Joshua S, Lisa B Kahn, and Matthew J Neidell.** 2019. "Incentivizing Learning-By-Doing: The Role of Compensation Schemes."Technical report, NBER.

**He, Shu, Liangfei Qiu, and Xusen Cheng.** 2018. "Wage Elasticity of Labor Supply in Real-Time Ridesharing Markets: An Empirical Analysis." *University of Connecticut School of Business Research Paper*(18-21): .

**Hsiaw, Alice.** 2018. "Goal Bracketing and Self-Control." *Games and Economic Behavior*, 111, DOI: http://dx.doi.org/10.1016/j.geb.2018.06.005.

**Ichniowski, Casey, and Kathryn Shaw.** 2003. "Beyond Incentive Pay: Insiders' Estimates of the Value of Complementary Human Resource Management Practices." *Journal of Economic Perspectives*, 17(1): 155–180.

**Imas, Alex, Sally Sadoff, and Anya Samek.** 2017. "Do People Anticipate Loss Aversion?" *Management Science*, 63(5): 1271–1284.

**Kahneman, Daniel, and Richard Thaler.** 1991. "Economic Analysis and the Psychology of Utility: Applications to Compensation Policy." *The American Economic Review*, 81(2): 341–346.

**Kahneman, Daniel, and Amos Tversky.** 1979. "Prospect Theory: An Analysis of Decision Under Risk." *Economet-*

*rica*, 47(2): 263–291.

**Köbberling, Veronika, and Peter P Wakker.** 2005. "An Index of Loss Aversion." *Journal of Economic Theory*, 122(1): 119–131.

**Koch, Alexander K, and Julia Nafziger.** 2016. "Goals and Bracketing Under Mental Accounting." *Journal of Economic Theory*, 162 305–351.

**Koch, Alexander K, and Julia Nafziger.** 2020. "Motivational Goal Bracketing: An Experiment." *Journal of Economic Theory*, 185 104–149.

**Kőszegi, Botond, and Matthew Rabin.** 2006. "A Model of Reference-Dependent Preferences." *The Quarterly Journal of Economics*, 121(4): 1133–1165.

**Kőszegi, Botond, and Matthew Rabin.** 2007. "Reference-Dependent Risk Attitudes." *American Economic Review*, 97(4): 1047–1073.

**Kuhn, Peter J, and Lizi Yu.** 2021. "Kinks as Goals: Accelerating Commissions and the Performance of Sales Teams."

**Laibson, David.** 1997. "Golden Eggs and Hyperbolic Discounting." *The Quarterly Journal of Economics*, 112(2): 443–478.

**Markle, Alex, George Wu, Rebecca White, and Aaron Sackett.** 2018. "Goals As Reference Points in Marathon Running: A Novel Test of Reference Dependence." *Journal of Risk and Uncertainty*, 56(1): 19–50.

**Martin, Vincent.** 2017. "When to Quit: Narrow Bracketing and Reference Dependence in Taxi Drivers." *Journal of Economic Behavior and Organization*, 144 166–187, URL: http://dx.doi.org/10.1016/j.jebo.2017.09.024, DOI: http://dx.doi.org/10.1016/j.jebo.2017.09.024.

**Meier, Stephan, and Charles Sprenger.** 2010. "Present-Biased Preferences and Credit Card Borrowing." *American Economic Journal: Applied Economics*, 2(1): 193–210.

**Menne, Matthew J, Imke Durre, Bryant Korzeniewski, Shelley McNeal, Kristy Thomas, Xungang Yin, Steven Anthony, Ron Ray, Russell S Vose, Byron E Gleason et al.** 2012. "Global Historical Climatology Network-Daily (GHCN-Daily), Version 3." *NOAA National Climatic Data Center*, 10, p. V5D21VHZ, DOI: http://dx.doi.org/10.7289/V5D21VHZ.

**Mitchell, Olivia S.** 1988. "Worker Knowledge of Pension Provisions." *Journal of Labor Economics*, 6(1): 21–39.

**Morgul, Ender Faruk, and Kaan Ozbay.** 2015. "Revisiting Labor Supply of New York City Taxi Drivers: Empirical Evidence from Large-Scale Taxi Data." In *Transportation Research Board 94th Annual Meeting*. 15.

**Nguyen, Quang, and Pingsun Leung.** 2013. "Revenue Targeting in Fisheries: The Case of Hawaii Longline Fishery." *Environment and Development Economics*, 18(5): 559–575.

**O'Donoghue, Ted, and Charles Sprenger.** 2018. "Reference-Dependent Preferences." In *Handbook of Behavioral Economics: Foundations and Applications 1*. eds. by B Douglas Bernheim, Stefano DellaVigna, and David Laibson: Elsevier, , Chap. 1 1–78.

**Oettinger, Gerald S.** 1999. "An Empirical Analysis of the Daily Labor Supply of Stadium Vendors." *Journal of Political Economy*, 107(2): 360–392.

**Park, R Jisung, Joshua Goodman, Michael Hurwitz, and Jonathan Smith.** 2020. "Heat and Learning." *American Economic Journal: Economic Policy*, 12(2): 306–39.

**Patnaik, Arpita, Joanna Venator, Matthew Wiswall, and Basit Zafar.** 2020. "The Role of Heterogeneous Risk Preferences, Discount Rates, and Earnings Expectations in College Major Choice." *Journal of Econometrics*.

**Pope, Devin G, and Maurice E Schweitzer.** 2011. "Is Tiger Woods Loss Averse? Persistent Bias in the Face of Experience, Competition, and High Stakes." *American Economic Review*, 101(1): 129–57.

**Schmidt, Marc-Antoine.** 2018. "The Daily Labor Supply Response to Worker-Specific Earnings Shocks."Technical report, Working Paper.

**Shefrin, Hersh M., and Richard H. Thaler.** 1992. "Mental Accounting, Saving, and Self-Control." In *Choice Over Time*. eds. by George Lowenstein, and Jon Elster: New York: Russell Sage Foundation Press.

**Sheldon, Michael.** 2016. "Income Targeting and the Ridesharing Market." *Unpublished manuscript. Available at: https://static1. squarespace. com/static/56500157e4b0cb706005352d*, 56, p. 1457131797556.

**Stafford, Tess M.** 2015. "What Do Fishermen Tell Us that Taxi Drivers Do Not? An Empirical Investigation of Labor Supply." *Journal of Labor Economics*, 33(3): 683–710.

**Thakral, Neil, and Linh T Tô.** 2021. "Daily Labor Supply and Adaptive Reference Points." *American Economic Review*, 111(8): 2417–43.

**Tversky, Amos, and Daniel Kahneman.** 1992. "Advances in Prospect Theory: Cumulative Representation of Uncertainty." *Journal of Risk and Uncertainty*, 5(4): 297–323.

**Warner, John T, and Saul Pleeter.** 2001. "The Personal Discount Rate: Evidence from Military Downsizing Programs." *American Economic Review*, 91(1): 33–53.

# Tables

Table 1: Robustness Check: Non-Linear Least Squares

| Model Parameters | (1)<br>Pr(Stop) | (2)<br>Pitches/Half Hour |
|---|---|---|
| Optimal Cutoff | 0.11 | 0 |
| Slope Before Cutoff | 0.00074*** | -0.0132** |
| | (0.00026) | (0.0056) |
| Slope Change After Cutoff | 0.0031*** | -0.0309*** |
| | (0.0012) | (0.0105) |
| Intercept Shift at Cutoff | 0.0054*** | 0.0057 |
| | (0.0022) | (0.0270) |
| Constant | -0.0011*** | -0.0097 |
| | (0.0004) | (0.0117) |
| Ratio of Slopes | 5.2 | 3.3 |

Robust standard errors in parentheses
*** p<0.01, **p<0.05, * p<0.1

Source: Author's calculations of data from a pest control sales company.
Notes: Estimates use the residualized outcome variables from a regression on all fixed effects and controls in the non-linear estimates. Standard errors clustered at the seller level.

Table 2: Robustness Check: Parametric Model Adding Exertion Margin as Control

| | (1)<br>Pr(Stop) | (2)<br>Pitches/Half Hour |
|---|---|---|
| Cumulative Pitches | -0.0005*** | 0.0261*** |
| | (0.000007) | (0.0013) |
| Slope Before Cutoff | 0.0019*** | -0.0175 |
| | (0.0005) | (0.0079) |
| Slope Change at Cutoff | 0.0035*** | -0.00005 |
| | (0.0013) | (0.012) |
| Intercept Shift at Cutoff | 0.0007 | 0.0144 |
| | (0.0018) | (0.0244) |
| Ratio of Slopes | 2.8 | 0.997 |

Robust standard errors in parentheses
*** p<0.01, **p<0.05, * p<0.1

Source: Author's calculations of data from a pest control sales company.
Notes: Results are from estimating Equation 2 but the model includes a control cumulative pitches that day. This adjusts for any effects of fatigue from working more intensely. Standard errors clustered at the seller level.

Table 3: Regression Evidence of Goal Setting

| Panel A: Average Daily Sales in Early Weeks | | |
|---|---|---|
| Total Sales at End of Season | Weeks 1–2 | Weeks 1–5 |
| Average Daily Sales | 95.91*** | 91.32*** |
| | (3.785) | (2.129) |
| Observations | 33,728 | 36,857 |
| R-squared | 0.752 | 0.872 |

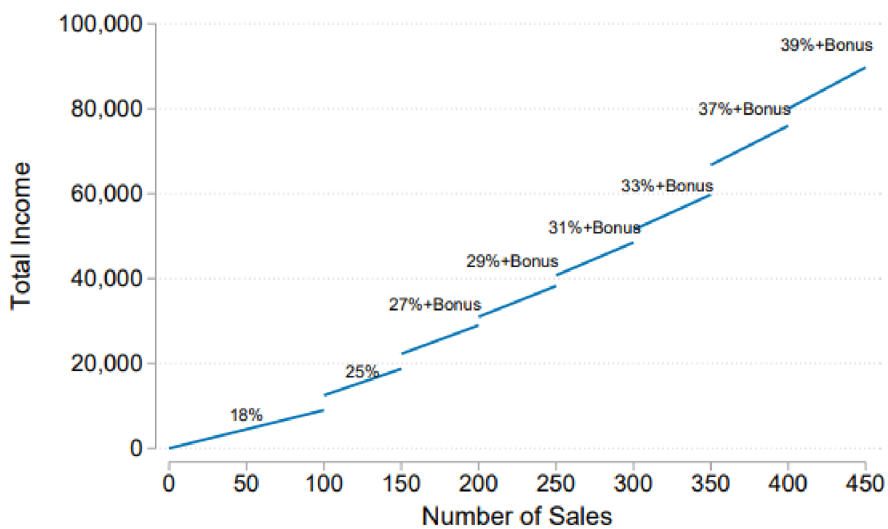
| Panel B: Sales and Hours, Days that Exceeded Expectations or Not | | |
|---|---|---|
| Hours Worked Per Day | Did Not Exceed | Exceeded Expectations |
| Sales | 0.441*** | 0.335*** |
| | (0.0179) | (0.0101) |
| Observations | 37,977 | |
| R-squared | 0.266 | |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Source: Author's calculations of data from a pest control sales company.
Notes: Panel A is from a regression of sellers' total sales at the end of the season on average daily sales during the first two or five weeks of the season. Panel B is from Equation 3 and includes fixed effects for seller, day of the week, week of the season, and year. Standard errors are clustered at the seller level.

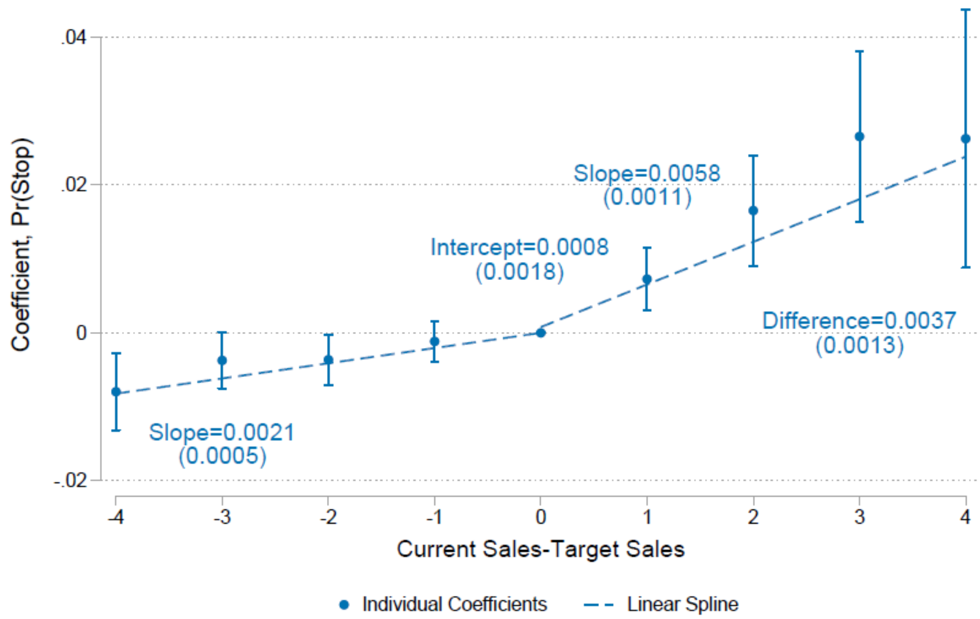
# Figures

Figure 1: Contract Structure: Total Income by Sales ($500 Contract Value)

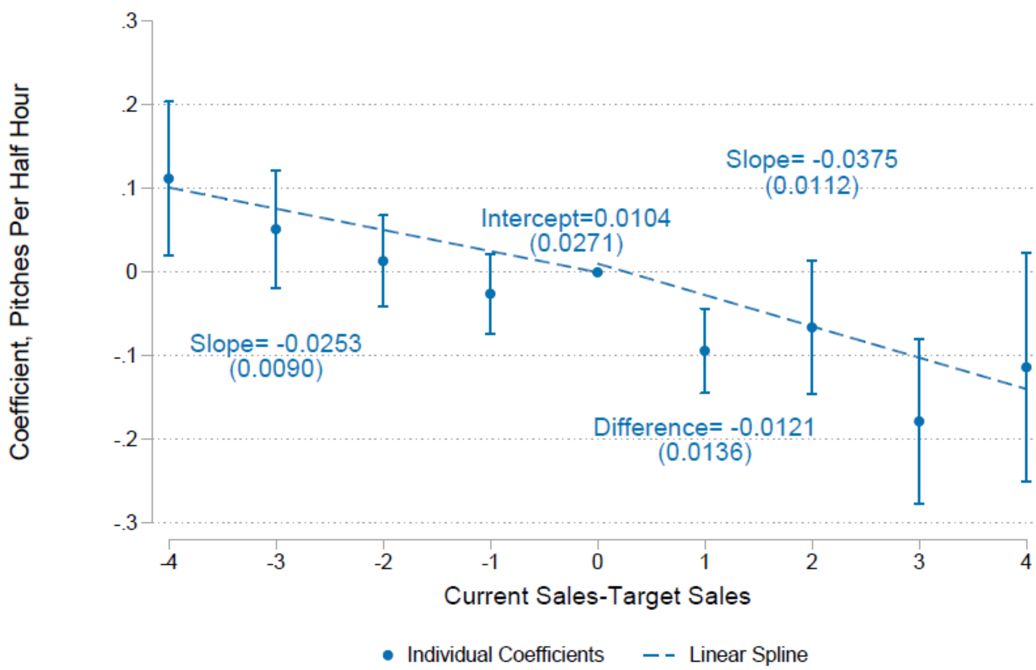



Source: Author's calculations of typical contracts from a pest control sales company.
Notes: Percentages indicate commissions as they are applied to each interval for all sales at the end of the season. At 150 sales, the "bonus" is that the company pays for the seller's rent for the summer in full (≈$2,000). At 250 sales, sellers qualify for the all-expenses-paid company vacation.

Figure 2: Labor Supply Around Expectations
Panel A: Probability of Stopping for the Day

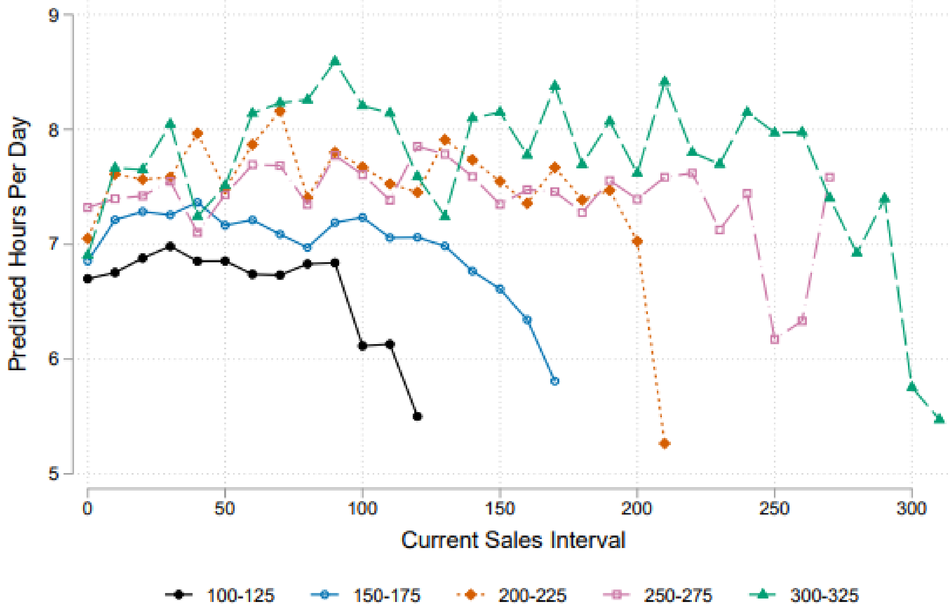



Panel B: Pitches Per Half Hour



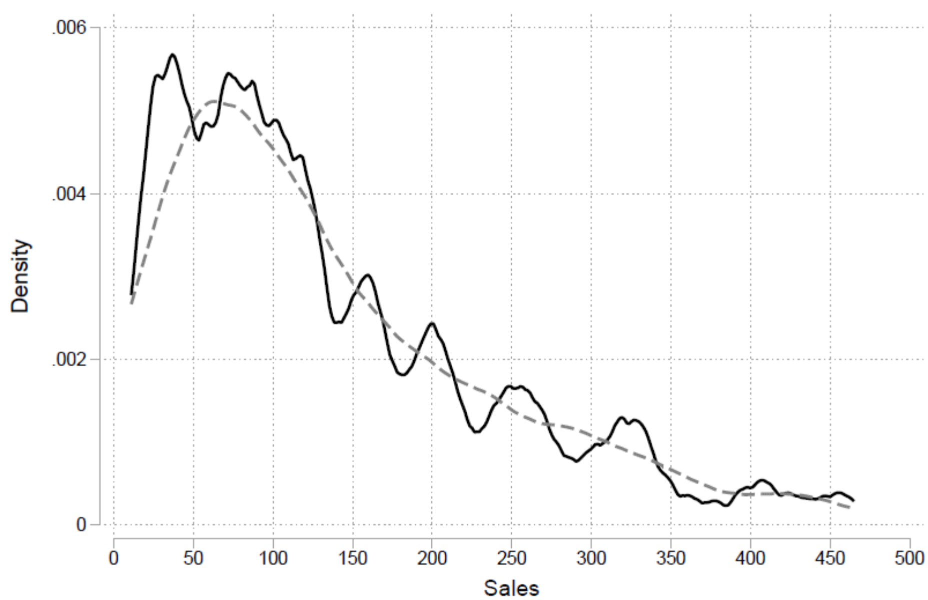

Source: Author's calculations of data from a pest control sales company.
Notes: Results are from estimates of Equations 1 and 2. Standard errors are clustered at the seller level. Individual coefficients are rounded to the nearest integer. The coefficient on +1 includes the interval (0,1.5).

Figure 3: Predicted Labor Supply Over Current Sales Interval, By Final Season Sales Interval



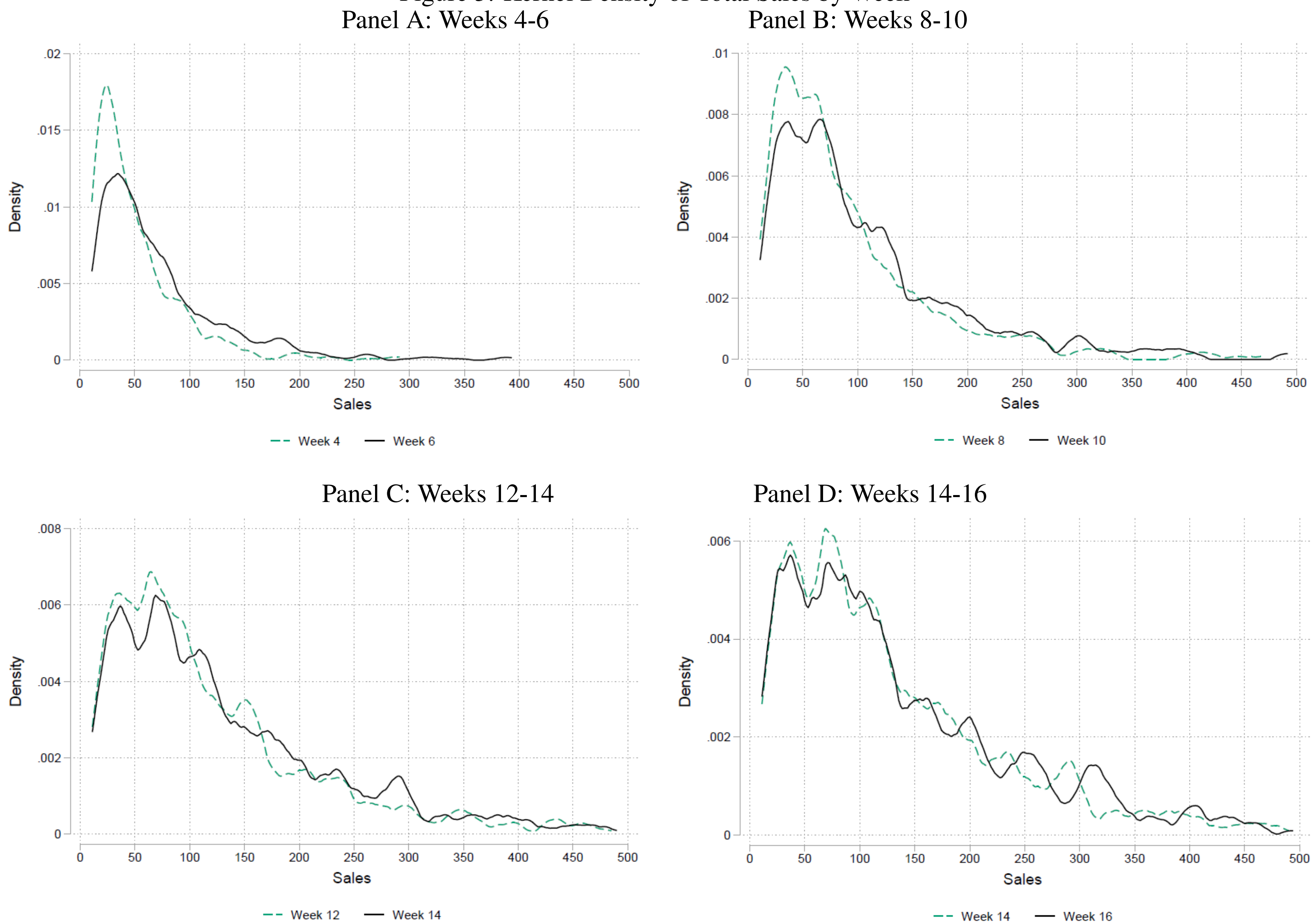

Source: Author's calculations of data from a pest control sales company.
Notes: Plot shows predicted hours from specification in Equation 4 for current sales interval (x-axis) separated by bins of total end-of-season sales.

Figure 4: Kernel Density of Total Sales at End of Season



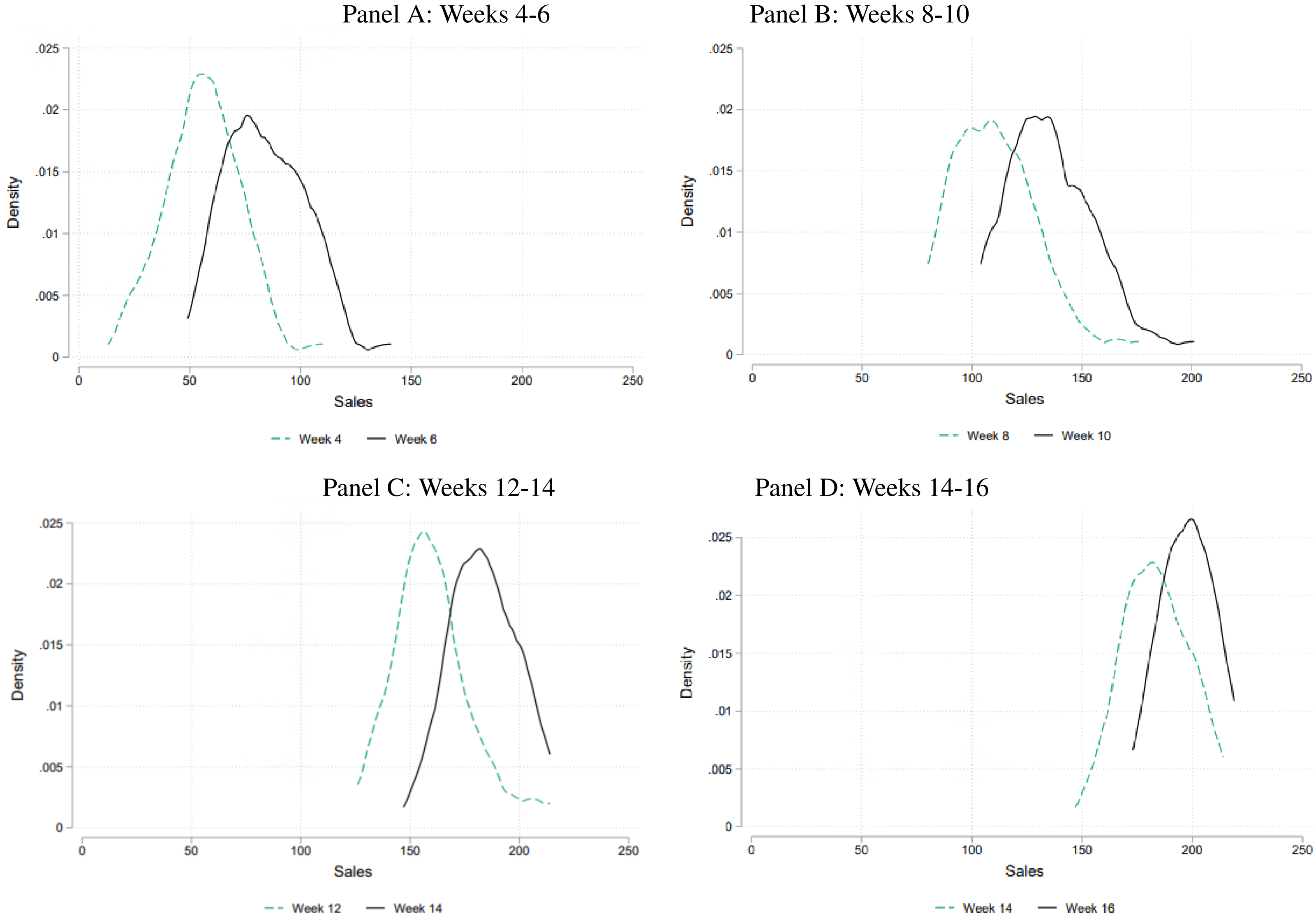

Source: Author's calculations of data from a pest control sales company.
Notes: Estimated using Epanechnikov kernel with bandwidth of 7 sales and 25 sales for sellers with at least ten sales and fewer than 500. The retroactive nature of the commission increases leads to a cash bonus upon hitting each 50-sale interval. At 150 sales, the company pays for the seller's rent for the summer in full (≈$2,000). At 250 sales, sellers qualify for the all-expenses-paid company vacation.

Figure 5: Kernel Density of Total Sales by Week



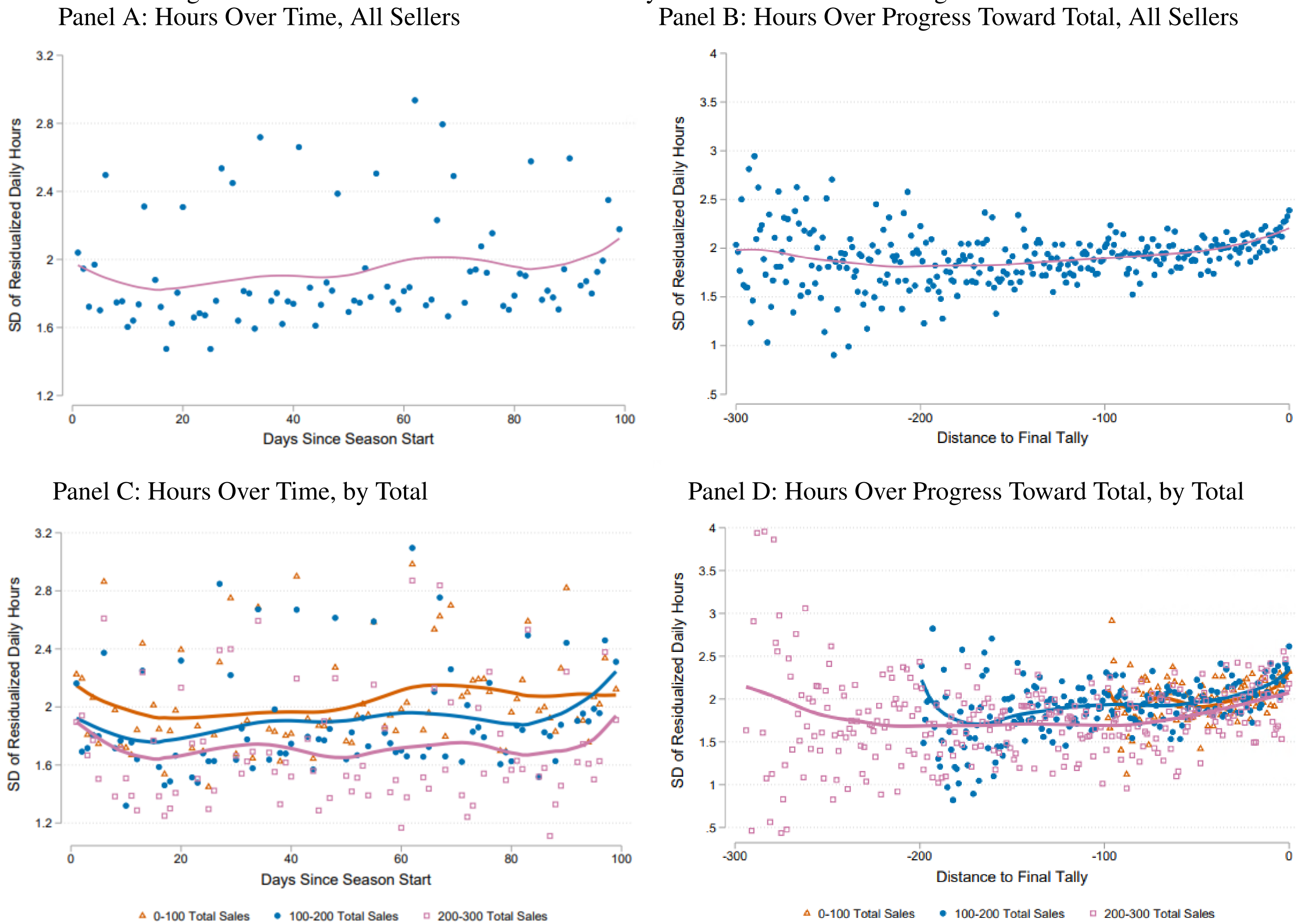

Source: Author's calculations of data from a pest control sales company.
Notes: Estimated using Epanechnikov kernel with bandwidth of 7 sales at the end of each estimated week among those with at least ten sales.

Figure 6: Kernel Density of Total Sales by Week
Workers with Total Sales of 175–225 at End of Season




Source: Author's calculations of data from a pest control sales company.
Notes: Estimated using Epanechnikov kernel with bandwidth of 7 sales at the end of each estimated week among those with at least ten sales.

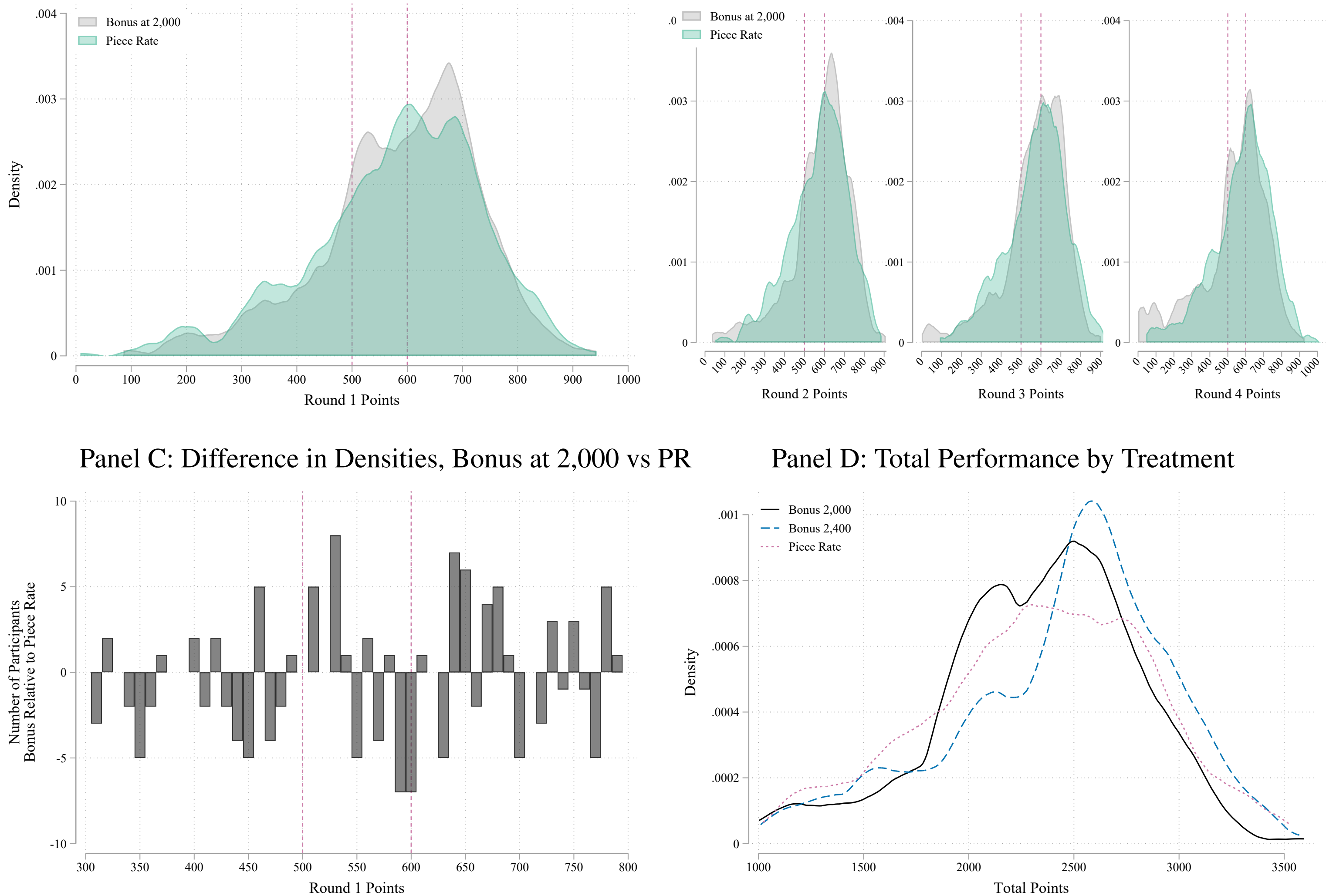
Figure 7: Standard Deviation of Residual Daily Hours Over Time and Progress Toward Total




Source: Author's calculations of data from a pest control sales company.
Notes: Residuals come from a regression of daily sales or hours on seller and day-of-the-week fixed effects. Lines represent LOWESS smoothing.

Figure 8: Density Estimates of Experimental Performance

Panel A: Round 1 Densities, Piece Rate vs Bonus at 2,000 Panel B: Densities: PR vs Bonus by Round

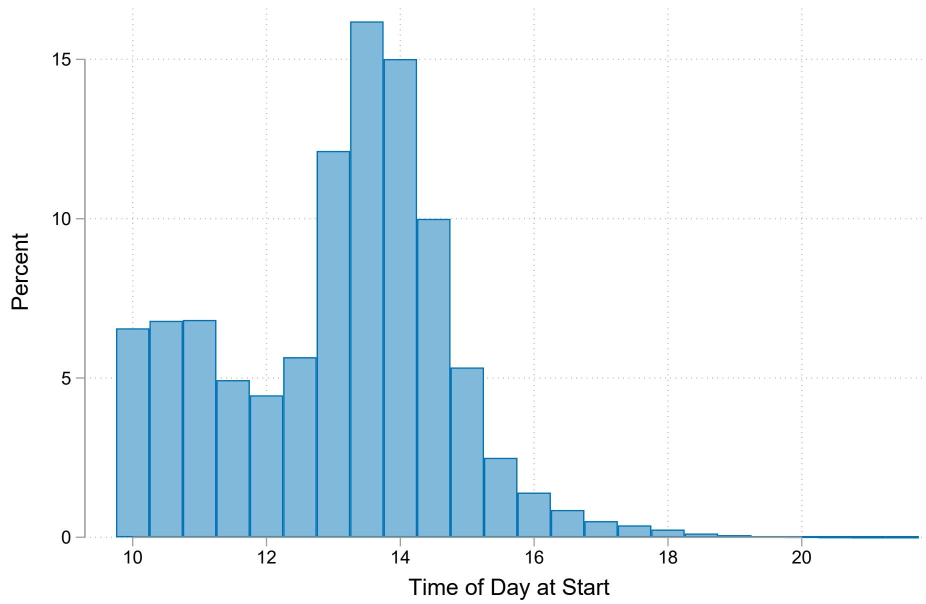



Source: Author's calculations of data from an online experiment.

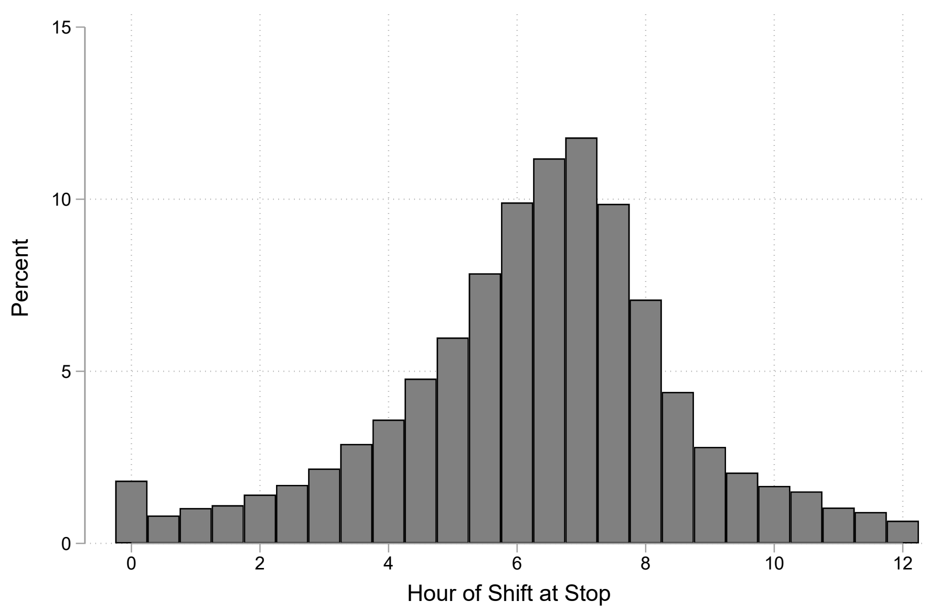
Notes: Panel C breaks frequencies down into discrete 10-point categories and presents the differences between the distributions in Panel A. In Panel C, the X-axis values reflect the minimum of each 10-point interval.

# A Online Appendix: Figures and Tables

Figure A1: Distribution of Start and Stop Characteristics
Panel A: Time of Day at Start of Shift

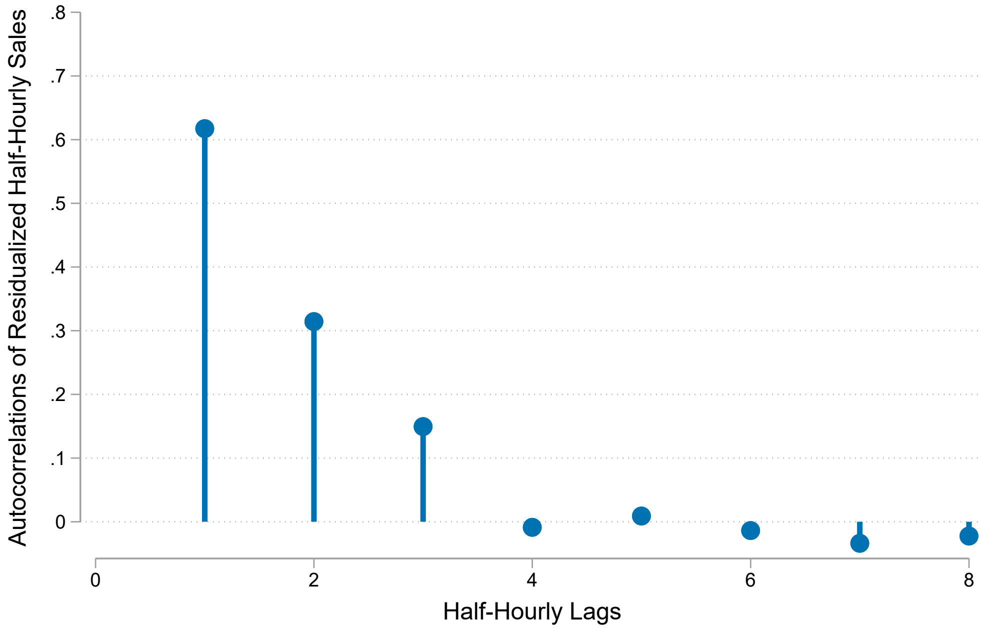



Panel B: Hour of Shift at Stop

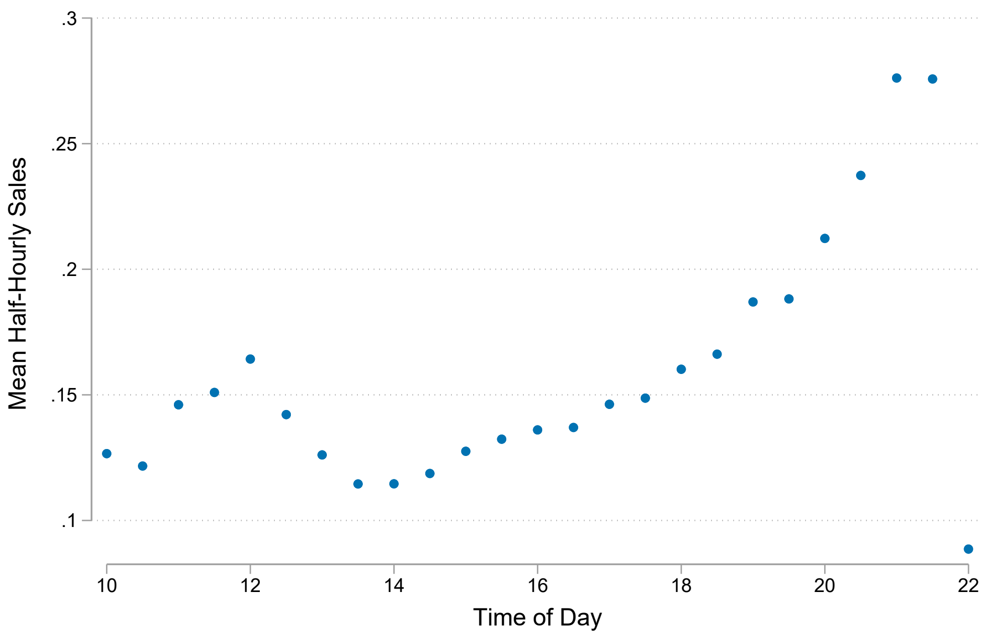



Source: Author's calculations of data from a pest control sales company.
Notes: Shifts begin during the half hour period when a seller first registers a knock or sale on each workday. Shifts end during the half hour they record their last sale or knock for the day.

Figure A2: Upward Pressures on Labor Supply During the Day
Panel A: Autocorrelation of Sales

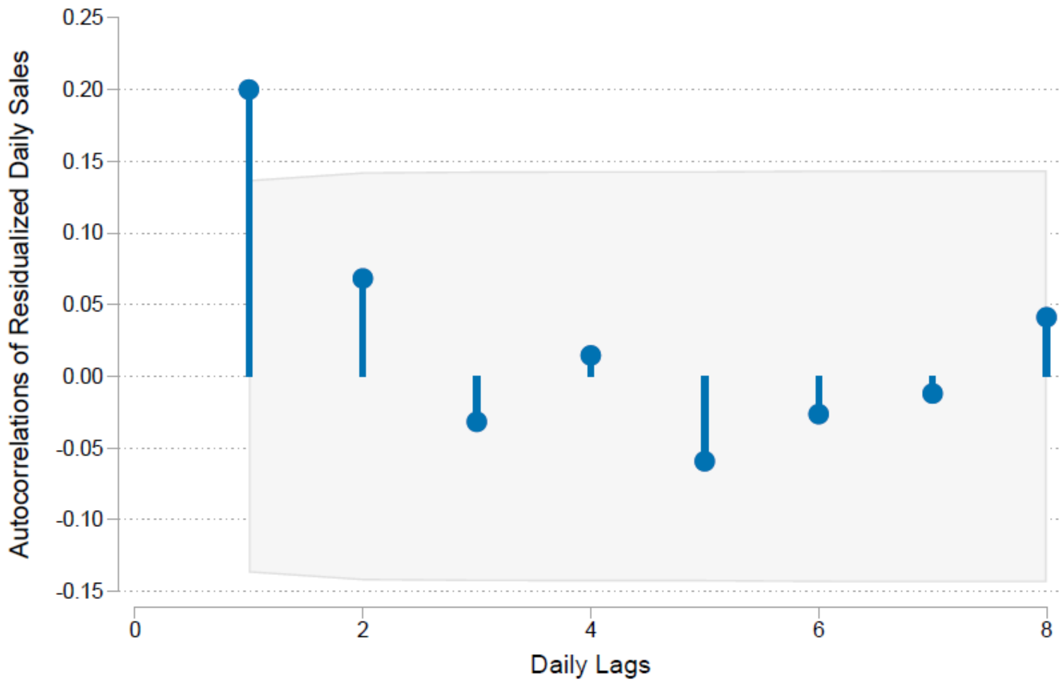



Panel B: Mean Sales by Half Hour



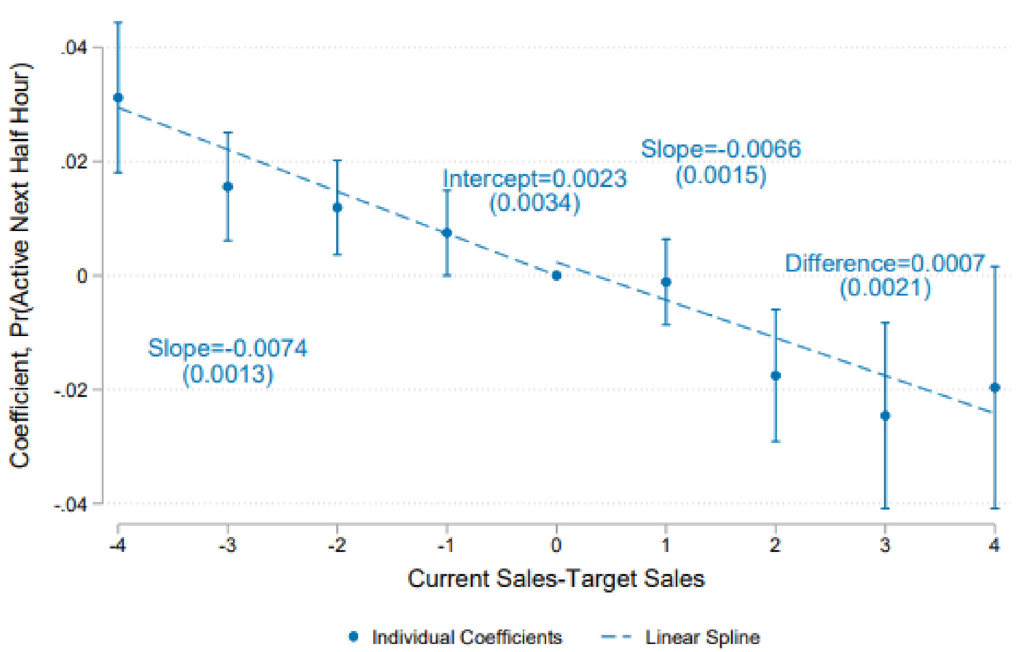

Source: Author's calculations of data from a pest control sales company.
Notes: In Panel A, residualized sales come from a regression of sales each half hour on seller, half-hour-of-the-day, day-of-the-week, week-of-season, and year fixed effects as well as controls for having any knocks recorded that half hour, weather, and ZIP code characteristics. I then calculate the autocorrelation for those predicted residuals for half hour lags of one through eight.

Figure A3: Autocorrelation in Daily Sales




Source: Author's calculations of data from a pest control sales company.
Notes: This figure uses the seller-day panel to calculate residualized sales. I regression of sales each day on seller, day-of-the-week, week-of-the-season, and year fixed effects as well as controls for having any knocks recorded that day, weather, and ZIP code characteristics. I then calculate the autocorrelation for those predicted residuals for lags of one through eight days. The shaded region shows Bartlett's formula for MA(q) 95% confidence bands. The low autocorrelation between days indicates that performance today is not strongly predictive of performance tomorrow, or that individual workdays come from independent draws.

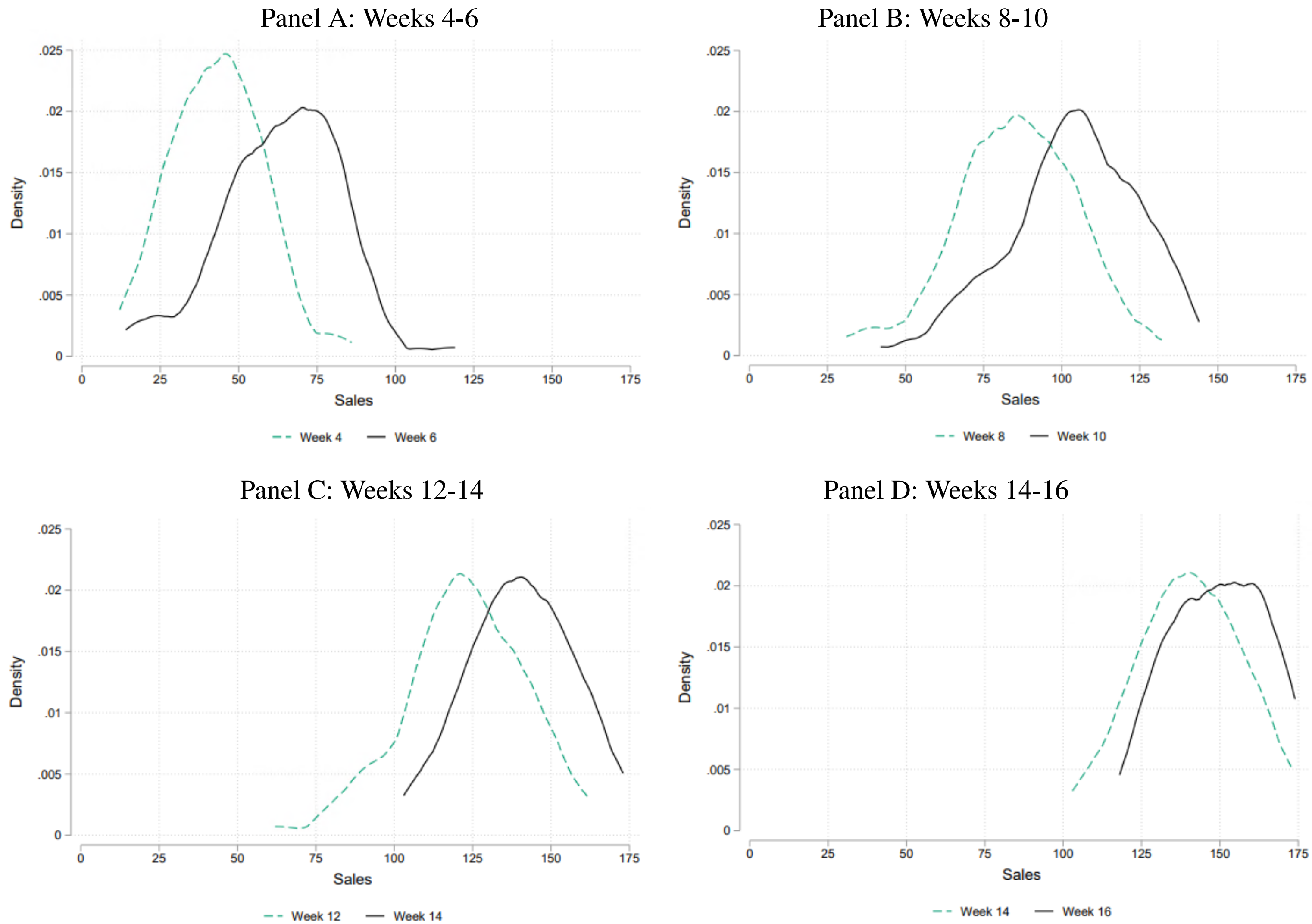
Figure A4: Reference Dependence and Probability of Active Work




Source: Author's calculations of data from a pest control sales company.
Notes: Results are from estimates of Equations 1 and 2 for the probability of working during the next half hour of the day. Standard errors are clustered at the seller level. Individual coefficients are rounded to the nearest integer. The coefficient on +1 includes the interval (0,1.5). Model also includes an additional control for actively knocking during the current half-hour period. At a base active knocking share of 80% of all half-hour periods, an increase of 1 sale above or below expectations decreases the probability of actively knocking by approximately 0.74 percentage points, or approximately 0.093%. Sellers are not more likely to take breaks during their work as a function of their position relative to expectations.

Figure A5: Kernel Density of Total Sales by Week

Workers with Total Sales of 125–175 at End of Season

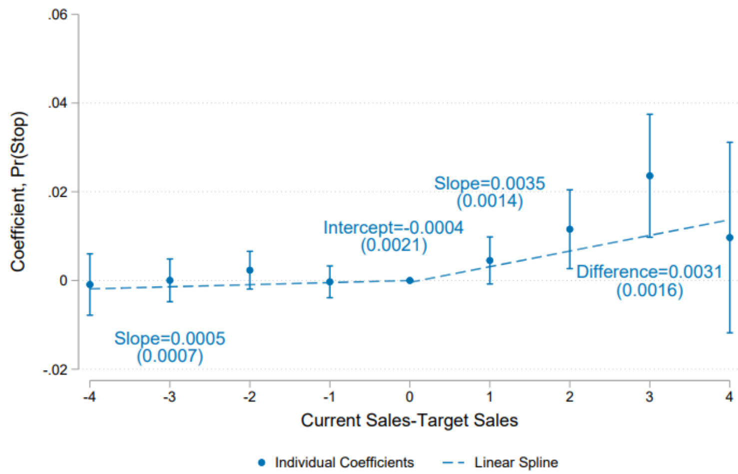



Source: Author's calculations of data from a pest control sales company.

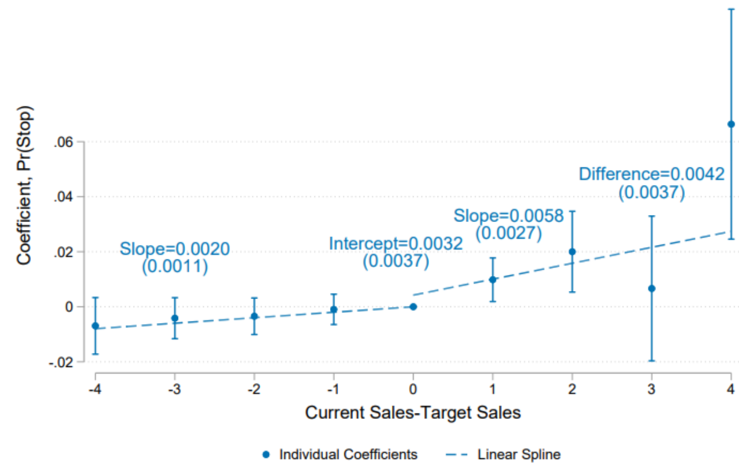
Notes: Estimated using Epanechnikov kernel with bandwidth of 7 sales at the end of each estimated week among those with at least ten sales.

Figure A6: Estimates of Stopping Probability by Month

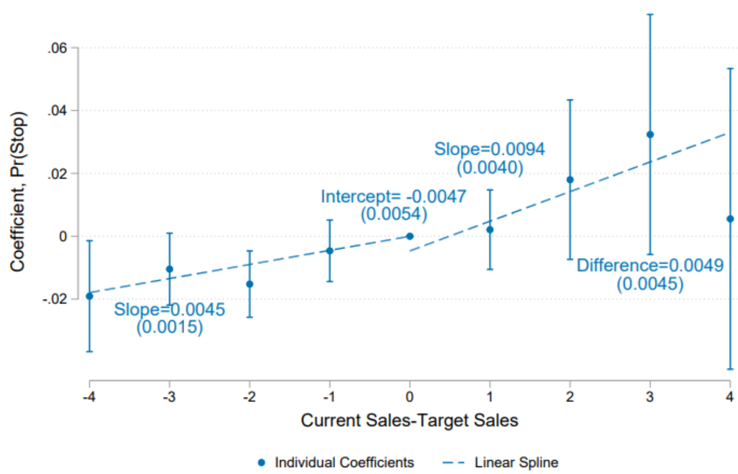
Panel A: May




Panel B: June



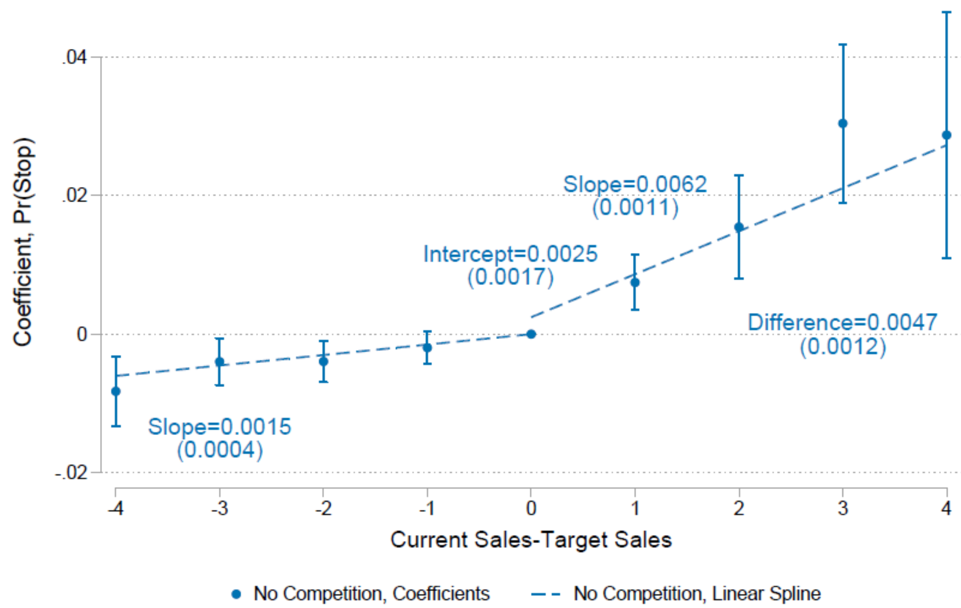

Panel C: July



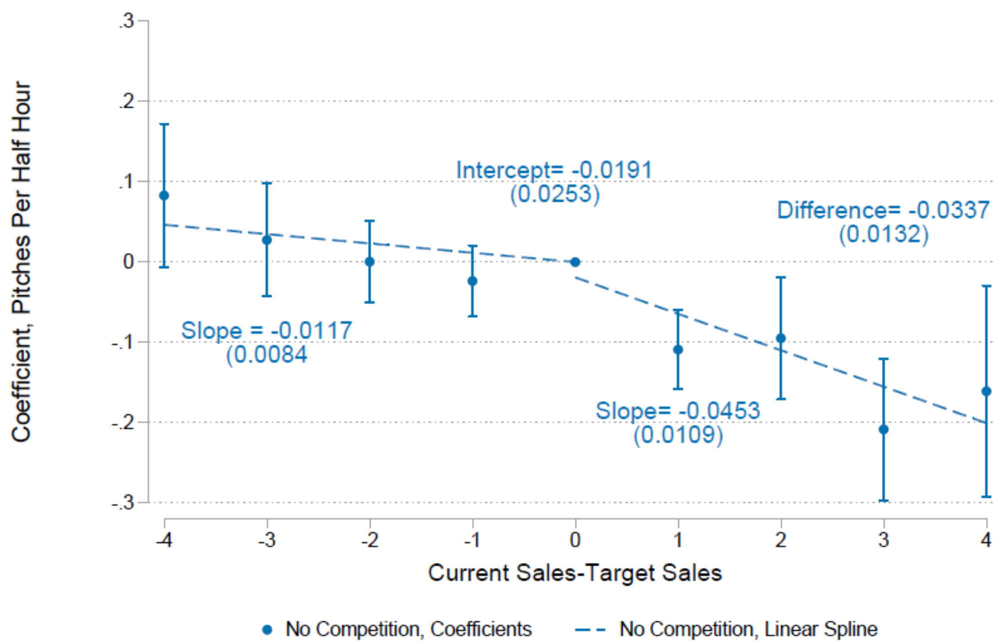

Source: Author's calculations of data from a pest control sales company.

Notes: Results are from estimates of Equations 1 and 2 separated by calendar month. Standard errors are clustered at the seller level. Individual coefficients are rounded to the nearest integer. The coefficient on +1 includes the interval (0,1.5).

Figure A7: Robustness Test: Pooled Estimates with Tournament/Non-Tournament Interactions
Panel A: Probability of Stopping for the Day



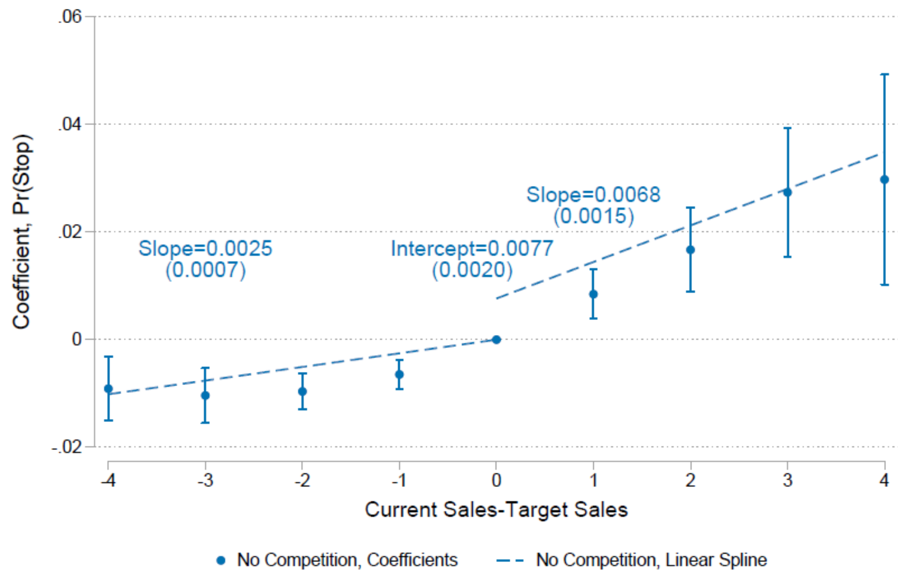

Panel B: Pitches Per Half Hour



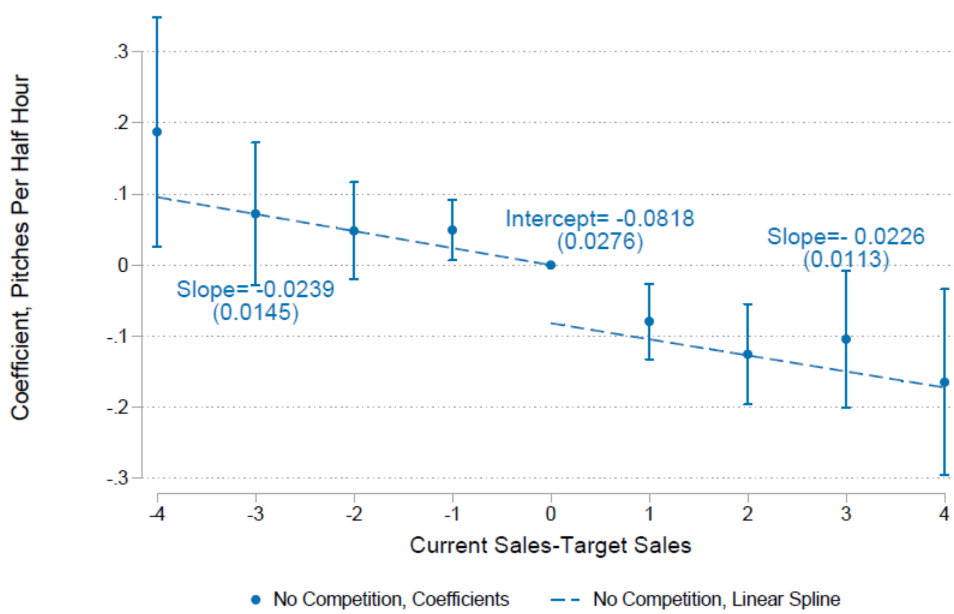

Source: Author's calculations of data from a pest control sales company.
Notes: Results are from estimates of Equations 1 and 2 but pooling all data and including interactions between indicators for tournament type or non-tournament days and distance to the target. Reported results are for non-tournament interactions. Standard errors are clustered at the seller level. Individual coefficients are rounded to the nearest integer. The coefficient on +1 includes the interval (0,1.5).

Figure A8: "Goal-Based" Reference

Panel A: Probability of Stopping for the Day



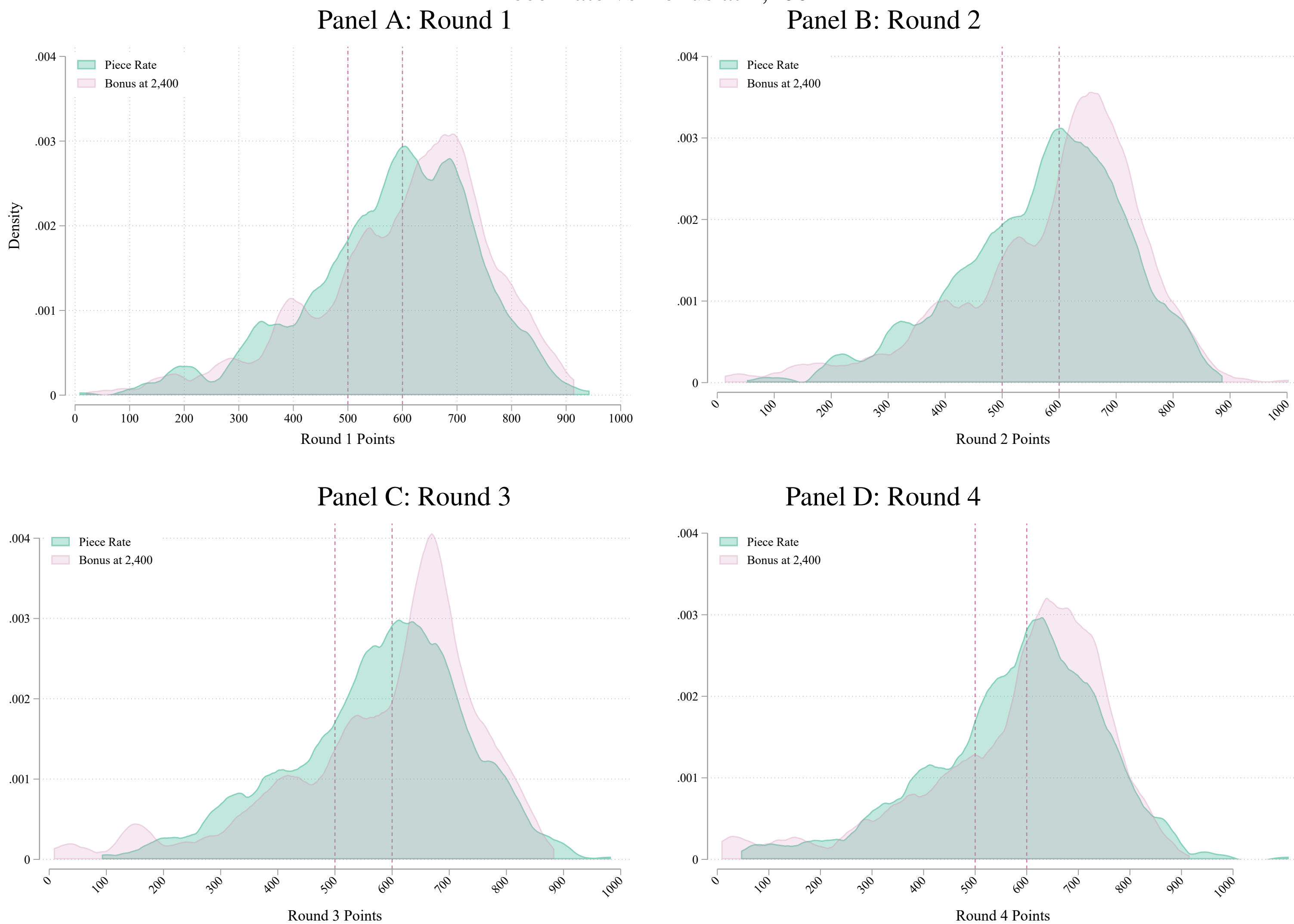

Panel B: Pitches Per Half Hour




Source: Author's calculations of data from a pest control sales company.

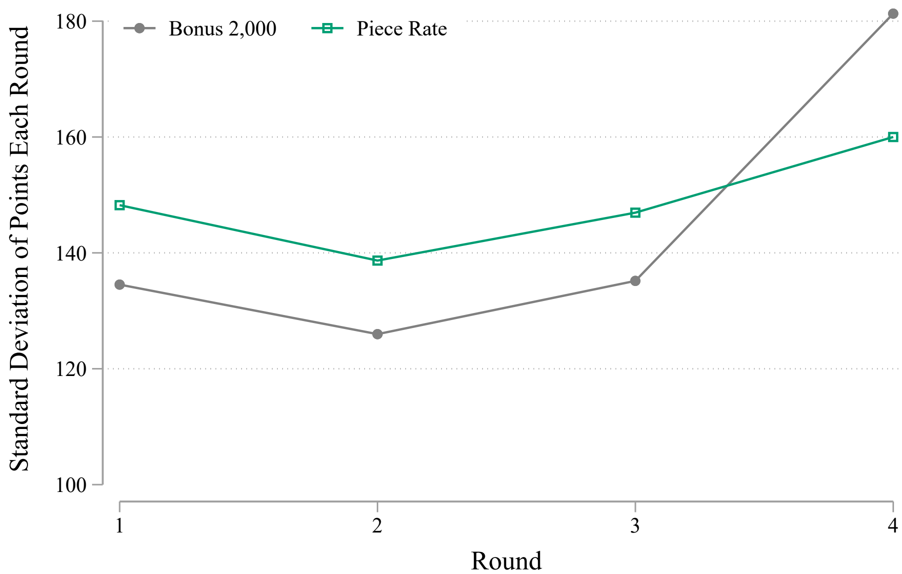
Notes: Results are from estimates of Equations 1 and 2 but pooling all data and including interactions between indicators for tournament type or non-tournament days and distance to the target. Reported results are for non-tournament interactions. The target in these models is a projection of the first two weeks of performance to the nearest bonus threshold at the end of the season. Standard errors are clustered at the seller level. Individual coefficients are rounded to the nearest integer. The coefficient on +1 includes the interval (0,1.5).

Figure A9: Density Estimates of Experimental Performance
Piece Rate vs Bonus at 2,400

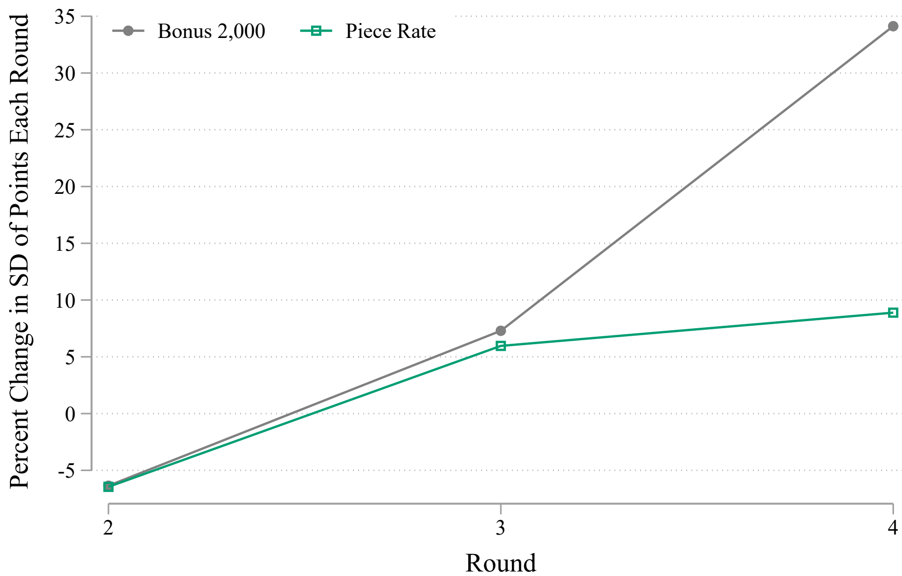



Source: Author's calculations of data from an online experiment.
Notes: These comparisons follow those in Figure 8 comparing the piece rate treatment to the bonus at 2,000 points treatment.

Figure A10: Variance in Each Round, Piece Rate vs Bonus at 2,000
Panel A: Standard Deviation by Round

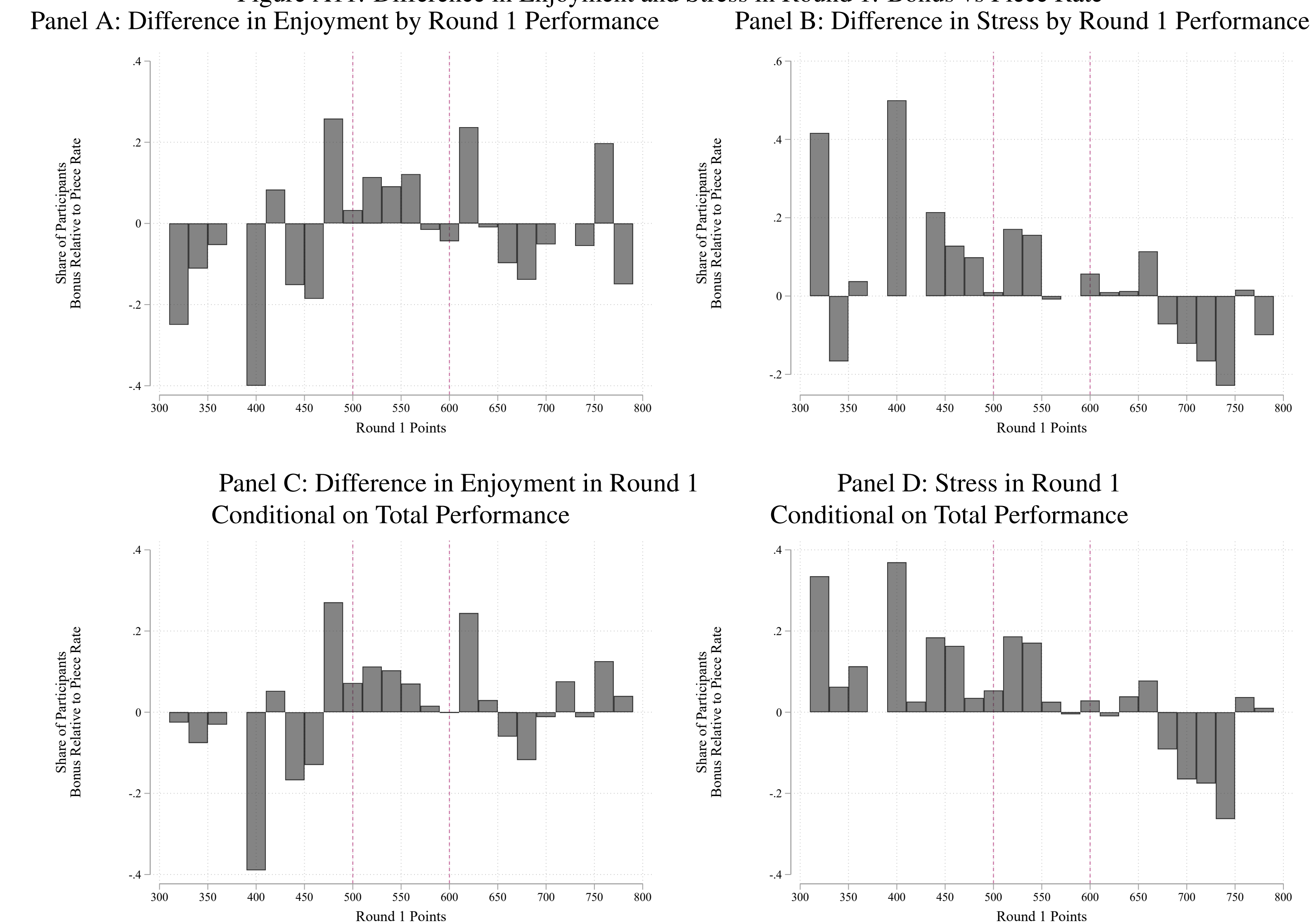



Panel B: Percent Change in Standard Deviation by Round




Source: Author's calculations of data from an online experiment.
Notes: Panel B changes are measured relative to the prior round.

Figure A11: Difference in Enjoyment and Stress in Round 1: Bonus vs Piece Rate




Source: Author's calculations of data from an online experiment.

Notes: The two measures of enjoyment and stress are based on answering "agree" or "strongly agree" that they enjoyed the task or felt stress. Total performance is controlled non-parametrically with bins for every 50 total points at the end of the task (Panels C and D).

Table A1: Summary Statistics of Key Variables

| Panel A: Half-Hourly Panel | Mean | SD |
|---|---|---|
| Pr(stop) | 0.074 | 0.262 |
| Pitches Per Half Hour | 2.281 | 2.498 |
| Sales Per Half Hour | 0.156 | 0.419 |

| Panel B: Daily Panel | Mean | SD |
|---|---|---|
| Sales Per Day | 2.02 | 2.20 |
| **Labor Supply** | | |
| Pitches Per Day | 31.21 | 19.63 |
| Hours Per Day | 6.94 | 2.23 |
| Average Sales Specific to Day of the Week | 1.99 | 1.60 |
| **Weather** | | |
| Precipitation (1/10th MM) | 4.00 | 8.52 |
| High Temperature (Celsius) | 26.85 | 5.00 |
| Low Temperature (Celsius) | 15.29 | 4.97 |
| **Select ZIP Code Characteristics** | | |
| Median HH Income | 85,945 | 25,385 |
| % HH Income $100,000-$150,000 | 19.49 | 4.69 |
| % Residents Living in Same Home From Last Year | 88.19 | 4.41 |
| Total Housing Units | 112,203 | 5,766 |
| % Housing Units Single-Family Homes | 80.08 | 11.85 |
| Median Home Value | 258,083 | 107,492 |
| % Non-Hispanic White | 80.36 | 13.71 |
| % Bachelors Degree or More | 44.93 | 14.74 |
| Total Sellers | 512 | |
| Total Days | 180 | |
| Total Half-Hourly Observations | 458,558 | |
| Total Daily Observations | 37,984 | |

Source: Author's calculations of data from a pest control sales company, NOAA daily weather data, and ACS data on ZIP codes.

Table A2: Test of Location Sorting

| Sales Per Day, All Significant Coefficients | (1)<br>ACS | (2)<br>Weather | (3)<br>Both |
|---|---|---|---|
| % Non-Hispanic Black | 0.0313* | | 0.0316* |
| | (0.0161) | | (0.0162) |
| % Single Mothers | -0.0833** | | -0.833** |
| | (0.0403) | | (0.0403) |
| % House Value $100,000-$200,000 | -0.0276** | | -0.0271* |
| | (0.0140) | | (0.0139) |
| Precipitation (1/10th MM) | | -0.00507*** | -0.00635*** |
| | | (0.00152) | (0.00147) |
| High Temperature (Celsius) | | 0.0209** | 0.0188** |
| | | (0.00774) | (0.00788) |
| Low Temperature (Celsius) | | -0.0131 | -0.0142 |
| | | (0.0107) | (0.0108) |
| Observations | 37,508 | 37,943 | 37,467 |
| R-squared | 0.029 | 0.013 | 0.031 |
| F-Statistic | 1.59 | 9.724 | 3.782 |
| prob>F | 0.054 | 0 | 0 |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Source: Author's calculations of data from a pest control sales company, the American Community Survey 2013-2017 5-year ZIP code estimates, and daily weather data from NOAA.
Note: Results are from regression of observed ZIP code characteristics from the American Community Survey (ACS) and dailiy weather data on sales generated per day, including day-of-the-week, week-of-the-seaon, and year fixed effects. Standard errors clustered at the seller level. Non-significant coefficients on % Non-Hispanic White, % Hispanic; % of households with income $50,000-$75,000, $100,000-$150,000, and >$200,000; median household income, poverty rate, unemployment rate, % adults with Bachelors degree or more, % households in the same home as last year; total housing units, % of housing units that are single-family homes; % homes with value $100,000-$200,000, $200,000-$300,000, $300,000-$500,000, and $500,000-$1 million and median home value.

Table A3: Non-Parametric Estimates
Expectations-Based References

| Distance to Expectations | (1) Pr(Stop) | (2) Pitches Per Half Hour |
|---|---|---|
| -8 | -0.0128** | 0.174 |
| | (0.00535) | (0.169) |
| -7 | -0.00577 | 0.00429 |
| | (0.00472) | (0.0765) |
| -6 | -0.00812* | 0.0403 |
| | (0.00421) | (0.0888) |
| -5 | -0.0130*** | 0.120* |
| | (0.00290) | (0.0688) |
| -4 | -0.00796*** | 0.112** |
| | (0.00264) | (0.0469) |
| -3 | -0.00373* | 0.0515 |
| | (0.00195) | (0.0360) |
| -2 | -0.00363** | 0.0134 |
| | (0.00174) | (0.0277) |
| -1 | -0.00115 | -0.0257 |
| | (0.00138) | (0.0242) |
| 1 | 0.00727*** | -0.0939*** |
| | (0.00213) | (0.0258) |
| 2 | 0.0165*** | -0.0659 |
| | (0.00381) | (0.0406) |
| 3 | 0.0266*** | -0.178*** |
| | (0.00586) | (0.0503) |
| 4 | 0.0263*** | -0.114 |
| | (0.00892) | (0.0699) |
| 5 | 0.00574 | -0.224** |
| | (0.0132) | (0.108) |
| 6 | 0.00874 | -0.244** |
| | (0.0177) | (0.107) |
| 7 | 0.0254 | -0.420** |
| | (0.0365) | (0.163) |
| 8 | 0.0473 | -0.209 |
| | (0.0426) | (0.254) |
| 9 | 0.0704 | 0.733 |
| | (0.0542) | (0.569) |
| 10 | 0.199*** | -0.819*** |
| | (0.0738) | (0.294) |

Robust standard errors in parentheses
*** p<0.01, **p<0.05, * p<0.1

Note: Results are from regression in Equation 1 and coincide with estimates from Figure A7. Standard errors clustered at the seller level.

Table A4: Parametric Estimates of Stopping Probability
Pooled Estimates with Interactions for Tournament/Non-Tournament

Panel A: Expectations-Based References

| | (1) Slope Below Reference | (2) Slope Change Above Reference | (3) Intercept Shift at Reference | (4) Ratio of Slopes [(Change Above + Below)/Below] |
|---|---|---|---|---|
| No Competition | 0.00151*** | 0.00470*** | 0.00244 | 4.113 |
| | (0.00042) | (0.00124) | (0.00170) | |
| Individual Competitions | 0.000333 | 0.00003 | 0.00379 | 1.090 |
| | (0.00058) | (0.00302) | (0.00399) | |
| Team Competitions | 0.00242*** | 0.00055 | 0.00997*** | 1.227 |
| | (0.00048) | (0.00194) | (0.00227) | |
| Benchmark Competitions | 0.0014** | 0.00011 | 0.00849** | 1.079 |
| | (0.00058) | (0.0030) | (0.00376) | |

Panel B: Goal-Based References

| | (1) Slope Below Reference | (2) Slope Change Above Reference | (3) Intercept Shift at Reference | (4) Ratio of Slopes [(Change Above + Below)/Below] |
|---|---|---|---|---|
| No Competition | 0.00252*** | 0.00431** | 0.00768*** | 2.710 |
| | (0.00073) | (0.00180) | (0.00206) | |
| Individual Competitions | 0.000238 | 0.00237 | 0.00311 | 10.958 |
| | (0.00099) | (0.00292) | (0.00433) | |
| Team Competitions | 0.00360*** | 0.00119 | 0.0029 | 1.331 |
| | (0.00086) | (0.0022) | (0.00258) | |
| Benchmark Competitions | 0.00316*** | -0.00114 | 0.00207 | 0.639 |
| | (0.00098) | (0.00271) | (0.00344) | |

Robust standard errors in parentheses
*** p<0.01, **p<0.05, * p<0.1

Note: Results are from regression in Equation 2 but include interactions between indicators for each tournament/non-tournament period and distance to the reference. Standard errors clustered at the seller level.

# B Online Appendix: Data

The pest control sales company data were obtained through a data use agreement prohibiting disclosure of the company's identity or intimate details of their operations.

The data cover the entirety of all sales and knocks recorded from January 2018 to January 2020. Sales in the "off-season" are not compensated the same way as they are during the summer, and there are very few recorded knocks in their system. Most sales the company generates during the off-season are renewals of current contracts for the following year as well as follow-ups with past customers, but those contacts are typically not done in person. Most off-season knocks are those done in the service of training new sellers. The knocking data are reported using their common application, which also shows leaderboards, team performance, and the performance of all other sellers in the company. The centralized sales website also contains sales information but does not include knocking information. Competition rules, dates, and prizes were collected from raw internal company documents as well as the company website usually available only to contractors and employees.

To correctly measure the incentives and behavior of these workers at the right time, I impose a few basic restrictions to my half-hourly panel. I limit my sample to the "summer sales season" each year, which is the period from the last week of April to the third week of August. This excludes trainees who arrive early, those who stay late into the end of August or early September (who are usually managers and those not enrolled in school), and off-season sales. I exclude the last two weeks of August because participation drops precipitously as sellers return to school. Less than 50% of sellers stay past August 17th-18th, and less than 25% of sellers stay past August 25th-26th. I then exclude any sellers who stopped working altogether before late May, which effectively excludes the least able sellers who averaged less than one sale per week and decided to go home after experiencing this lack of success. This group also includes managers who record knocks for training purposes during the first month. Off-season sales during the September to April months entail a different compensation structure, and many of the sales are generated by full-time employees of the company rather than the independent contractors that work during the summer.

In my half-hourly panel, I exclude observations with no previous expectations, i.e. the first week a seller is active. In all, my half-hourly panel consists of approximately 459,000 observations for 512 sellers across 180 days in 2018-2019.

# C Online Appendix: Further Background

The company whose data I use (which I call "PestCo") operates a full-service pest control service operation. In addition to removing insects, spiders, and rodents, they apply preventative treatments to prevent pests from returning or growing larvae near an individual home. There is a range of services they provide, and sellers are encouraged to "upsell" for more comprehensive services whenever they see an opportunity. Sellers are given the responsibility to generate new contracts and schedule the service with a separate wing of the company that performs the service. Most contracts last 12–18 months. Commission rates are based on the annualized value of the contracts the seller generates.

PestCo is not markedly different from the rest of the sales industry in terms of its use of incentive schemes. Their independent contractor agreements and practices are all in line with industry standards.

Sellers are paid an up-front portion ($75) of their commissions during the two-week period each sale is made, similar to a regular paycheck. The balance of commission payments is calculated at the end of the season after the status of all contracts is known. Final payouts for Spring sales are given in the Fall, and late Summer sales payouts are given at the end of the year. Most contractor agreements include penalties for leaving the selling area before the official end of the sales season or for not recording knocking activity a minimum number of days. The penalties typically stipulate that regardless of the number of sales, the commission the seller earns will return to some low base rate (usually 18–20%).

Prior to leaving for their assigned metro area, sellers at PestCo are trained in sales techniques and are given a detailed manual of behavioral tools to help them over the course of the summer. These include training on proper body language, handshaking, standards for appearance, overcoming customer objections, rephrasing customer concerns, interacting with upset neighbors, and how to look for and identify pests before approaching a door. They are provided with video examples of strong sales performance and are encouraged to review their training materials on a daily basis.

PestCo takes an active role in trying to motivate their workers. In training materials, the company encourages their sellers to be physically active and healthy, to be honest about their performance and goals, and to take accountability for their own performance and summer experience. These training materials are especially important because approximately half of the sellers who are working any given day are brand new to the company and the industry. Sellers are encouraged to learn advanced sales techniques from their teammates. Sellers share an apartment with other sellers from the company, and new sellers are asked to seek feedback from their more experienced roommates.

Work neighborhoods for each seller are assigned by a local team leader. Metro areas are divided into sections for each team, and within their section, team leaders assign sellers to a neighborhood. Work in each neighborhood continues until approximately 75% of doors have been marked in their tracking software, after which the seller can request a new area. Managers insist that "work area does not matter" in their training materials, and the evidence I present supports this argument. Area assignments, while not strictly random, are not correlated with sales outcomes in any meaningful way either across or within seller (see A2) and Section 3. Managers emphasize that making assignments to work areas based on perceived skill or other seller attributes is costly to them as managers and generates unclear returns, which undermines the business case for sorting. For example, assigning a better seller to a "harder" neighborhood may generate sales that would otherwise not take place. However, the marginal cost of achieving those may be high and the benefits may be smaller than the difference in sales *speed* across sellers in an "easier" neigbhorhood. In practice, managers do not typically spend large amounts of time on these assignments.

In addition to the high-powered cash incentives built into their contractor agreements, PestCo also runs frequent short-run tournaments for prizes valued from $300 to $3,000. These take three forms: individual rank-order, team rank-order, and what I call "benchmark" competitions. Individual rank-order tournaments pit sellers against each other for a single day, and the seller with the most head-to-head daily "wins" at the end of the two-week tournament gets a prize. Team rank-order tournaments have a similar structure but are based on wins against another team, and "wins" are based on per-seller team revenue. During "benchmark" competitions, if a seller generates more revenue during the week-long competition period than he did during any prior week in the season, he will get a prize. Prizes include merchandise like Bluetooth headphones, apparel, and expensive grills as well as "experiences" like a cruise, resort stay, or annual ski passes, though sellers have the option to cash out the value of the prize. These tournaments are not the focus this study. However, they are important for contextualizing my empirical models because they modify the incentive structure within particular sales days.

# D A Simple Model of Reference Dependence with Loss Aversion

The basic insight of models of reference dependence and loss aversion propose that losses loom larger than gains.

O'Donoghue and Sprenger (2018) present a simple model of this idea that is instructive. A worker can choose an effort level $e$, which yields output $x(e)$ and has a cost of effort $c(e)$. The function $c(e)$ is increasing and convex. Utility is linear in $x(e)$. Suppose there is an output or income reference, $r$, which can be endogenously determined by rational expectations or exogenously imposed. Distance from the reference, $x(e) - r$, enters the utility or value function:

$$U(e) \equiv x(e) + \mu\left(x(e) - r\right) - c(e) \tag{5}$$

where

$$\mu(z) = \begin{cases} \eta z & \text{if } z \geq 0 \\ \eta\lambda z & \text{if } z \leq 0 \end{cases}$$

The $\mu$ function captures "gain-loss utility." The equilibrium labor supply for this utility function with gain/loss utility is given by:

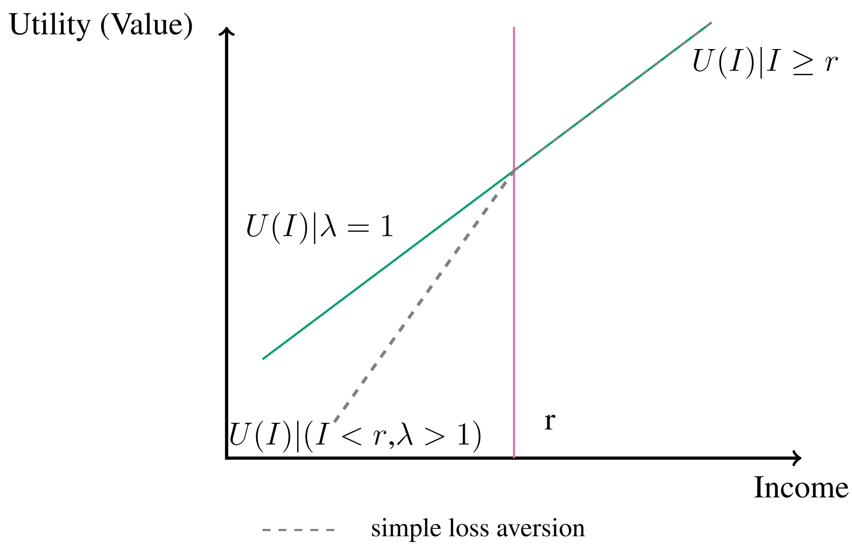
$$\begin{array}{ll} (1+\eta)x'(e) - c'(e) = 0 & \text{if } x(e) - r > 0 \\ (1+\lambda\eta)x'(e) - c'(e) = 0 & \text{if } x(e) - r < 0 \end{array} \tag{6}$$

The shift across the reference threshold reflects the difference in the marginal value of income. At the same level $e$, the marginal benefit on the left side of the reference ($x(e) < r$) is scaled by a factor of $\lambda > 1$ relative to the right side of the reference ($x(e) > r$). This parameter is the coefficient of loss aversion. The parameter $\eta$ is the weight of gain-loss utility in the utility function. This simple model with linear utility implies that, if current earnings, $x(e)$, are below the reference, equilibrium labor supply will be higher than if earnings are above the reference for the same value of $e$. For a loss-averse worker, upon reaching the reference, $r$, there is a downward kink in the marginal value of income, so labor supply will also kink downward, holding constant effort costs at $c(e)$. Figure D12 shows an illustration of this concept. The marginal utility when $\lambda = 1$ is the same on either side of the reference. However, when $\lambda > 1$ and income is below the reference, the marginal utility is higher and overall utility is lower because being below the reference creates a sense of loss. In the standard case in Equation 6, $\lambda = 1$ or $\eta = 0$, and there is no discontinuous change in marginal benefit across the reference.

The prior literature on labor supply has almost exclusively focused on daily references. This focus simplifies the theoretical tests of reference dependence by limiting the role of income effects, which standard theory suggests may be notable in the long-run but will be negligible each day because daily income plays such a small role in long-run or lifetime earnings (O'Donoghue and Sprenger, 2018; Dellavigna, 2009). This justifies the use of linear utility in Equation 5. Reference dependence with loss aversion predicts in my context that when a seller surpasses her daily reference, the probability she stops working for the day will kink upward, holding other factors constant.

On the other hand, the standard model predicts that if the wage return, $x'(e)$, shifted upward for the same value of $e$, the worker would unambiguously work more hours regardless of which side of $r$ she is on. When daily wages are high, the standard worker will increase daily labor supply, and when daily wages are low, the worker will stop working earlier in the day. These labor supply decisions will be a smooth function of $x(e)$ and $c(e)$.

Figure D12: Illustration of Basic Model of Reference Dependence with Loss Aversion




Notes: Illustration of basic loss aversion with linear utility over income. When $\lambda = 1$, the marginal utility above the reference $r$ is the same as marginal utility below the reference.