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

Samuel Dodini
Cornell University
snd46@cornell.edu

November 12, 2020
Click Here for Latest Version

Abstract

The standard and reference-dependent models of labor supply make opposing claims about worker decisionmaking that have important implications. This paper uses novel, high-frequency administrative data from a sales company to test for reference-dependent daily labor supply in a new setting: door-to-door sales. I show evidence that recent expectations, which theory predicts act as important references, are selected by workers based on long-run objectives and form around the lump-sum bonuses paid by the firm at the end of the sales season. I test for reference-dependent labor supply around expectations using a regression kink and discontinuity design. I find that extensive margin labor supply shifts downward substantially at a worker's expectations-based reference, consistent with the reference-dependent model with loss aversion. Long-run goal setting combined with reference dependence supports the theory that expectations-based references must be based on optimally planned choices. It also supports the theory that narrow goal setting and reference dependence together may act as a commitment device. These results have broad implications for how firms motivate their workers and show how long-run contract incentives can drive short-run labor supply choices.

Keywords: reference dependence, loss aversion, non-linear compensation
JEL Codes: D9, J22, J33, M52

*I am grateful to Mike Lovenheim, Maria Fitzpatrick, Evan Riehl, Ted O'Donaghue, Alex Rees-Jones, and Seth Sanders for helpful comments and guidance. 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.

1 Introduction

In recent years, evidence of deviations from the neoclassical labor supply model has led to the emergence of alternative theories of worker decisionmaking and a growing empirical literature. Knowing where worker choices deviate from standard theory is essential for our understanding of human behavior in a variety of important economic interactions.

Kahneman and Tversky (1979) propose a framework that differs from the neoclassical model in important ways. One core piece of the theory is "reference-dependent preferences," which posits that a worker's utility depends not only on her income for consumption but also on where her current income stands relative to some mental reference point or target. Under reference dependence with loss aversion, her marginal benefit of income is higher when she perceives her income to be below her target (in a "loss domain") compared to above her target (in a "gain domain"). When a reference is defined by a worker's expectations (Kőszegi and Rabin, 2006), incomes below expectations are interpreted as a "loss," and incomes above expectations as a "gain." The implication for daily labor supply is that when daily earnings are high, the worker will work *fewer* hours, not more hours as the standard model predicts.[1]

Evidence of reference-dependent preferences in daily labor supply has been mixed across different studies, mainly due to data and modeling differences. Several papers examining the daily labor supply of taxi and rideshare drivers have found either a negative relationship between daily wages and hours worked or downward shifts in labor supply at particular earnings levels (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; Thakral and Tô, Forthcoming). These results are contrary to the standard model. A competing set of studies on taxi drivers finds a positive relationship between daily wages and hours worked, which is consistent with the standard model (Farber, 2005; 2008; 2015; Sheldon, 2016). These studies find little evidence of reference dependence.[2] No clear consensus has yet emerged. None

[1]Standard theory suggests income effects operate over the long run or over the life cycle. Because daily earnings are such a minuscule portion of lifetime earnings, an increase in the daily wage should lead to a non-negative change in labor supply within a day (O'Donoghue and Sprenger, 2018; Dellavigna, 2009).

[2]Other papers that find evidence of reference-dependent labor supply analyze the behavior of bike mes-

of these papers has addressed the question of what factors influence the formation of daily references or expectations in the first place. Notably, prior empirical work has not examined worker behavior outside a limited set of manual occupations.

This paper approaches the question of reference dependence from a new context: door-to-door sales. I use high-frequency data from a company whose contractors sell pest control services purely on commission during a specific sales season. I make two main contributions to this literature. First, I provide evidence that supports the predictions of Kőszegi and Rabin (2006): when a worker plans for what she perceives to be the optimal path forward based on future expectations, the planned choice becomes her reference. Kőszegi and Rabin (2006) call this "personal equilibrium" and posit that the firm may have a substantial influence on a worker's choice of expectations under this condition. This is the first field study of which I am aware to examine this condition and describe how the firm affects the choice of expectations in real-world data.

In my second contribution, treating an individual's daily sales expectations as the reference, I test for reference-dependent labor supply on both the extensive margin (the probability of stopping for the day) and the intensive margin (effort conditional on working), which is new to the literature. In support of the Kőszegi and Rabin (2006) model of expectations-based reference dependence, I provide clear evidence that door-to-door sales workers exhibit reference dependence with loss aversion around expectations in their extensive margin labor supply choices. Because intensive margin responses to the reference are qualitatively small, I show that the choice when to stop working is the margin at which reference dependence is operative. The combination of long-run goal setting and reference-dependent daily labor supply supports the theory in Koch and Nafziger (2016) that loss aversion may act as a commitment device to achieve long-run objectives. I provide the first empirical evidence of this theory in the field.

sengers (Fehr and Goette, 2007; Goette et al., 2004) and fishermen in Hawaii (Nguyen and Leung, 2013). Analyses that find evidence supporting the neoclassical model examine day laborers in Malawi (Goldberg, 2016), stadium vendors (Oettinger, 1999), and fishermen in Florida and India (Stafford, 2015; Giné et al., 2016).

To analyze the formation of expectations-based references and personal equilibrium, I use a panel of each seller's daily performance outcomes and work hours to examine the relationship between a worker's average number of daily service contracts generated, which I call "sales" throughout the paper, and the firm's lump-sum bonuses paid at the end of the season. My hypothesis is that the firm's contract structure incentivizes the worker to optimize around a long-run goal at a bonus threshold and workers distribute that goal into their daily expectations. For example, if a seller sets a goal to sell 300 contracts by the end of a 100-day season, she will set daily goals to sell three service contracts per day, and her cumulative sales will track closely to that trajectory from the beginning of the season to the end. This becomes her daily reference and personal equilibrium. In the neoclassical model, sales across days are fungible, so she will work more hours when the three-sale target is achieved earlier in the day because she knows that high-sales days will need to make up for some low-sales days. Under reference dependence, however, once she hits her three-sale target, the seller will work fewer hours.

I show three pieces of evidence to support this hypothesis. The first two are tests of the existence of a long-run goal that could fit both the standard and reference-dependence model. First, I show that there is substantial "bunching" in the distribution of cumulative sales around bonus thresholds. This bunching emerges early in the season, persists throughout, and is centered on the average daily sales necessary to achieve the relevant bonus. This is a signal that workers have similar long-run objectives that inform similar daily choices. Second, I show that a seller's average daily sales performance in the first five weeks of the sales season explains almost 90% of the variation in total sales at the end of the season. This is an indication that the daily goals chosen by the sellers early on are borne out in their subsequent cumulative outcomes. Third, as an initial test of the standard model against a reference-dependent model, I show that the relationship between sales and hours worked within sellers is more strongly positive on days that were below average than for days that were above average. This casts doubt on the standard model prediction that workers will increase their labor supply on high-earnings days.

In my second contribution, I explicitly test for reference dependence around daily expectations. I use a detailed panel of observations in half-hour increments with each seller's location, cumulative sales, pitches presented to a prospective customer, and the probability of stopping work for the day. This measure of intensive margin effort, pitches per half hour, is new to the literature. I use the discrete number of sales a seller generates as my focal measure of performance and as a proxy for a worker's income. A kink or discontinuity in labor supply at the reference point indicates the presence of loss aversion due to a shift in marginal utility or a fixed utility cost of not reaching the reference, respectively.

To test for these changes in labor supply, I use these detailed observations in a regression discontinuity and kink design using each individual's own expectations as the cutoff. I define the reference as a sample proxy of "recent expectations" (Kőszegi and Rabin, 2006): 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 estimate the effects of approaching and crossing the reference on both the probability of stopping work and effort conditional on continuing to work. My estimation model addresses a shortcoming in the prior literature by controlling for each worker's baseline hazard through rich interacted fixed effects. This accounts flexibly for unobserved heterogeneity correlated with differences in typical stopping behavior or effort, which has been a source of bias in past work.[3]

I find significant evidence of reference dependence with loss aversion in stopping behavior. Upon reaching their expectations-based reference, the probability a worker stops for the day increases significantly. Moving from the reference to one sale above the reference raises the probability of stopping by 7.3–8% while moving from the reference to one sale below the reference reduces the probability of stopping by only 1.9–2.7%. On the intensive margin, when moving one sale above the reference, pitches presented the next half hour decline by approximately 1.5–1.9% compared to a 0.5–1% change from moving down to one sale

[3] Thakral and Tô (Forthcoming) include a driver-specific hazard in their estimates of taxi driver behavior and compare it to the estimates generated by Farber (2015). After adjusting for this baseline hazard, they conclude that the evidence for reference dependence is stronger than that found in Farber (2015) and the inclusion of a baseline hazard is essential for unbiased estimates.

below the reference. The results at the intensive margin are smaller in magnitude and more sensitive to specification choice than in my stopping models. My results indicate that the choice of when to stop working is the key margin at which reference-dependent daily labor supply operates.

Using my estimates, I calculate coefficients of loss aversion in my sample of approximately 2.8–4.1 and 1.5–3.9 for the probability of stopping and the intensive margin, respectively. These coefficients indicate that the marginal utility of income before reaching the reference is 2.8–4.1 times higher than after reaching the reference, or that losses loom larger than gains by a factor of approximately three to four. Because daily income in my sales context can be very high, the stakes of each day's work are high as well, making daily labor supply decisions more financially consequential than in the past literature.[4]

This paper contributes to the literature on reference-dependent labor supply in multiple ways. The novel context and data used in my analysis alone provide two contributions. First, the literature thus far has been limited to examining occupations such as taxi drivers that are manual and routine in nature, in part, due to data availability. There has been no work on occupations that require advanced cognitive or social skills. The distinction between these occupations is important if the types of workers who select into these manual occupations differ in meaningful attributes from those who select into primarily social or cognitive occupations.[5] This is particularly important if reference dependence and loss aversion vary with those attributes. My analysis provides new evidence of reference dependence in a more cognitive/social occupation.

Second, the prior literature has generally considered occupations in which income is a smooth function of hours worked, and deviations from average income are relatively minor. For example, a standard deviation in wages for a taxi driver is only about 10% of the mean

[4]As an example, encountering one extra resident willing to purchase pest control services leads to an increase in income of \$100–\$250. The 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 as an entire shift for a taxi driver but takes roughly the same amount of time as one or two taxi trips (16–32 minutes) (Thakral and Tô, Forthcoming).

[5]These might conceivably include discount rates, preferences for consumption and leisure, cognitive capacity, myopia, etc.

(Thakral and Tô, Forthcoming). In my setting, 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 income references than in the past literature.[6]

My analysis of long-run objectives and short-run expectations is the first field study of which I am aware to analyze the conditions around which sellers choose expectations in "personal equilibrium," and I show that the firm plays a role in that choice. Importantly, the empirical literature has not considered the interaction between long-run objectives and daily targets in reference dependence, in part because of a nebulous definition of the "long run" in other contexts.

The neoclassical model views reference dependence as irrational because workers are not taking advantage of high-wage days by working more hours. However, if the personal equilibrium condition holds and workers are setting goals for their long-run outcomes that are optimal given their future expectations, Koch and Nafziger (2016) theorize that evaluating goal progress at the daily or weekly level can act as a commitment device for achieving long-run objectives, particularly if the worker is present-biased. This operates by intentionally inducing loss aversion until the daily objective is met. My analysis provides evidence of this phenomenon by showing how daily expectations serve long-run objectives.

Finally, I contribute to the literature that examines the use of non-linear incentives and bonuses in the workplace and their effects. Recent work presents evidence that bonus payments for reaching a performance threshold increases worker effort and productivity (Freeman et al., 2019; Graff-Zivin et al., 2019), particularly when framed as a loss (Fryer Jr et al., 2012; Levitt et al., 2016). I provide evidence that one mechanism through which these may increase productivity is by influencing the formation of goals and expectations in the long run and short run.

The question of reference-dependent labor supply is central to our understanding of

[6]For example, it is much easier to count "1, 2, 3 contracts sold" than to calculate total income earned net of tips over a shift when driving a taxi.

the power of incentives to induce workers to provide optimal effort. 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. This means that simply increasing wages is a blunt and imperfect instrument for motivating worker effort, and the efficacy of a wage increase depends on the worker's reference. Reference dependence also has major implications for management practice. If firms can use non-linear payments or other tools to exert determinative power over what goals the worker sets for herself, then the firm can anchor the worker's reference-dependent labor supply at a higher level in pursuit of greater profits and higher worker income.

My results have important implications for public policy. For example, if a transfer benefit is contingent upon earning some minimum income or if benefits are maximized at a kink point, the benefit threshold or kink may become a reference akin to the bonus payments in my sales setting.[7] In choosing these various thresholds and kink points, my results suggest policymakers may be affecting the personal equilibrium condition for workers and therefore their daily labor supply choices by influencing the formation of their daily expectations. This suggests that policymakers have substantial power over labor markets and daily labor supply choices through their choice of tax and transfer thresholds.

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.[8] 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 incentives are also extremely common in a variety of sales occupations across industries.

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

[7] Examples include benefit cliffs in the Affordable Care Act (Kucko et al., 2018), and kinks in the Earned Income Tax Credit and Child Tax Credit schedules (Mortenson and Whitten, 2020).

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

to sell their products and services.[9] These include solar panels, pest control services, knives, and home automation and security systems. Industry practice is relatively homogeneous across these products, which managers attribute simply to historical inertia.[10] Prospective sellers meet with managers and current sellers, 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 what city they will call home for the summer. The work itself is unpleasantly hot in the summer and sometimes 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. 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. This leads to a large variance in income across sellers. 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 details of the contract are recorded electronically, and the first service is scheduled usually within one to three days. 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 modest 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.

[9] 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: the ability to approach strangers and strike up a conversation, the ability to connect quickly with others, and the ability to command conversations to move them toward a specific goal. Recruiters understand this dynamic and seek to capitalize on this recent skill development before it depreciates.

[10] In conversations with managers, when asked why they structure their compensation the way they do, a common answer is, "That's industry standard." Many companies who employ these workers were started by those who were former sellers themselves, so they rely on their past experience in setting their own company policies.

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. 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 are calculated at the end of the season based on final performance and paid out thereafter.

Figure 1 characterizes the total income a seller earns at the end of the summer season depending on their total sales at an assumed average service 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. In addition to this de facto bonus from the commission change, the seller receives a flat "rent bonus." This bonus is about \$2,000 and covers the seller's apartment rental costs for the summer. At 250 sales, sellers qualify for the company vacation: an all-expenses-paid trip that includes airfare, hotel accommodations, food, and excursions. 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. The increasing commissions can rightly be interpreted as price discrimination that induces individuals to reveal their "type" as in the non-linear pricing literature (Wilson, 1993). If a worker expects to end in a particular 50-sale interval, the operative incentive is a linear piece rate with a bonus.

The skills required for success on the job are broad. Sellers must be able to strike up a conversation with a stranger, understand and respond to objections that a potential customer may share, communicate the value of the product in improving the life of the customer, 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. These requirements make this setting unique in the literature.

Work areas are assigned by local team managers tasked with overseeing operations within a portion of a metropolitan area. Prior to the start of the season, company general managers decide on a set of neighborhoods as candidates for sellers to canvas. After the start of the season, sellers are assigned to a work neighborhood within their portion of the metro area by their team manager. Sellers continue to knock on doors in their assigned area until approximately 75% of doors in the neighborhood are marked in their tracking software, after which the team leader assigns a new area. This process, while not strictly random, does not appear to differentially sort sellers based on skill into "easier" or "harder" neighborhoods (see Section 3 and Appendix Table A1 for a discussion and evidence).[11] See Appendix C for more details on industry practice and contracts.

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 and "experiences" like a cruise, resort stay, or annual ski passes, though sellers have the option to cash out the value of the prize.[12]

[11]In some cases, it may be beneficial to send more talented sellers into more "difficult" areas because they could likely generate sales that less talented sellers might not be able to generate. Any sorting by talent and expected difficulty by neighborhood would be costly for managers with unclear returns. Managers insist to all sellers that "work area does not matter," and the evidence supports that argument.

[12]While these tournaments are of independent interest, they are not the focus of my analysis here. I do,

Finally, there is an important information innovation to note regarding these sellers. Through the company's centralized website and mobile application, sellers can view their own performance as well as that of all others in the company. 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 to the seller. 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.

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 for 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." Observations in this panel are contingent upon showing up to work, and I omit days in which the seller records no knocks or sales. While this eliminates one aspect of the labor supply choice, almost all sellers record work activity each day. Their contractor agreements also stipulate penalties for not knocking a minimum number of days.

Similar to the past literature (Crawford and Meng, 2011), I calculate a proxy for each

however, analyze behaviors during these tournaments in a separate study.

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 as sellers further develop their skills or update their personal goals with new information, though the measure is stable for most sellers.[13]

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, Forthcoming). For each seller in each half hour, I create a measure of their current distance to their daily reference: their number of cumulative sales so far that day minus their average sales for that day of the week. For values of this measure 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 across 180 days in 2018-2019 covering the late April to mid August selling 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 daily total precipitation, high temperature, and low temperature for the weather station nearest to the ZIP code in which each seller is working as controls. These factors are important because door-to-door sales is an almost exclusively outdoor job. During these summer months, heat is an especially important factor to consider, as heat has demonstrable negative effects on cognitive ability and learning (Park et al., 2020). Heavy rain and intense heat also

[13] Using various definitions of recent expectations such as sales in the prior five weeks yields similar results. Sales during the first five weeks of the season are highly predictive of the rest of the season, explaining almost 90% of the variation in total sales at the end of the season.

greatly increase the marginal cost of effort.

One obvious concern in this context is that sellers might sort non-randomly into neighborhoods that are more or less receptive to someone selling pest control services. For this reason, 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. These variables are median household income, the unemployment rate, the poverty rate, the share of home values in specific ranges, the share of households with a married couple, the share of adults with a Bachelor's degree or more, total housing units, the share of units that are owner-occupied, the share of units that are single-family homes, and the share of the population that has not moved in the past year.[14]

Summary statistics for my two panels are in Table 1. 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 2 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 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

[14]These neighborhood characteristics explain less than three percent of the variation in sales both between and within sellers. Appendix Table A1 provides details of a regression of daily sales on weather and ZIP code characteristics. Only three coefficients are statistically significant at the 10% level. There is little evidence of sorting behaviors correlated with seller performance.

their sixth hour of work.

In the data, there are a few important trends that put upward pressure on labor supply as success increases during the day or as the evening progresses, working against the downward shift in labor supply predicted by Prospect Theory. One such trend is positive autocorrelation in sales. 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. I present this autocorrelation in Panel A of Figure 3. The results suggest that there is high autocorrelation in residualized sales for just under one hour, or that success now is predictive of success in the next half hour or so. If sellers understand this, the standard model predicts they will increase their labor supply rather than reducing it when performance crosses their reference.

Panel B of Figure 3 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. If sellers find success that puts them in their gain domain earlier in the day, the prospect of more productive work and more sales later in the evening would lend itself to increasing labor supply rather than reducing it because the cost to generate sales later in the day is lower than earlier in the day. Quitting early because of early success would lead to foregoing the most productive hours of the day.

This context and the availability of comprehensive data provide a unique opportunity to test for the existence of reference-dependent labor supply and to study how reference dependence is affected by the firm. To adequately connect my context to the theory, I frame the predictions of the standard labor supply model and Prospect Theory within this sales context.

4 Theory and Predictions

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 reference) relative to what they supply in a "gain domain" (after achieving a reference) because they are averse to feeling a sense of loss.

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:

$$U(e) \equiv x(e) + \mu\left(x(e) - r\right) - c(e) \tag{1}$$

where

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

The μ function captures what is called "gain-loss utility." The equilibrium labor supply for this utility function with gain/loss utility is given by:

$$\begin{aligned} (1+\eta)x'(e) - c'(e) = 0 \quad & \text{if } x(e) - r > 0 \\ (1+\lambda\eta)x'(e) - c'(e) = 0 \quad & \text{if } x(e) - r < 0 \end{aligned} \tag{2}$$

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 η 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 4 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 neoclassical case in Equation 2, $\lambda = 1$ 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 1. 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 neoclassical 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 $c(e)$, and they will be opposite those in the reference-dependence model. In short, there is nothing in the standard labor supply model that would predict a discontinuous change in labor supply around a particular point in the distribution. My empirical approach, which I explain in Section 7, tests for discontinuities and kinks in the distribution of daily sales that would indicate the presence of reference dependence.[15]

4.2 Goals and Daily References

As important as the parameter of loss aversion (λ) is to the model of reference dependence, equally important is the definition of r. In an essential theoretical paper, Kőszegi and Rabin (2006) (the KR model) theorize that "recent expectations" act as important references. But how do expectations for daily outcomes form? 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 choice is made, the planned path becomes her reference. This state is a 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 neoclassical prediction. 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 gain new information about their ability, they can quickly adjust their future goals to a new bonus and then adjust their daily reference. 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 only exhibit gain-loss utility over outcomes that deviate from expectations. After setting her plan for the path ahead, each seller responds each day to whether or not her performance is below or above what she expects for the day. These expectations may also be specific to each day of the week. For example, door-to-door sales workers can generate more sales per hour on Saturdays because more residents are home, which makes for plentiful contact opportunities. The KR model predicts that workers

[15] The presence of a discontinuity or kink is not sufficient to conclude that neoclassical behavior is completely absent from the worker's labor supply decision. I show in the next section that an anticipated wage change via a bonus leads to behaviors consistent with the neoclassical model and with the Kőszegi and Rabin (2006) model.

will form expectations specific to a Saturday. Then, given their expectations about their own typical Saturday earnings, they will experience gain-loss utility if their performance on a Saturday is above or below expectations.[16]

Building on the concept of personal equilibrium, Koch and Nafziger (2016), 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. Recent lab experiments support this interpretation (Koch and Nafziger, 2020). 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 (Gilpatric, 2008).[17] To combat present bias, as a commitment strategy, a worker will separate (or "bracket") her broad or long-run objectives into narrow pieces with the *intention* of inducing loss aversion if she is below her reference. Not achieving performance expectations in narrow tasks or periods induces a sense of loss, and to avoid a negative comparison, the worker will increase her effort toward expectations, leading to behaviors consistent with reference dependence. No other study of which I am aware investigates this hypothesis in the field.

An example from my sales setting 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 have two effects. First, she may raise her objective for total sales at the end of the season from 190 to 200 because she believes it is attainable and the $2,000 bonus is worth the extra effort. Next, 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. She then works each day with these two sales per day in mind, which satisfies the personal equilibrium condition. If she is present biased, she may shirk today in favor of putting in more effort tomorrow because performance is fungible across days. However,

[16] In my data, expectations for each day of the week are highly correlated within seller across days (approximately 0.90). If there were a larger difference in expectations across days, I could use each day's expectations as a falsification test for performance on other days of the week.

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

if she views each day in a separate mental account, unexpectedly being below her two sales creates a sense of loss, so she will work harder or work extra hours into the evening 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 is opposite the standard model, which predicts that because income is fungible across days, she will work more hours on days with above-average earnings and will work fewer hours on days with below-average earnings. Around the same average of two sales, her effort and total income will fluctuate more from day to day as she expects low-income and high-income days to even out over time.

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 and pre-determined end date that is known to all workers. A second reason the prior literature has been unable to examine long-run goals is that the occupations under study provide no exogenous or experimental manipulation of goal choices. PestCo, through its use of lump-sum bonuses, provides exogenous incentives for workers to sort themselves around specific outcomes. These bonuses increase the salience of particular points in the distribution to act as goals or targets.

An important prerequisite for a worker to be able to have a personal equilibrium is that she must have a realistic forward-looking view of what she can plausibly accomplish. 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 indicate 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 evidence of forward-looking expectations in Section 5.

Let us suppose that a worker has a feasible long-run objective in mind at the beginning of the sales season. What empirical evidence would confirm the existence of such goal setting, and what do the standard and reference-dependence models predict about their behaviors? Holmstrom and Milgrom (1987) suggest in their model of optimal contracts that when workers choose an outcome for the end of a contract period they will control a Brownian process with a constant drift rate. In other words, workers draw a mental line between their starting point and their choice of some cumulative outcome at the end of the contract. In order to attain their chosen endpoint, a worker under the standard model will work more on high-wage days and less on low-wage days with performance centered around that mental line. Because income is assumed to be fungible across days in this model, high-wage days will make up for low-wage days on average. This kind of "broad bracketing" of goals across days provides insurance against feeling a sense of loss for any one day, but it weakens the incentives to achieve a daily goal (Koch and Nafziger, 2016).

Figure 5 provides a visual illustration of the neoclassical model using simulated data relevant to my sales setting. In the simulation, I show the growth in sales corresponding to goals at the end of a 90-day season of 200 and 250 sales. Panel A shows the evolution of the various individual paths over the season. The thick lines correspond to the average daily sales necessary to achieve each goal, while the other lines of the same color represent the experience of 100 individual workers as they work through the season. I allow for deviations each day from the average with mean zero and standard deviation of 2.0, similar to the structure of the data in Table 1. Deviations from the mean each day in the simulation are the result of working more hours on high-wage days and fewer hours on low-wage days. Some workers will have high-earning days followed by low-earning days, while others will experience runs of multiple high- and low-earnings days. The individual worker time paths incorporate these highs and lows. Panels B and C show the histogram of cumulative sales at day 50 as well as at the end of the season corresponding to the different individual outcomes in Panel A.

The difference in the distributions in Panel B and Panel C demonstrates that the two distributions are moving further apart as time passes. On day 50, there is some separation of the two distributions, but at the end of the season, there is a notable separation between the two. This creates “bunching” or “heaping” in the distribution that becomes more defined until the end of the season. For example, relatively few sellers end the season near 225 sales, while many end at 200 or 250. The variance of total sales within groups with the same goal increases because some workers will have numerous high-wage/high-hours days that will bring their total sales far above the mean, while others will have numerous low-wage/low-hours days that will bring their total far below the mean.

This simulation makes clear three predictions about the behaviors of workers with a common goal in the standard model. First, the variance in the distribution of sales around each distinct goal’s average path will grow with time. This is because some workers will consistently have more low-wage days than high-wage days and vice versa. This behavior pulls their realized total sales further from the mean. Second, there will be volatility in the distributions over time because peaks in daily outcomes are high and troughs are low. And third, the relationship between hours and income will be more strongly positive on days that were above average than for days that were below average. This is because workers are taking advantage of days with above-average wages by working more hours.

On the contrary, Koch and Nafziger (2016) and Kőszegi and Rabin (2006) predict bunching behavior in the distribution of sales but for a different underlying reason with different predictions. The key assumption in the standard model is that performance across days is fungible. However, rather than working more on high-wage days and less on low-wage days and expecting the two will average out over time, a reference-dependent worker evaluates her performance every day against her daily goals. Outcomes below expectations create a feeling of loss, so she will work more hours in order to achieve her daily expectations. Unexpectedly high performance leads to lower motivation because the marginal value of additional income has fallen in the gain domain. Hence, she will stop working earlier on days that are better than expected. Bunching or heaping emerges not because similarly skilled workers make similar average effort choices and have a similar long-run target, but because workers have similar *daily* outcome targets.

The empirical prediction is that, contrary to the standard model, there should be little volatility in the distribution of sales within groups with a similar goal over time. This is because peaks and troughs in daily outcomes are muted by reference-dependent labor supply when there is upward pressure on hours from the loss domain and downward pressure in the gain domain toward expectations. Sellers will work in narrow bands around the average daily sales necessary to achieve their goal. The variance of the distribution around expectations will remain relatively constant for the same reason. Contrary to the standard model, the relationship between earnings and hours worked will be more strongly positive on below-average days than on above-average days due to loss aversion. Finally, in both the reference-dependence model and the standard model, if workers have a long-run goal in mind, early performance in the season should have strong predictive power regarding the final total cumulative sales of the worker because they follow through on the goals and expectations they set early on.

5 Empirical Strategy: Goal Setting and Personal Equilibrium

I first use my daily panel to examine goal setting by sellers. To test for forward-looking behavior, a prerequisite for setting long-run goals and the personal equilibrium condition in Kőszegi and Rabin (2006), I use my daily panel to 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 significantly change their daily labor supply because they have already optimized over their chosen long-run outcome.

To test this hypothesis, I use my daily panel to regress hours worked per day on indicators for 25-sale intervals of current cumulative sales interacted with indicators for 100-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:

$$\begin{aligned} y_{idwa} = & \beta_0 + \sum_{e=[0,25)}^{[875,900)} \sum_{f=[0,100)}^{[800,900)} \beta_{ef} \boldsymbol{I}_e * \boldsymbol{I}_f \\ & + Efficiency_{idwa} + \alpha X_{zdwa} + \sigma W_{zdwa} + \mu_i + \nu_d + \omega_w + \tau_a + \varepsilon_{idwa} \end{aligned} \tag{3}$$

The outcome variable y is the number of hours worked per day as a measure of labor supply. 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, β_{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.[18] The X vector controls for the characteristics of the ZIP code in which the seller is working discussed in Section 3. These include median household income, the unemployment rate, the share of housing units that are owner-occupied, the share of housing units that are single-family homes, etc. The W vector controls for weather, including the high and low temperature and precipitation each day at the weather station closest to 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 (μ_i), day of the week (ν_d), week of the season (ω_w), and year (τ_a). These fixed effects ensure that the β coefficients characterize within-seller choices holding constant other characteristics of the sales season, fatigue, or learning effects.

Next, to visualize how workers respond to bonus incentives, I present kernel density estimates of total cumulative sales at the end of the season as well as throughout the season for all workers. I present densities at two-week intervals to draw the evolution of sales over

[18] Without these interactions for total sales f, differences in the composition of workers over sales intervals may confound the coefficients of interest in an example of Simpson's paradox. See Appendix Figure A2 for the result of this exercise.

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. I examine two groups: those whose total sales at the end of the season were between 175 sales and 225 sales, putting them around the bonus threshold at 200 sales, and those with total sales between 125 and 175, putting them around the 150 sale bonus threshold. If workers with the same apparent goal at the end of the season preserve the same shape of the performance distribution throughout the season with little volatility, this suggests that workers set their eyes on a long-run expectation and work each day with those goals in mind with reference-dependent labor supply.

Finally, I perform two regression exercises to test for indications of goal setting and 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, which provides evidence of goal setting early in the season. The second regression is 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. Following the past literature (Crawford and Meng, 2011), I define expectations 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 estimate:

$$
\begin{aligned}
y_{idwa} = &\beta_0 + \beta_1 Sales_{idwa} * \boldsymbol{I}^{+}_{Expectations} + \beta_2 Sales_{idwa} * \boldsymbol{I}^{-}_{Expectations} \\
&+ \alpha X_{zdwa} + \sigma W_{zdwa} + \mu_i + \nu_d + \omega_w + \tau_a + \varepsilon_{idwa}
\end{aligned} \tag{4}
$$

The model includes all the same fixed effects and controls as Equation 3. 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 wages are high, β_1 will be more strongly positive than β_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).[19]

The combination of all these pieces of information supports the hypothesis that these door-to-door sales workers are optimizing their long-run expectations around the lump-sum bonuses paid by the company, which then determines their personal equilibrium. Strong evidence of reference-dependent labor supply around the expectations that arise from this choice of long-run goal would confirm the elements of the personal equilibrium condition as well as the predictions of the KR model.

6 Results: Bonuses, Goals, and Forward-Looking Labor Supply

First, I show that sellers have realistic forward-looking expectations and are not myopic. Estimates from Equation 3 are summarized in Figure 6, which shows the predicted hours worked per day over 25-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. After what appears to be an initial adjustment period, those whose sales totaled over 300 quickly began working 7.5-8 hours per day, while those with fewer than 200 sales worked less than 7 hours per day consistently over the season. 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 money they have currently banked, suggesting a singular focus on long-run performance expectations.

Next, I present evidence of bunching behavior in the distribution of sales throughout the sales season. Figure 7 shows the results of a kernel density estimate for bandwidths of 25 and 7 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 free rent and

[19] This should be especially true in this sales setting where sales late in the day are less costly to achieve.

the company vacation. This indicates that the bonuses are salient for the sellers and that they are forward-looking. Figure 8 breaks down the density of total cumulative sales for each seller in two-week increments over the season.[20] 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. It is clear that groups of workers have very similar trajectories in the growth of their cumulative sales, and these trajectories eventually bring them to specific bonus thresholds by the end of the season.

An even starker pattern emerges when examining groups of workers with similar total performance at the end of the season. Figure 9 presents the kernel density estimates of cumulative sales over the same two-week intervals as Figure 8, 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. Recall from the simulation in Figure 5 that the standard model predicts growing dispersion in the distribution as time passes in the season for those with similar long-run objectives as well as volatility in the density if workers substitute sales performance across days. Contrary to that prediction, the dispersion of performance among those who finished around the neighborhood of 200 sales remains constant or even narrows over time. These figures confirm that sellers are particularly cognizant of and responsive to these lump-sum bonuses. A similar pattern is visible for those who finish the season around the 150-sale bonus threshold, as seen in Figure 10. The densities are remarkably stable, and the variance is stable or narrows over time. 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.

The results of my regression exercises using my daily panel are in Table 2. 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

[20] 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 weeks progress, so the relatively high density below 100 includes those who only worked a portion of the season.

explains over three-quarters of the variation in total cumulative sales at the end of the season. Expanding this 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, which predicts that workers will work more hours on days that have high wage returns, even when the returns are unexpectedly high. The results of this regression exercise suggest that sellers are putting in more hours in response to sales when sales are below expectations. This provides initial evidence of reference dependence.

Taken together, the results of each of these exercises 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. That all four of these conditions hold empirically is consistent with the conditions of the KR model's personal equilibrium definitions. If combined with strong evidence of reference dependence within days, these factors also support the intuition behind the use of narrow goals in pursuit of long-run objectives in Koch and Nafziger (2016).

7 Empirical Strategy: Reference-Dependent Labor Supply

I next use my half-hourly panel to rigorously test for the presence of reference dependence in daily labor supply decisions. I focus on the probability of 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 conditional on continuing to work. This is new to my setting.

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 (and therefore income earned). 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 time is as good as random. Given the full set of controls and fixed effects I present, it is not obvious what other factors might confound the results I present.

I first estimate a non-parametric model of labor supply with respect to each seller's distance from their sales target to trace out any differences around their reference without imposing a functional form. I note here that PestCo runs a number of 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 reference, I separately analyze behavior during non-tournament days and present those results in my tables and figures.

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:

$$\begin{aligned} y_{ithdwa} =& \beta_0 + \sum_{e=-k, e\neq 0}^{k} \beta_e * \boldsymbol{I}_e\{sales_{ithdwa} - \overline{Sales}_{idwa} = e\} \\ & + \alpha X_z + \sigma W_{dwa} + \mu_{it} + \eta_h + \nu_d + \omega_w + \tau_a + \varepsilon_{ithdwa} \end{aligned} \tag{5}$$

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, β_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 include values rounded to the nearest integer, with the (0,1.5) interval being included in β_1.[21] Under reference dependence, beginning with β_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 (μ_{it}, η_h, ν_d, ω_w, τ_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, μ_{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 (Forthcoming). 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 also used in Equation 3.

In my main models of interest, I fit parametric estimates that impose a functional form to match the non-parametric estimates in Equation 5. This equation is a regression kink and discontinuity design with linear splines on each side of the reference.

$$\begin{aligned} y_{ithdwa} = &\beta_0 + \beta_1\{sales_{ithdwa} - \overline{Sales}_{idwa}\} \\ &+ \beta_2\{sales_{ithdwa} - \overline{Sales}_{idwa}\} * \boldsymbol{I}_{sales \geq \overline{Sales}} \\ &+ \beta_3 * \boldsymbol{I}_{sales \geq \overline{Sales}} \\ &+ \alpha X_z + \sigma W_{dwa} + \mu_{it} + \eta_h + \nu_d + \omega_w + \tau_a + \varepsilon_{ithdwa} \end{aligned} \tag{6}$$

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 β_1 defines the slope of the relationship between one's current

[21] 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, Forthcoming). 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 A2.

distance to average sales and labor supply in the loss domain. The β_2 coefficient captures the change in slope upon crossing the reference and entering the gain domain. Finally, β_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 5. In a neoclassical 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, β_2 will be significantly positive in the stopping model. The coefficient β_3, while not predicted by simple loss aversion, represents a discrete penalty for "losing," or for falling short of expectations, which suggests reference dependence.[22] If β_2 and/or β_3 are significant and positive in the stopping model, this represents strong evidence of reference-dependent labor supply.[23]

8 Results: Daily References and Daily Labor Supply

I now present the results from my half-hourly panel estimates of labor supply responses to daily references. Figure 11 shows each of the coefficients from the non-parametric estimates from Equation 5 as well as the linear estimates from Equation 6. 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. For context, the average probability of stopping right at the reference is 0.079, so an increase in this probability of 0.0058 for each sale past the reference represents a 7.3% increase. These are substantial changes. Were labor supply in the gain domain to match labor supply in the loss domain, sellers would continue to be active longer into the day and would have greater

[22] I also estimate these models allowing for curvature on either side of the reference. The second-order polynomials are small and not statistically significant on either side for either outcome.

[23] As a robustness check, also I estimate my parametric models using a pooled sample with across all tournament and non-tournament periods and interact my coefficients of interest with indicators for tournament periods. These estimates are in Appendix Table A3. I report the non-tournament coefficients in my figures (see Section 8.1.

opportunities to take advantage of peak earning hours later in the evening.

Panel B of Figure 11 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 intensive margin effort and sales each day. There is a minimal change in intensive margin effort at the reference. The size of the decline is small in percentage terms: each sale reduces intensive margin effort by approximately 1% from a baseline mean of 2.38 pitches at the reference. These results suggest that reference dependence in intensive margin 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 intensive effort is similar.[24]

Next, I use my estimates to calculate the parameter of loss aversion, λ. While there are various methods that attempt to measure this parameter, my setting requires an approach 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: $\lambda = U'(0)_{\uparrow}/U'(0)_{\downarrow}$. Because my empirical model partials out all covariates correlated with effort costs, 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 intensive margin have a ratio of 1.5. My estimate of 2.8 in my baseline models is the most conservative of my stopping model estimates. Other specifications, which I detail in Section 8.1 yield estimates as high as 4.1 or 5 for the stopping model and 3.9 at the intensive margin. In their review, Gächter et al. (2007) find loss aversion coefficients of approximately of 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 consistent with the prior literature. There is some evidence in Engström

[24] In two of my specifications, there is a downward kink and a discontinuity in intensive margin effort. However, even with a downward shift in effort in the gain domain, the effects are small in percentage terms: less than 2% per sale above the reference.

et al. (2015) that younger people and lower-income people are more loss-averse than their older or higher-income counterparts. This is relevant given the youth of my sample. In addition, the high stakes of my setting as well as the connection between daily outcomes and long-run objectives may explain a higher degree of loss aversion in this context.

8.1 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 whether or not there is a tournament that day (or what kind of tournament). This allows the effect of crossing the reference to differ based on what type of tournament/non-tournament period the seller is working in. The results of this specification are in Figure 12. 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 intensive 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. At the intensive margin, the results are still qualitatively small at a change of less than 2% per sale above the reference.

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. Conceivably, there may be other cutoffs and other parameters that better fit the data. As a robustness check against incorrect specifications of the cutoff, 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 a chosen cutoff at zero, I allow the cutoff itself to be a parameter of the model. The model minimizes the sum of squared residuals at different slope and cutoff parameters. The results from estimating this model are in Table 3. 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. There is also a statistically significant discontinuity of 0.0054 at the cutoff. 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, 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. The ratio of the slopes across the reference is 3.3.

As an additional test, I estimate my baseline model but include intensive effort on the right-hand side. As a control, I include cumulative pitches that day as a measure of total exerted effort. If a worker is exerting a high level of effort in 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 4 presents these estimates for my parametric models. 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. The ratio of slopes in this specification is 2.8, the same as my baseline model.

At the intensive 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. The ratio of slopes is 0.997. The slope in the loss domain, -0.0175, indicates a very small change in pitches as sales increase (a 0.7% per sale decline).

Finally, I create an alternative measure of each worker's daily reference and 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 presented earlier. 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 A3, I show that the ratio of slopes across the reference is 2.7, consistent with my baseline results. There is also a substantial discontinuity in this specification, which is also consistent with reference dependence.

8.2 Discussion

Setting narrow, daily goals in the service of broader, long-run goals as a method of accountability is a direct implication of recent theories on goal setting (Koch and Nafziger, 2016; Hsiaw, 2018; Koch and Nafziger, 2020). Goals instill a sense of loss for not meeting a daily target. Narrow bracketing of goals therefore frequently induces effort above that of similarly sized gains by keeping workers in a loss domain at the start of each day. This is particularly important if workers are present-biased. That individuals are present-biased in a number of contexts is supported in the empirical literature (DellaVigna and Malmendier, 2006; Ashraf et al., 2006; Meier and Sprenger, 2010). Given the unpleasantness in door-to-door sales, self-control problems resulting from present bias may be nearly universal in the occupation.

These results suggest that not only are sellers setting specific long-run targets around bonus thresholds, but they persist toward these targets as the sales season progresses, leading to bunching throughout the season. Given this response to long-run incentives, if the principal were to shift those bonus thresholds upward, how would workers respond? Freeman et al. (2019) report that when a Chinese insurance sales company shifted their bonus threshold up and increased the size of the bonus, the result was significant increases in sales performance, company profitability, and stock price. The entire distribution of sales shifted from bunching around the initial bonus to bunching around the new bonus. My results suggest that one mechanism through which this change occurred is that the new bonus schedule shifted worker expectations upward, both in the long run and in daily performance. The implication of that paper for my setting is that PestCo could conceivably generate more effort by setting individual daily expectations higher by shifting out their own bonuses—essentially influencing sellers to set their personal equilibrium to a higher level. While the Freeman et al. (2019) paper does not report daily productivity for workers, my results predict that were hourly labor supply choices observed, workers in that insurance firm might exhibit reference dependence in the daily sales necessary to achieve the new bonus threshold.

I argue my results have strong external validity and are important and applicable to other contexts for several reasons. First, sales as an industry is a large and expanding market in the US, 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. Second, other industries and labor markets make use of these types of incentives as well. 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 school sports or performing arts programs. That these incentives are widely used across occupations and contexts indicates that a broad set of actors acknowledge the power of these incentives for motivating people. Finally, the types of people who work in sales at some point in their lives are not particularly unique. They go on to perform work across all types of occupations, and the skills used in sales are transferable to many occupations. In other words, there is no reason to think that sales workers are somehow not comparable to workers in other occupations that use cognitive or social skills.

9 Conclusion

Using novel, comprehensive data from a door-to-door sales company, this paper provides strong evidence of reference-dependent preferences in daily labor supply in a new setting. The sales job is unique in the literature because it entails socially adaptive and cognitively challenging work. In contrast, the prior literature has focused exclusively on manual and routine jobs. This difference in job skills is important if selection into the manual jobs in the past literature is correlated with reference dependence or loss aversion. Door-to-door sales workers exhibit behaviors around expectations-based references consistent with loss aversion in their labor supply choices. I confirm this behavior when examining the choice of when to quit working for the day as well as the choice of how much effort to expend conditional on continuing to work. I find that the extensive margin choice (when to stop working) is the margin at which reference dependence is operative. Effort conditional on continuing to work is less responsive to surpassing expectations. This reference dependence puts upward pressure on labor supply while in the loss domain and downward pressure on labor supply in the gain domain.

I show that firms, through 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). These sellers plan for outcomes they perceive to be optimal given their expectations for the future, and their daily expectations become their references. Sellers' average daily sales in the first five weeks of the season explain 87% of the variation in total sales at the end of the season, meaning workers are targeting early on the number of sales each day that reflects where they want to be by the end of the contract. Furthermore, there is significant bunching in the distribution of performance around bonuses at the end of the season. This bunching behavior emerges near the beginning of the season almost immediately, and workers perform in narrow ranges around the average daily performance they need to achieve their long-run targets at bonus thresholds.

When combined with evidence of reference-dependent labor supply each day, these behaviors are attributable to workers setting daily goals and references for themselves throughout their contracts. The formation of daily goals in pursuit of long-run goals to induce loss aversion is a prediction of recent work on goal setting (Koch and Nafziger, 2016; Hsiaw, 2018; Koch and Nafziger, 2020). Daily goal setting of this nature suggests that reference dependence may be a rational response to self control problems. This analysis supports the behavioral intuition of those theories and provides real-world evidence of this phenomenon. Daily goals and expectations become the references around which workers exhibit reference-dependent labor supply. When firms influence those goals, they have substantial power to determine expectations for each day's performance.

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 (Camerer et al., 1997; Chou, 2002; Crawford and Meng, 2011; Farber, 2015; Morgul and Ozbay, 2015; Agarwal et al., 2015; Martin, 2017; Thakral and Tô, Forthcoming). I innovate in this area by considering the influence of the employer in the making of reference-dependent preferences through their effect on a worker's personal equilibrium. Lastly, I demonstrate that one area in which non-linear compensation schemes and other behavioral interventions affect workers is through the setting of expectations in the first place.

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 contract would introduce noise into my analysis. Because the conditions for my analyses are not set by experimental manipulation, it is possible that some sort of non-random selection pressure may influence the results, though, given the comprehensive battery of fixed effects and controls in my model, this is unlikely.

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. Future work using data from PestCo will examine the role of the company's rank-order tournaments in shifting the reference from one's own expectations to the expected performance of the opponent, thereby incorporating reference-dependent preferences into tournament theory.

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, simply increasing daily wages will not have the effect the neoclassical model would predict. The effectiveness of a wage increase depends on the worker's reference. My results suggest that the firm, rather than trying to motivate *around* a reference, can influence the *positioning* of the reference itself. If firms can use non-linear payments or other tools to shape the long-run targets and successfully influence the worker to set her reference higher, this will lead to great worker effort, great worker income, and greater profit for the firm because she spends more time in a loss domain.

My results also have implications for public policy. There are many examples of benefit cutoffs and kinks in the tax and transfer system. Each of these cutoffs and kinks could conceivably serve as a reference similar to the bonus payments in my sales setting. The policy choice of where to place these thresholds may affect a worker's annual income targets and therefore short-run labor supply choices. Policymakers exert considerable influence over labor markets through these policy choices.

References

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

Ariely, Dan, and Klaus Wertenbroch. 2002. "Procrastination, Deadlines, and Performance: Self-Control by Precommitment." *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.

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.

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.

Engström, Per, Katarina Nordblom, Henry Ohlsson, and Annika Persson. 2015. "Tax Compliance and Loss Aversion." *American Economic Journal: Economic Policy*, 7(4): 132–64.

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.

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.

Fryer Jr, Roland G, Steven D Levitt, John List, and Sally Sadoff. 2012. "Enhancing the efficacy of teacher incentives through loss aversion: A field experiment."Technical report, National Bureau of Economic Research.

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

Gilpatric, Scott M. 2008. "Present-Biased Preferences, Self-Awareness and Shirking." *Journal of Economic Behavior & Organization*, 67(3-4): 735–754.

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, National Bureau of Economic Research.

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): .

Holmstrom, Bengt, and Paul Milgrom. 1987. "Aggregation and Linearity in the Provision of Intertemporal Incentives." *Econometrica*, 55(2): 303–328, DOI: http://dx.doi.org/10.2307/1913238.

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.

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

Kucko, Kavan, Kevin Rinz, and Benjamin Solow. 2018. "Labor Market Effects of the Affordable Care Act: Evidence from a Tax Notch." *Available at SSRN 3161753*.

Levitt, Steven D, John A List, Susanne Neckermann, and Sally Sadoff. 2016. "The Behavioralist Goes to School: Leveraging Behavioral Economics to Improve Educational Performance." *American Economic Journal: Economic Policy*, 8(4): 183–219.

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.

Mortenson, Jacob A, and Andrew Whitten. 2020. "Bunching to Maximize Tax Credits: Evidence from Kinks in the US Tax Schedule." *American Economic Journal: Economic Policy*, 12(3): 402–32.

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.

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. Forthcoming. "Daily Labor Supply and Adaptive Reference Points."

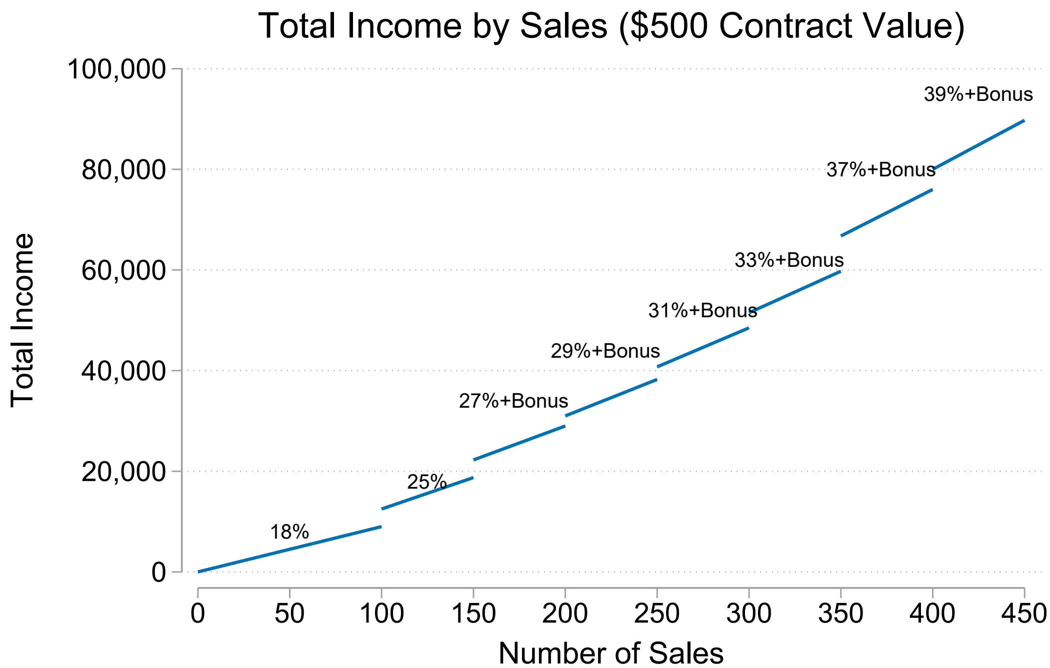
American Economic Review.

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

Wilson, Robert B. 1993. *Nonlinear Pricing*.: Oxford University Press on Demand.

Figures

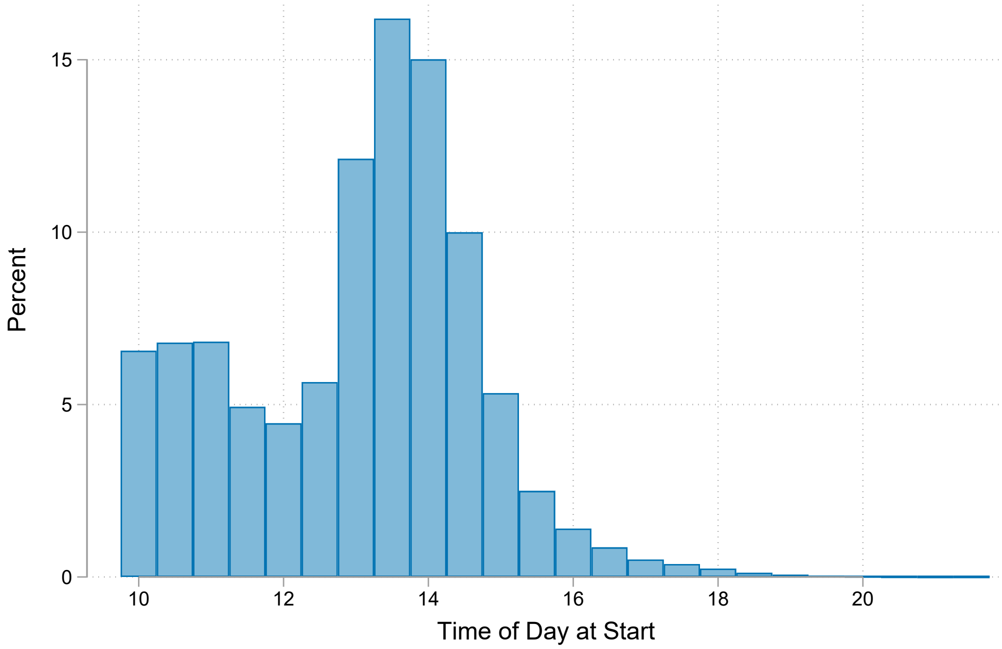
Figure 1: Contract Structure



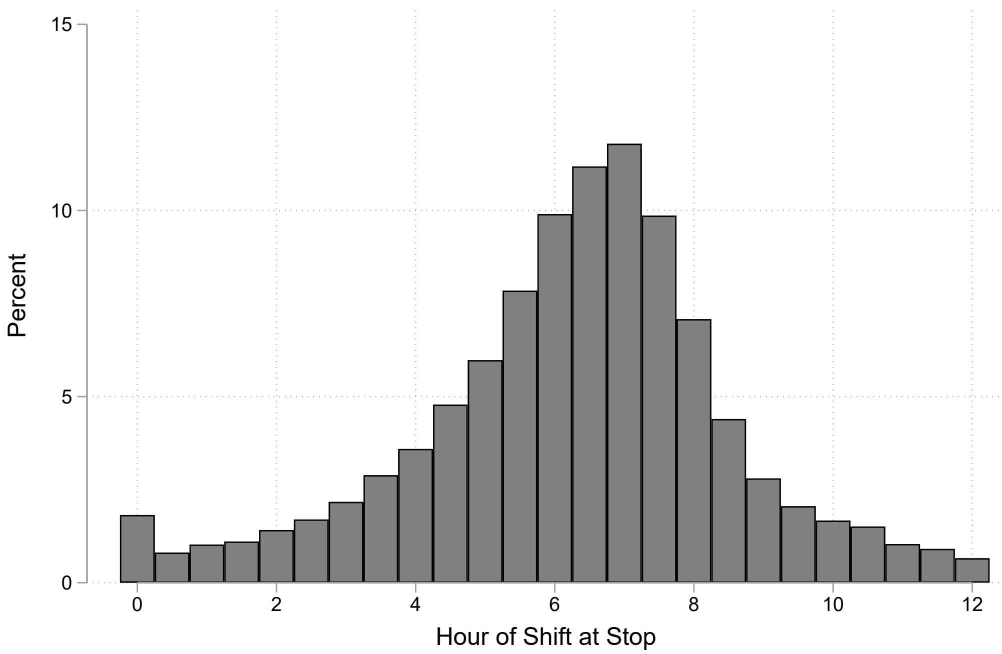
Source: Author's calculations of typical contracts from a pest control sales company.
Notes: Percentages indicate commissions as they are applied to each interval for all sales at the end of the season. At 150 sales, the "bonus" is that the company pays for the seller's rent for the summer in full (≈\$2,000).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

Figure 2: Distribution of Start and Stop Characteristics

Panel A: Time of Day at Start of Shift

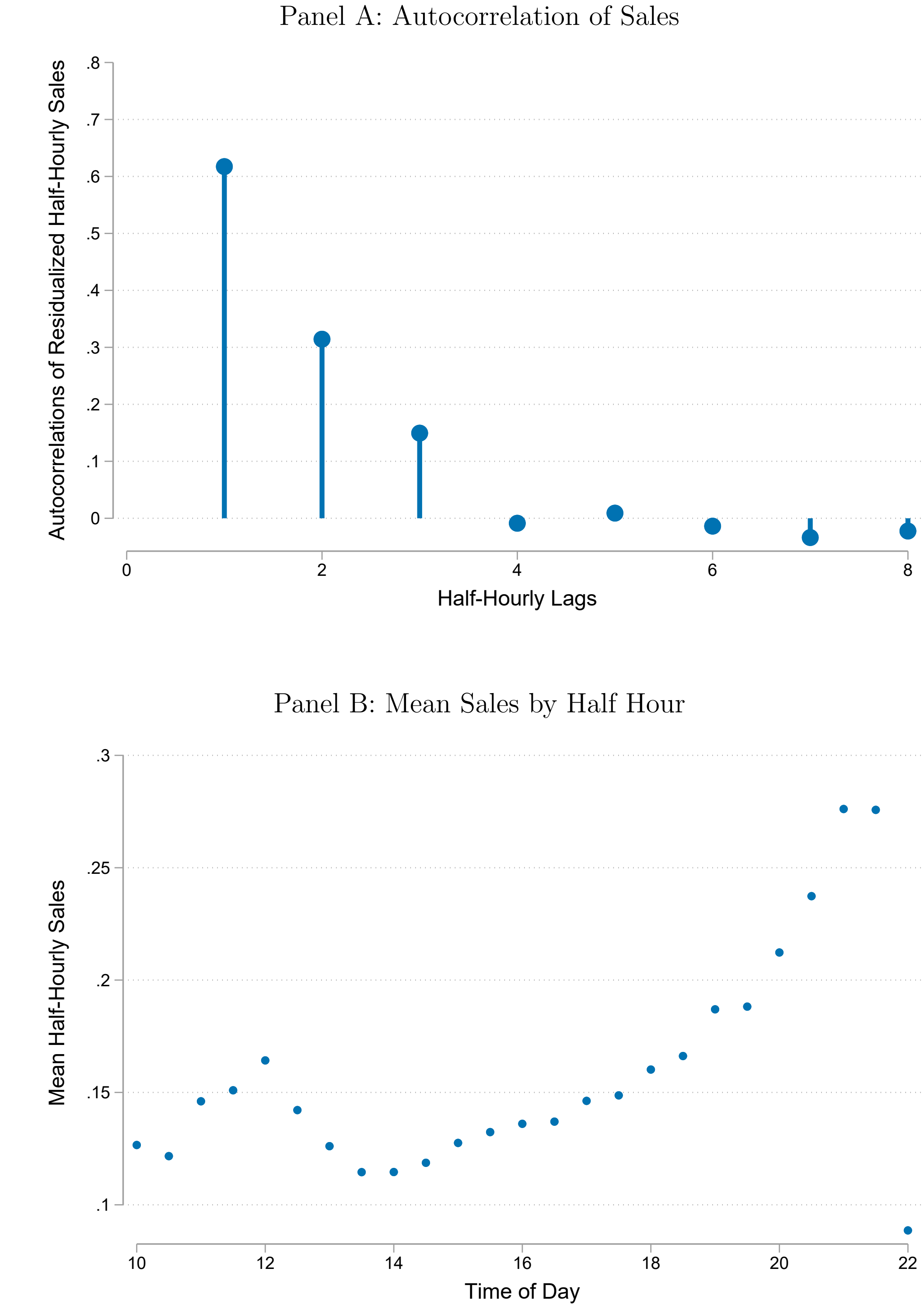


Panel B: Hour of Shift at Stop



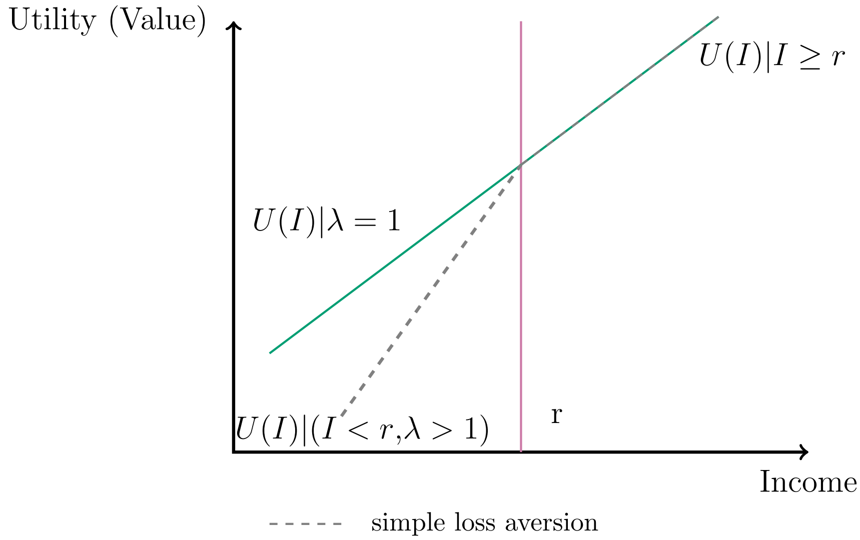
Source: Author's calculations of data from a pest control sales company.
Notes: Shifts begin during the half hour period when a seller first registers a knock or sale on each workday. Shifts end during the half hour they record their last sale or knock for the day.

Figure 3: Upward Pressures on Labor Supply During the Day



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.

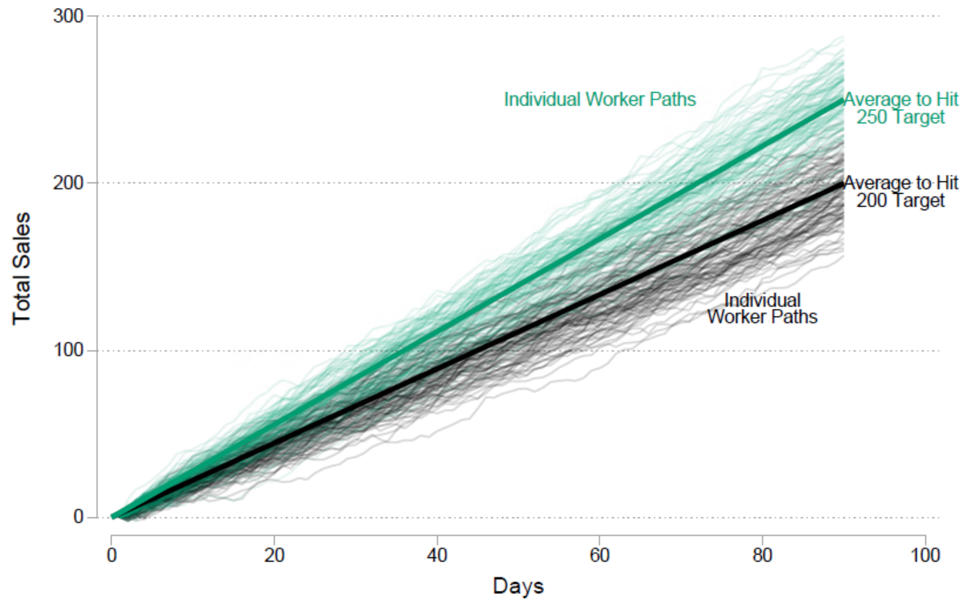
Figure 4: Illustration of Basic Model of Reference Dependence with Loss Aversion



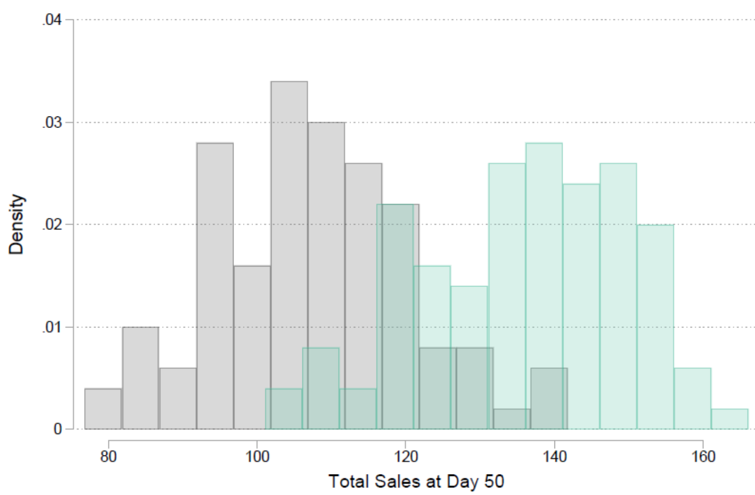
Notes: Illustration of basic loss aversion with linear utility over income. When $\lambda = 1$, the marginal utility above the reference r is the same as marginal utility below the reference.

Figure 5: Predictions of Standard Model with Two Goals

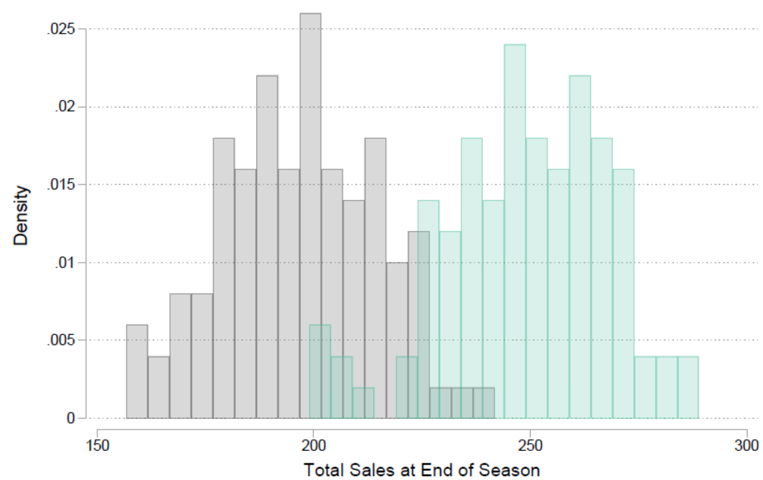
Panel A: Evolution of Sales Over Time



Panel B: Distribution of Sales at Day 50

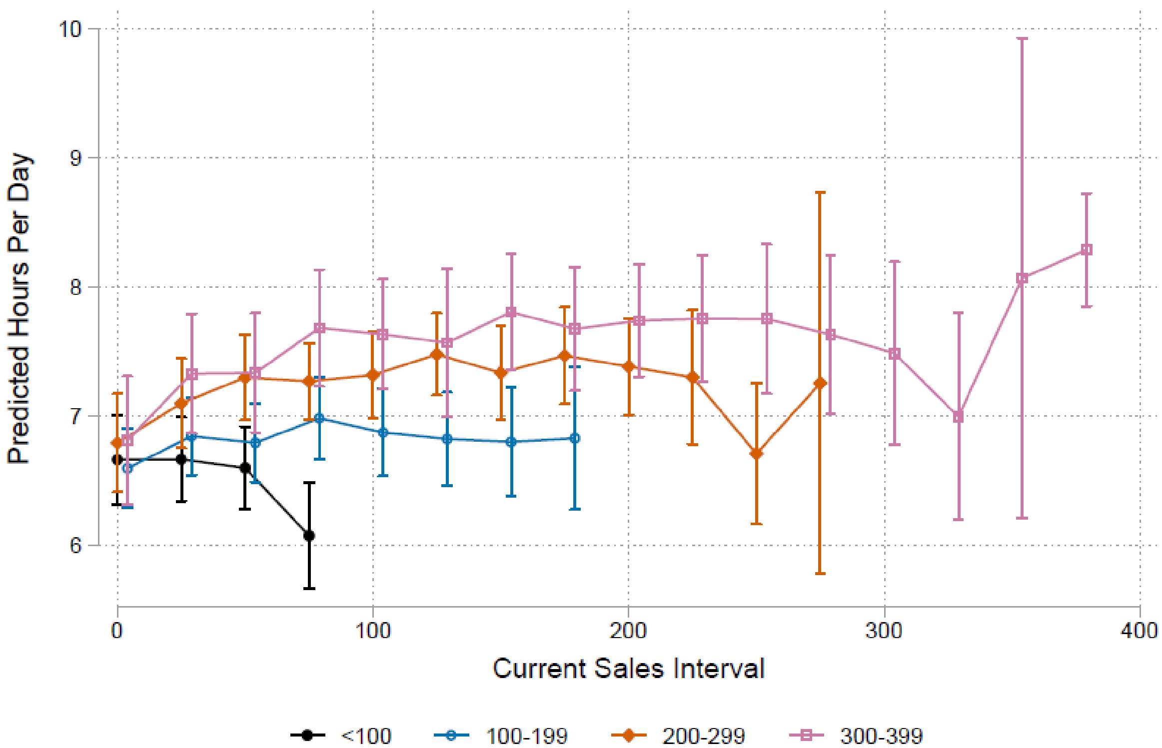


Panel C: Distribution of Sales at End of Season



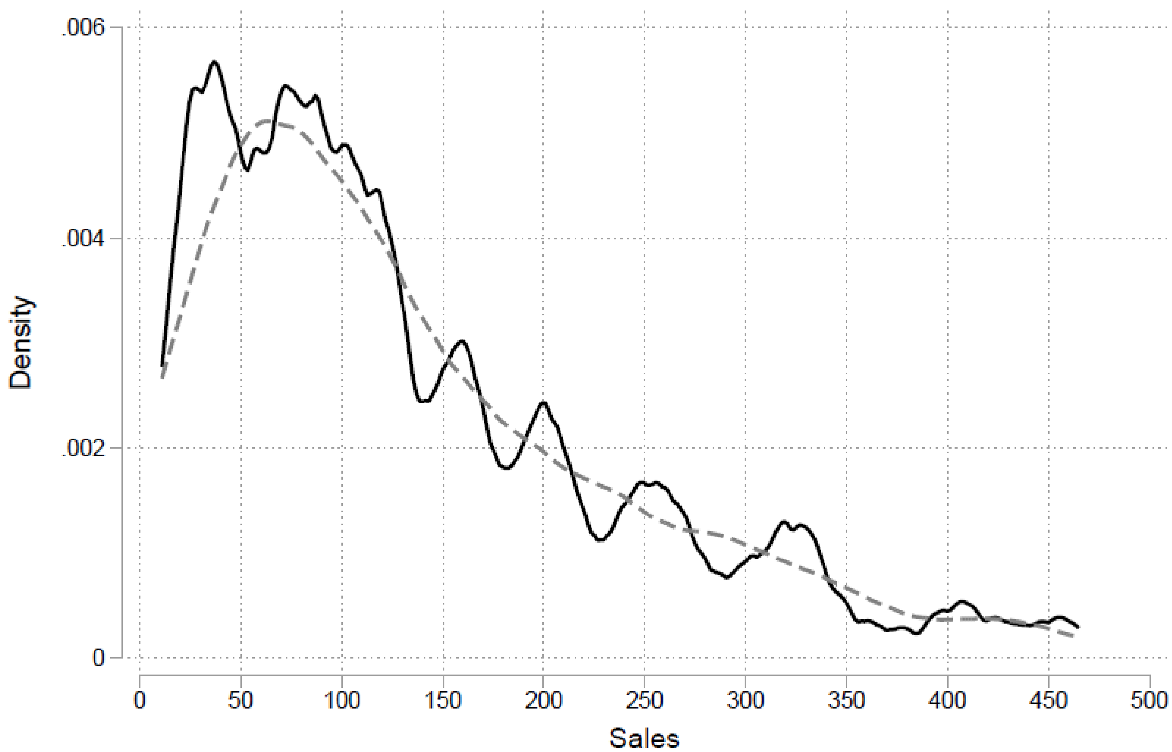
Source: Author's simulations of standard model. The simulations follow a Brownian motion with drift rates of 2.22 and 2.78 and a scale parameter of 2.0 as described in the text.

Figure 6: Predicted Labor Supply Over Current Sales Interval, By Final Season Sales



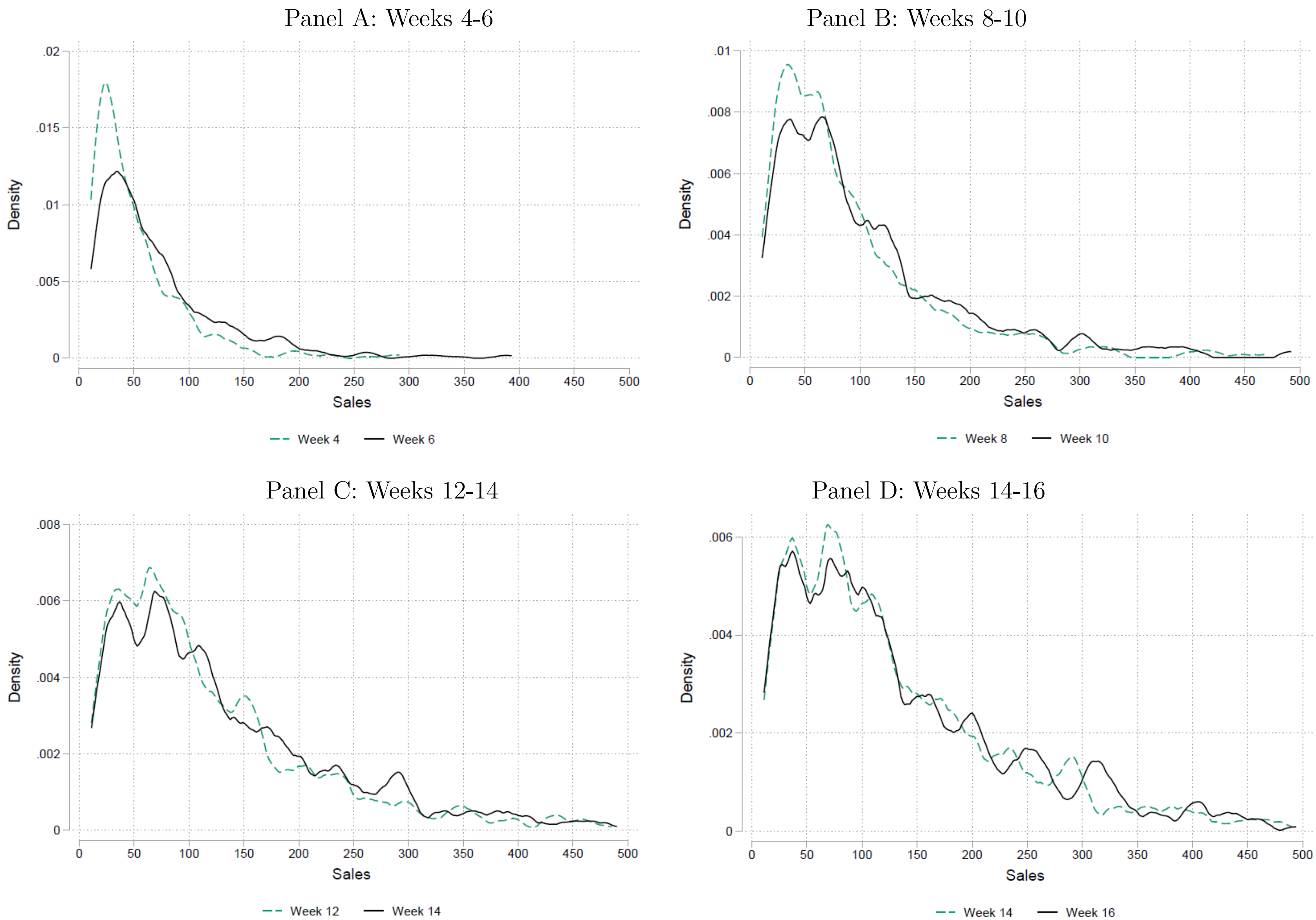
Source: Author's calculations of data from a pest control sales company.
Notes: Plot shows predicted hours from specification in Equation 3 for current sales interval (x-axis) separated by bins of total end-of-season sales. Standard errors are clustered at the seller level.

Figure 7: Kernel Density of Total Sales at End of Season



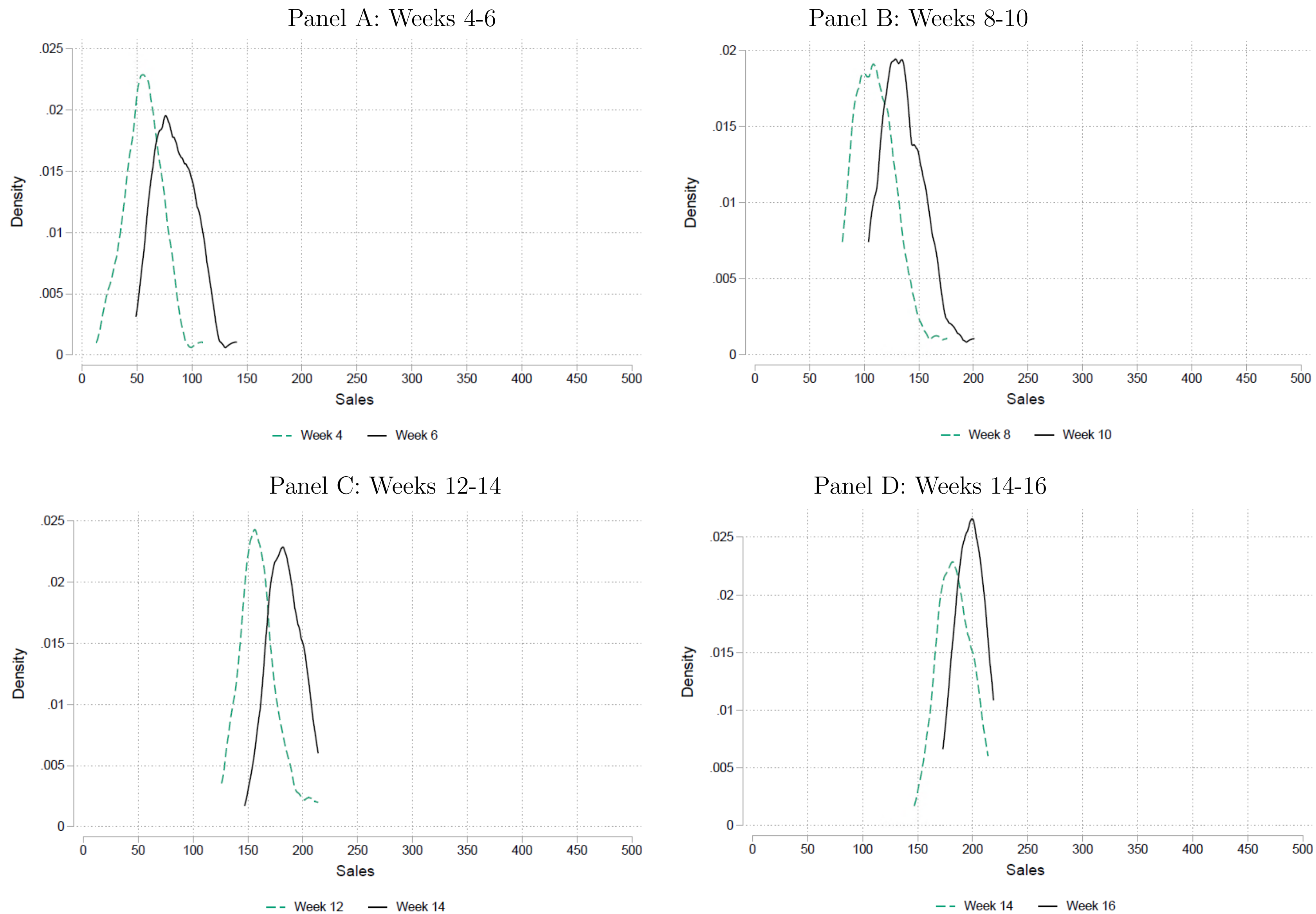
Source: Author's calculations of data from a pest control sales company.
Notes: Estimated using Epanechnikov kernel with bandwidth of 7 sales and 25 sales for sellers with at least ten sales and fewer than 500. The retroactive nature of the commission increases leads to a cash bonus upon hitting each 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 8: Kernel Density of Total Sales by Week



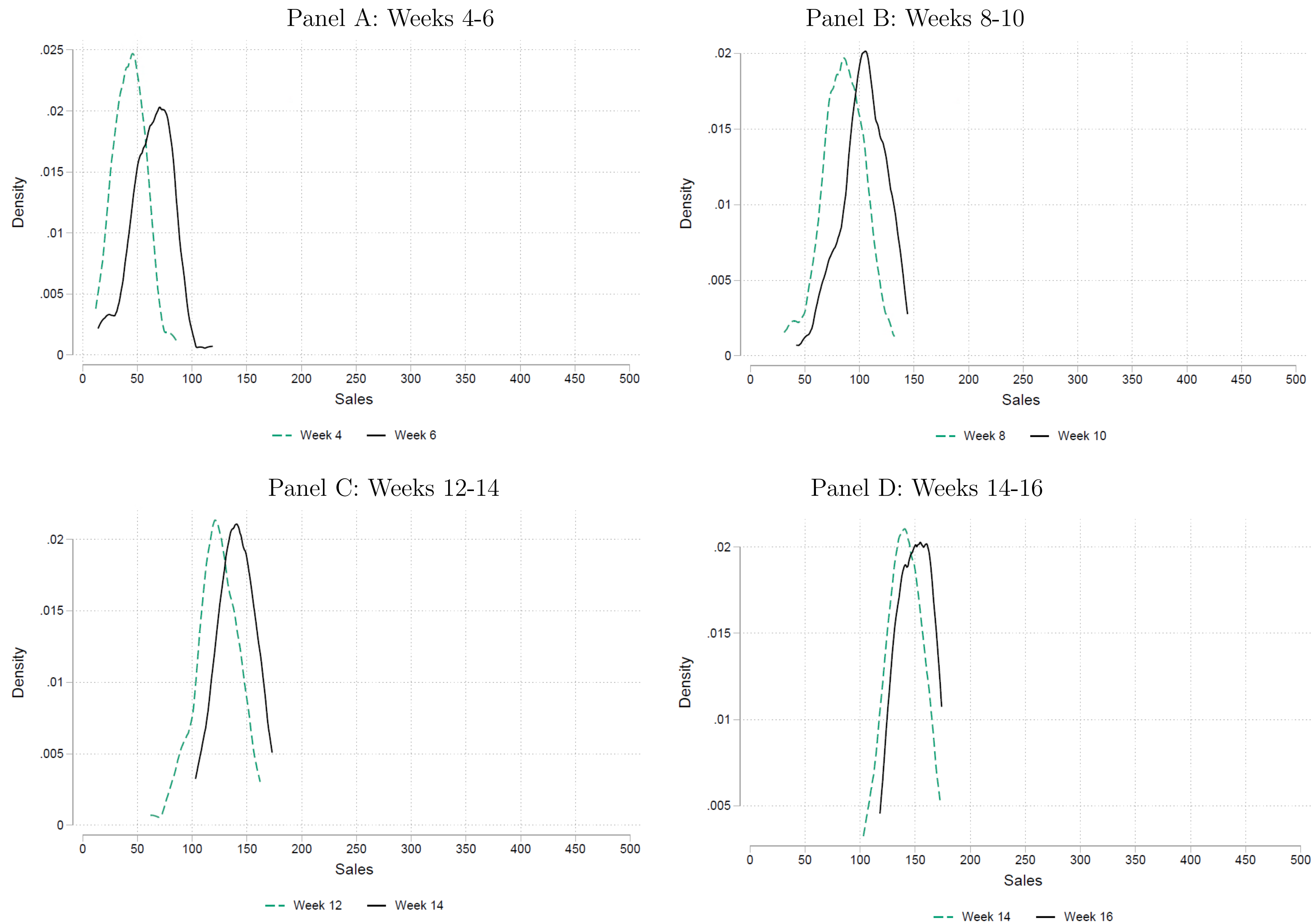
Source: Author's calculations of data from a pest control sales company.
Notes: Estimated using Epanechnikov kernel with bandwidth of 7 sales at the end of each estimated week among those with at least ten sales.

Figure 9: 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.

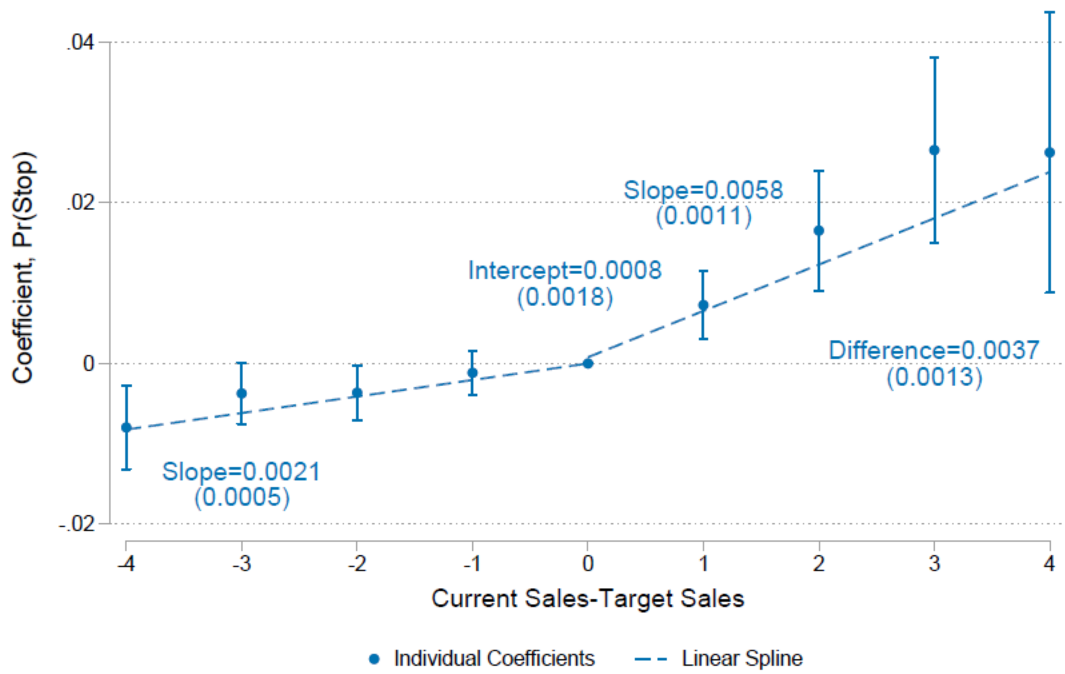
Figure 10: Kernel Density of Total Sales by Week
Workers with Total Sales of 125–175 at End of Season



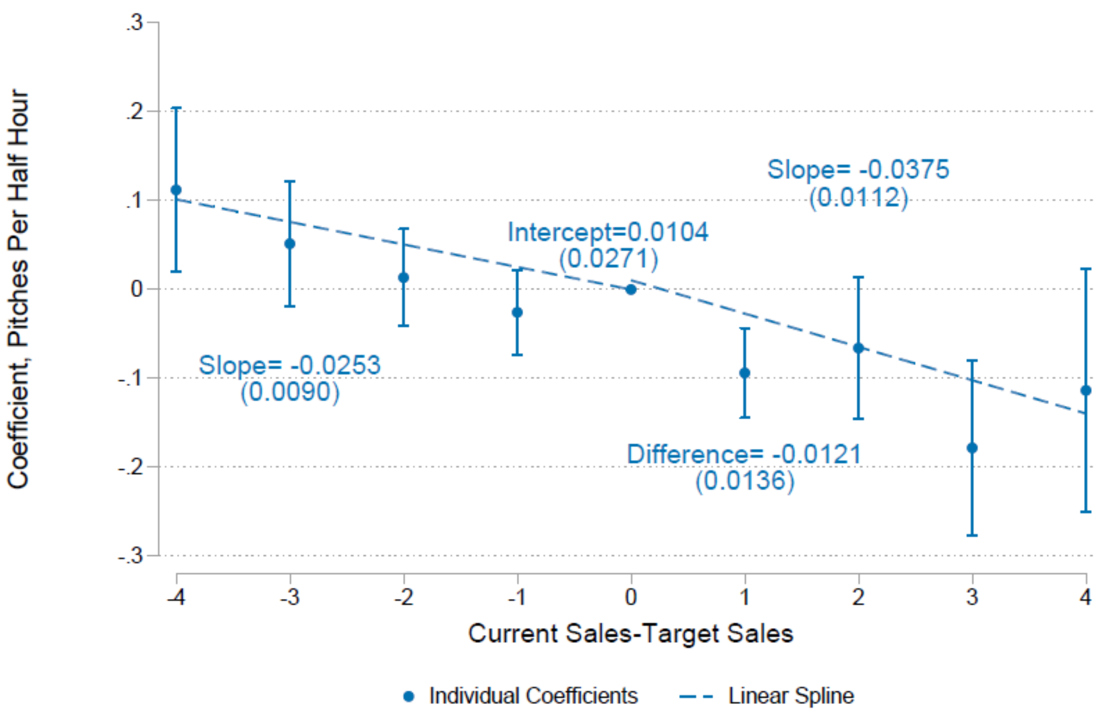
Source: Author's calculations of data from a pest control sales company.
Notes: Estimated using Epanechnikov kernel with bandwidth of 7 sales at the end of each estimated week among those with at least ten sales.

Figure 11: Labor Supply Around Expectations

Panel A: Probability of Stopping for the Day



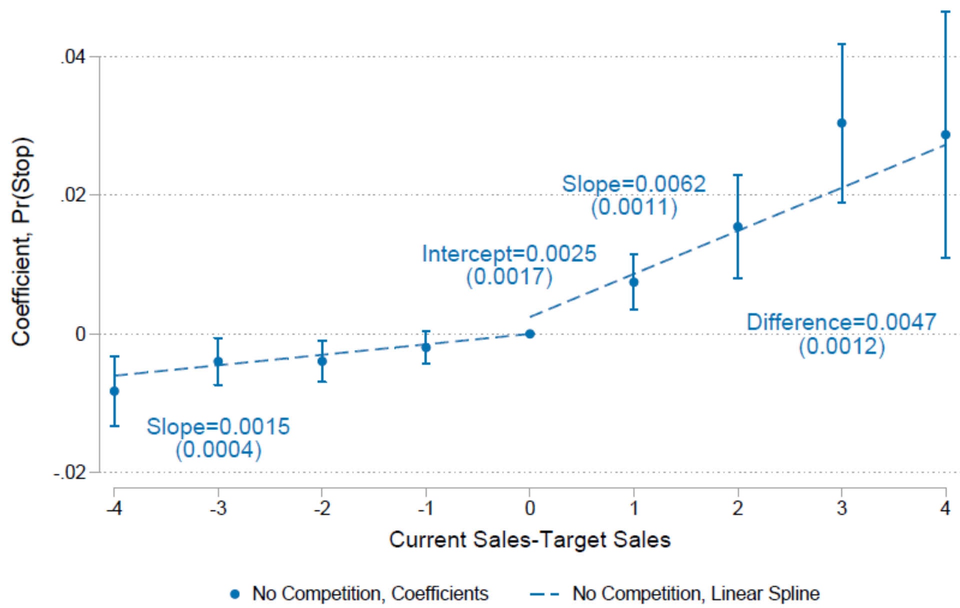
Panel B: Pitches Per Half Hour



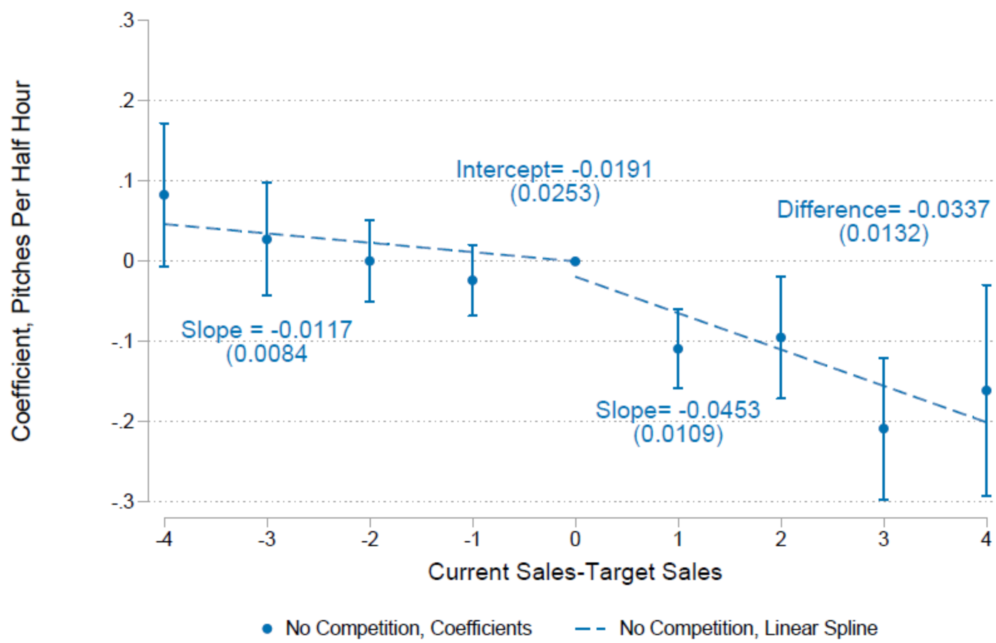
Source: Author's calculations of data from a pest control sales company.
Notes: Results are from estimates of Equations 5 and 6. Standard errors are clustered at the seller level. Individual coefficients are rounded to the nearest integer. The coefficient on +1 includes interval (0,1.5).

Figure 12: 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 5 and 6 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 interval (0,1.5).

Tables

Table 1: 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 2: 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 on Days that Exceeded or Did Not Exceed Expectations

Hours Worked Per Day	Did Not Exceed Expectations	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, NOAA daily weather data, and ACS data on ZIP codes.

Panel A estimates are from a regression of each seller's total sales at the end of the season on average daily sales during the first two weeks or five weeks of the season. Panel B estimates are from Equation 4 and include fixed effects for seller, day of the week, week of the season, and year as well as controls for weather and ZIP code characteristics. Standard errors are clustered at the seller level.

Table 3: Robustness Check: Non-Linear Least Squares

Model Parameters	(1) Pr(Stop)	(2) Pitches Per 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: The choice of cutoff is a parameter in the model. My non-linear least squares model selects the cutoff, slope, and intercept change parameters that minimize the sum of squared errors. Estimates use the residualized outcome variables from a regression on all fixed effects and controls to incorporate all controls and uses the residuals in the non-linear estimates. Standard errors clustered at the seller level.

Table 4: Robustness Check: Parametric Model Adding Intensive Margin as Control

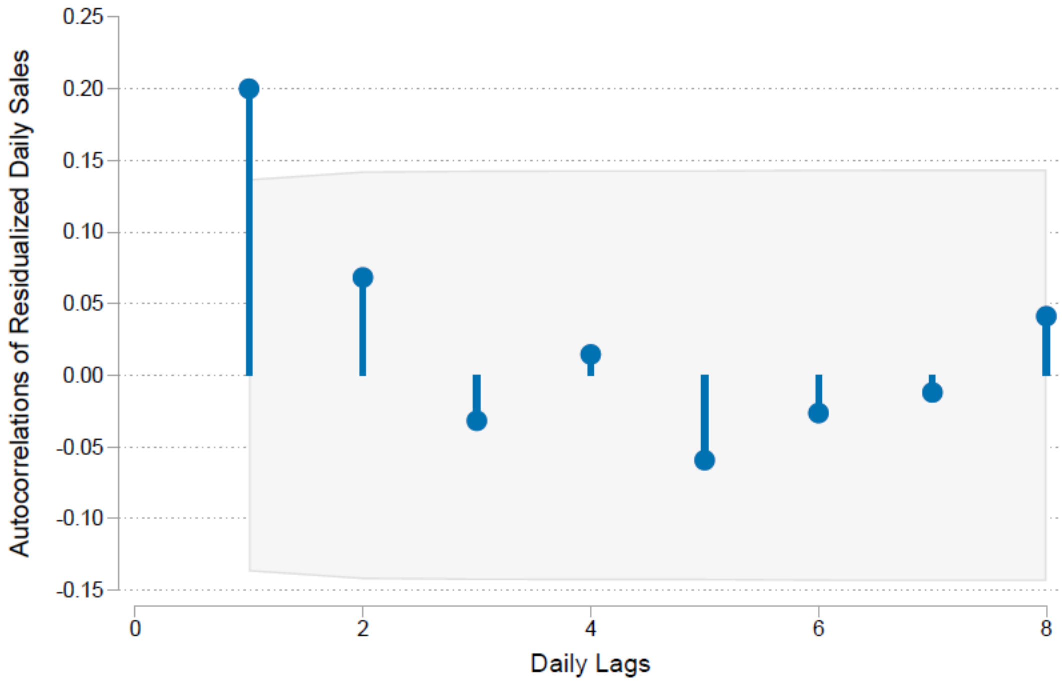
	(1) Pr(Stop)	(2) Pitches Per Half Hour
Cumulative Pitches	-0.0005***	0.0261***
	(0.000007)	(0.0013)
Slope Before Cutoff	0.0019***	-0.0175
	(0.0005)	(0.0079)
Slope Change at Cutoff	0.0035***	-0.00005
	(0.0013)	(0.012)
Intercept Shift at Cutoff	0.0007	0.0144
	(0.0018)	(0.0244)
Ratio of Slopes	2.8	0.997

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 6 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.

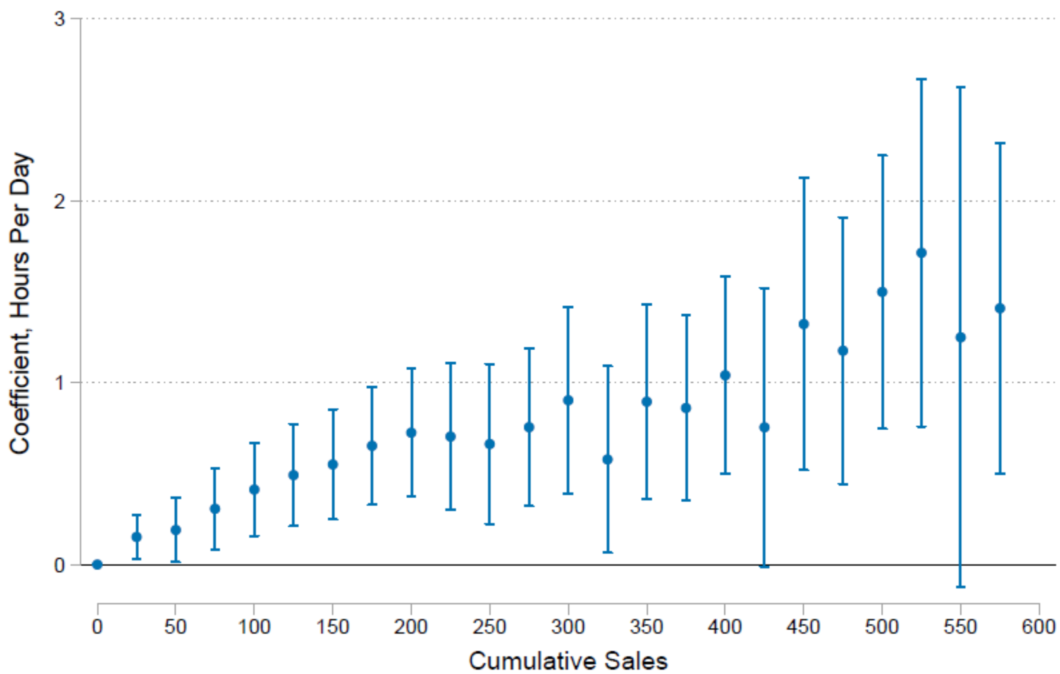
A Figures and Tables Appendix (Not For Publication)

Figure A1: 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.

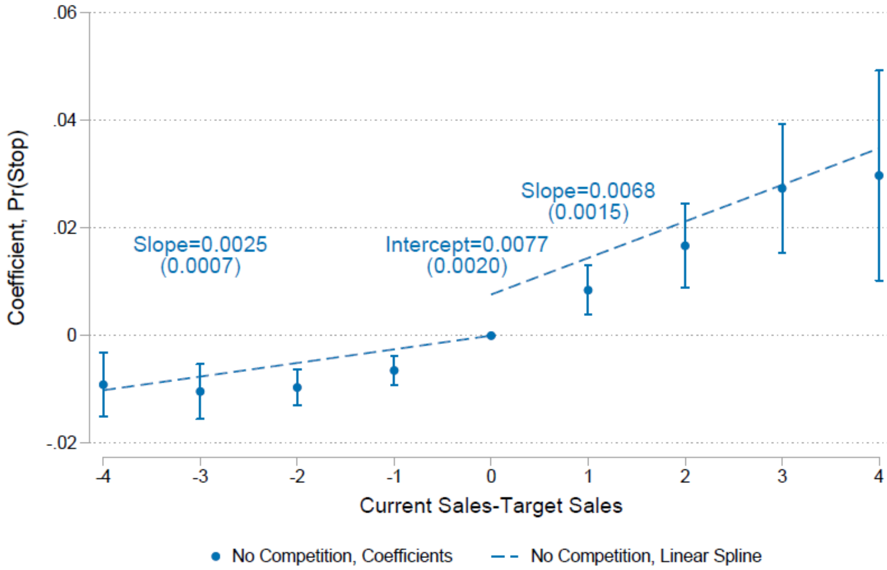
Figure A2: Daily Work Hours Over Cumulative Sales Intervals For All Sellers



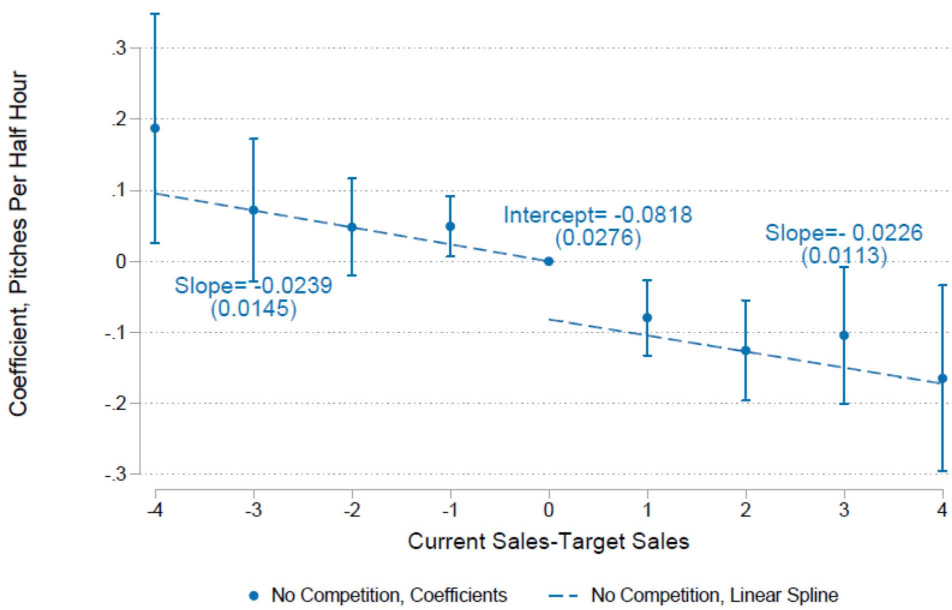
Source: Author's calculations of data from a pest control sales company.
Notes: Results are from estimates of Equation 3 except omitting fixed effects and interactions with final sales interval. Standard errors are clustered at the seller level. The differences between the results of this model and those in Figure 6 provide an example of Simpson's paradox. Because high-productivity workers work more hours on average across all sales intervals, the composition effects drive higher coefficients moving up the sales interval. Separately interacting the effects by subgroup reveals the flat trends noted in Figure 6.

Figure A3: "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 5 and 6 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 interval (0,1.5).

Table A1: Test of Location Sorting

Sales Per Day, All Significant Coefficients	(1) ACS	(2) Weather	(3) 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 A2: 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 5 and coincide with estimates from Figure 12. Standard errors clustered at the seller level.

Table A3: 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 6 but include interactions between indicators for each tournament/non-tournament period and distance to the reference. Standard errors clustered at the seller level.

B Data Appendix (Not For Publication)

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. Every sales contract in the data contains cancellation information, and I only include sales in my calculations that were not labeled as "not commissionable" due to immediate cancellation. In all, my half-hourly panel consists of approximately 459,000 observations for 512 sellers across 180 days in 2018-2019.

C Further Background Appendix (Not For Publication)

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 charged with generating contracts and scheduling 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 in terms of its use of incentive schemes. Their independent contractor agreements and practice are all in line with industry standards.

In calculating sellers' final commission rates, sales contracts generated must be effectual for at least two quarterly pest control treatments, or a period of six months. 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 are calculated at the end of the season after the status of all contracts are known. Final payouts for Spring sales are given in the Fall, and late Summer sales payouts are given at the end of the year. It is, therefore, in the best interest of the sellers to make sure that all sales are of high "quality," i.e. that the contracts will not likely be canceled early. 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 motivate them over the course of the summer. Sellers are trained 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 psychologically. In training materials, the company encourages their sellers to set effective schedules for themselves every day, to be physically active and healthy, to be brutally 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 taught to learn advanced sales techniques from their teammates. Unmarried 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. Area assignments, while not random, are not correlated with sales ability. Sellers often complain to managers that their work area is "hard." Managers will sometimes work with sellers making the claim and sell in their neighborhood for a few hours to demonstrate that sales can be achieved in the area if the seller works hard and refines his sales approach. Managers insist that "work area does not matter" in their training materials, and the evidence I present supports this argument.