

NBER WORKING PAPER SERIES

KINKS AS GOALS: ACCELERATING COMMISSIONS AND
THE PERFORMANCE OF SALES TEAMS

Peter J. Kuhn

Lizi Yu

Working Paper 28487

<http://www.nber.org/papers/w28487>

NATIONAL BUREAU OF ECONOMIC RESEARCH

1050 Massachusetts Avenue

Cambridge, MA 02138

February 2021, Revised June 2021

We thank Youssef Benzarti, Clement de Chaisemartin, Takao Kato, Alex Rees-Jones, Kathryn Shaw, Alisa Tazhitdinova, Gonzalo Vasquez, Guangli Zhang, and participants at the 2019 Trans-Pacific Labor Seminar (Tokyo) for helpful comments. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2021 by Peter J. Kuhn and Lizi Yu. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Kinks as Goals: Accelerating Commissions and the Performance of Sales Teams

Peter J. Kuhn and Lizi Yu

NBER Working Paper No. 28487

February 2021, Revised June 2021

JEL No. J33,M52

ABSTRACT

We study the performance of small retail sales teams facing an incentive scheme that includes both a lump sum bonus and multiple accelerators (kinks where the piece rate jumps upward). Consistent with standard labor supply models, we find that the presence of an attainable bonus or kink on a work-day raises mean sales, and that sales are highly bunched at the bonus. Inconsistent with those models, we find that teams bunch at the kinks instead of avoiding them. Teams' responses to the kinks are consistent with models in which the kinks are perceived as symbolic rewards, or where kinks induce a sufficient amount of loss aversion. Since bunching at both the target and kinks is strongest in larger, more experienced teams, we argue that these reference points are best understood as shared goals, used by teams to motivate and co-ordinate their members.

Peter J. Kuhn

Department of Economics

University of California, Santa Barbara

2127 North Hall

Santa Barbara, CA 93106

and IZA

and also NBER

pjkuhn@econ.ucsb.edu

Lizi Yu

lizi.yu@uq.edu.au

1 Introduction

Putting an accelerator or accelerators into the commission plan is the most common method applied. Put simply, the rate of commission paid is accelerated for every dollar earned above the set sales target. This encourages the salesperson to not only meet their sales targets, but to ‘smash their sales targets’! (MyFirstSalesJob, 2020)

Despite the well known potential costs associated with non-linear incentive pay schemes (Holmstrom and Milgrom, 1987), bonus schemes –which give workers a lump sum reward for attaining a threshold level of output– are widely used.¹ A rich empirical literature both confirms bonuses’ effectiveness in improving worker performance (Freeman et al., 2019), and demonstrates a wide variety of distortionary effects consistent with Holmstrom and Milgrom’s (1987) early concerns. These include mis-pricing of the product (Larkin, 2014; Owan et al., 2015), inefficient personnel retention decisions and distortions of subordinates’ incentives (Benson, 2015), and manipulating sales timing in ways that harm a company’s reputation. Theoretical work has also identified some conditions under which bonuses could have better incentive properties than linear rewards (Oyer, 2000).

Perhaps surprisingly, another form of nonlinear incentives that is widely used in workplaces –accelerators– has received much less theoretical and empirical attention.² Accelerators are performance thresholds at which the *slope* of the pay schedule jumps upward. In principle, accelerators could affect workers’ effort and performance through at least three channels. First, accelerators are sometimes combined with a lump sum reward. In these cases, both the discrete reward and the change in slope will affect performance– the discrete reward incentivizing workers to attain the threshold and the accelerator incentivizing them to exceed it. Second, accelerators divide the set of feasible performance levels into salient, disjoint ranges, which frequently have labels attached to them, such as “Tier1” and “Tier 2”, or “Good”, “Excellent, and “Superior”. These ranges could function as reference points or symbolic rewards that affect worker behavior. Finally, the increasing marginal pay associated with accelerators directly encourages workers to exceed the threshold, but –at least in the absence of attached bonuses or ‘behavioral’ determinants of effort– actually incentivizes workers to *avoid* output levels just around the threshold.

The goal of this paper is to study the effect of accelerating reward schemes on worker performance from both an empirical and theoretical perspective. The setting for our study is 103 small retail stores operated by a large clothing firm in China (hereafter “Firm A”) whose sales teams face a monthly pay schedule that include three thresholds. At the lowest of these

¹Douthit (1976) and Joseph and Kalwani (1998) respectively report that 80 and 72 percent of firms use bonuses to incentivize their salespeople. Oyer (2000) reports that 89 percent use sales quotas.

²O’Connell (1989) states that 53 percent of sales compensation plans have increasing commission rates, and Parrinello (2017) reports that over-quota accelerators are used in most sales compensation plans. Basu et al. (1985) provide theoretical arguments in favor of convex compensation plans. While the performance pay plan (PPP) studied by Lazear (2000) is an accelerating reward scheme (where the initial commission rate is zero), Lazear did not study its effects on output bunching or the role of behavioral factors in explaining its effects.

(henceforth the *target*) the commission rate rises and a small lump-sum bonus (worth about 3.5 percent of mean monthly pay) is awarded. At the other two thresholds, (henceforth the *kinks*) only the commission rate rises. The four team output intervals corresponding to these three thresholds are labeled as “Standard”, “Good”, “Excellent”, and “Superb” on a worker’s monthly pay stub. That said, no other monetary rewards, such as promotion are linked to attaining these ranges. Indeed, because the bonus is team-based and the company does not collect information on individual worker performance, there is no way for an individual worker to single herself out for promotion by performing better than her teammates. Nevertheless, workers may attach a symbolic value to attaining these ranges as a team.

We start our analysis by deriving the implications of the standard labor supply model for the distribution of team output around these two types of thresholds.³ We then compare these predictions with empirical estimates of output distributions at both the monthly and daily levels, in the latter case applying a *difference in density differences* (DDenD) estimator to quasi-random variation in the presence and location of an accelerator on the last day of a sales month. Consistent with the standard labor supply model, we find strong bunching of output levels at the target (where a bonus is paid). Also consistent, we also find that the presence of an attainable target *or* kink on the final day of a month raises mean sales on that day. Inconsistent with the standard model—which predicts missing density in the sales distribution at concave kinks—we also find strong bunching at the two pure kinks.⁴ This bunching is robust to controlling for a rich set of cyclical effects and team heterogeneity. Finally, we assess the ability of two behavioral models—one in which the kink acts as a loss aversion-inducing reference point, and one in which workers perceive attaining the kink as a symbolic reward—to account for this bunching. We show that the *loss-aversion* (LA) model can account for our results only if the induced loss aversion is sufficiently strong relative to the rise in commission rates at the kink. The *symbolic rewards* (SR) model can account for all the qualitative patterns we observe, regardless of the size of the induced symbolic reward.

In view of the well-documented gaming of bonus incentive schemes, it is noteworthy that several of the common unintended effects of non-linear pay are not relevant in our context. For example, mis-pricing is not possible because employees have almost no discretion on price. Workers’ ability to manipulate the timing of sales is limited because all sales are instantly recorded electronically, and relationships with customers are mostly short-term. Hasty sales will not result in returned items because purchases made at physical stores are not refundable at Firm A, which is typical for physical stores in China.⁵ At the same time, the sales target performs an additional role

³Throughout the paper we will model the behavior of sales teams as if they were individual workers. In Section 6 we estimate the effects of team size on team behavior and argue that Firm A’s piecewise linear commission schedule actually helps the teams address coordination and free riding problems.

⁴Kleven (2016) distinguishes between convex kinks (which maintain the convexity of the budget set and induce bunching) and non-convex kinks, which should produce gaps. We use the term *concave* for the second type of kink. Kleven reports that existing research on concave kinks is rare, and has not found much evidence of missing mass there. Seibold (2021) is the only study we know of that, like us, finds bunching at concave kinks. Those kinks appear to be driven more by a lack of information about Germany’s retirement system than by the symbolic incentives we argue are at work in our context.

⁵It is of course still possible for unsatisfactory purchases to hurt the firm’s reputation, but Firm A does not consider

for the firm that we do not model formally in the paper: In a high worker turnover context where sales traffic differs widely between locations and exhibits high fluctuation, the target –which equals each store’s sales in the same calendar month from the previous year– is a simple and effective way to communicate expected performance levels to workers.⁶ Overall, our analysis suggests that (a) accelerators are effective in motivating workers, and (b) their effect is not limited to the direct financial consequences of increasing marginal rewards. Accelerators have an additional motivational effect that is tied to the direct utility value of attaining different ranges of the compensation schedule.

In addition to the literature on non-linear rewards, our paper contributes to the literatures on symbolic rewards and goals. While a number of authors have studied symbolic rewards, much of the existing work focuses on *relative* symbolic rewards that induce a tournament among employees (Kosfeld and Neckermann, 2011; Ashraf et al., 2014). Awards of this type can have negative spillover effects on the workers who do not receive them, which can wash out their incentivizing effects on the winners.⁷ In our case, the symbol is activated when the team (comprising all the employees of the company at a particular workplace) attains a pre-specified *absolute* level of performance. Since everyone in a worker’s natural reference group either receives or does not receive the symbolic reward, our rewards should induce co-operation, not competition. Thus, an additional finding of our paper is that *team-based* symbolic rewards based on *absolute* performance can be motivating, without the negative spillover effects associated with awards for relative, individual performance.

The literature on goals studies situations like ours where the set of feasible performance levels is divided into intervals with potentially meaningful labels attached to them. In our reading, the existing work focuses mostly on individual performance targets that are chosen by workers themselves. Typically, these *commitment contracts* impose a monetary cost on workers who do not attain them (Royer et al., 2015; Kaur et al., 2015).⁸ More closely related to our work, Agarwal et al. (2017) study exogenously-assigned goals for a prosocial activity–resource conservation– with no financial consequences.⁹ Like us, they find that goals work best when they are neither too easy nor too hard, but they only consider loss aversion as a possible explanation for the goals’ effects.

Our study of accelerators also connects to a literature on non-linear pricing in other contexts, including retail sales. One aspect of the retail pricing literature focuses on the quantity discounts that are associated with different package sizes (Cohen, 2008). This situation is similar to ours in the sense that the consumer’s budget constraint is not convex– consumers face increasing marginal

this to be a major issue with its pay scheme.

⁶While a target based on last year’s sales raises the potential for ratchet effects, this is not a concern during our data period. Please see Section 2.2 for details.

⁷Such spillovers have also been observed when awards are given for high *absolute* levels of performance, such as perfect attendance (Gubler et al., 2016).

⁸Dobronyi et al. (2019) study such goals when no monetary cost is attached, and find they have a zero effect.

⁹Our context is also similar to the conversion of numeric to letter-based academic grades (Oettinger, 2002), though letter grades have real ‘lump sum’ consequences for academic promotion. In addition, letter grading schemes frequently have a competitive, zero-sum feature that is absent in our context.

incentives to buy. The bunching issue does not arise, however, because there is a small and fixed number of package sizes. Some papers in this literature (e.g. [Gu and Yang \(2010\)](#)) do however introduce non-standard preferences by letting the price per unit enter directly to transaction utility. Another aspect of the nonlinear retail pricing literature studies the effects of *rising* marginal prices for products like electricity and cell phone service ([Lambrech et al., 2007](#)), which would correspond to *decelerators* in incentive pay schemes.

Finally, our paper contributes to a growing applied literature devoted to estimating the amount of bunching, or ‘excess density’ in distributions of economic outcomes at theoretically relevant values. A number of papers in the tax literature ([Chetty et al., 2011](#); [Saez, 2010](#); [Kleven and Waseem, 2013](#); [Mortenson and Whitten, 2019](#)) address this question using large samples of draws from what is assumed to be a single distribution, in which there is a fixed potential bunching point of interest. For example, one may have millions of observations on the level of pre-tax income, and one is interested in the extent to which taxpayers bunch at the maximum EITC entitlement. The authors then fit a high-dimensional polynomial to the entire frequency distribution –excluding a small region of interest– and use the implied smoothness assumptions to identify the excess mass in that region.

In our case, while we observe team sales on a large number of days, our models suggest that we should not see any bunching or gaps on most of those days (either because there is no threshold within the team’s reach, or because reaching the threshold on that day is not a necessary condition for attaining it for the month). This gives us a large sample of *control days*. On the other days (which are a subset of the last days of the month), we expect bunching or gaps at particular sales levels that vary quasi-randomly from day to day, and from team to team. Our setting is complicated by the fact that other observable factors, such as day of week and store location, also affect the daily sales distribution and should be controlled for. As noted, we address this issue using a “difference in density differences” (DDenD) estimation approach that allows us to non-parametrically estimate the entire frequency distribution of daily team sales under a particular set of conditions.¹⁰ While our approach is not optimal for all contexts, it allows us, for example, to measure the extent to which the presence of an attainable kink in the \$1100-\$1199 bin on a given day and store shifts the entire distribution of output, not just in the area around the bin itself. These features may prove useful in other studies of the effects of nonlinear incentive plans, including bonuses, reference points, and payday effects, among others.

¹⁰[Pierce et al. \(2020\)](#) use a similar approach to estimate the effects of loss-framed performance incentives on sales distributions; unlike us, however they apply kernel smoothing to their density estimates. Our approach is also closely related to [Cengiz et al. \(2019\)](#), who directly estimate frequencies in all the bins of a wage distribution when a theoretically relevant quantity –in their case the minimum wage– falls into different bins in different times and places.

2 Background and Data

2.1 Retail Jobs and Firm A

Our data represent 103 retail stores operated by Firm A, a large manufacturer and retailer of men’s clothing in China.¹¹ We study the store performance of these stores during 2016, when the incentive scheme we analyze was in place.¹² Dropping observations in the month of store openings, closures or remodeling, we have 34,863 daily observations of team-level sales. The company does not maintain individual employee sales data, nor does it use them in setting pay.

Firm A operates retail stores in two types of locations: department stores and shopping malls. A typical store includes a counter and a display of products, and is operated by a sales team of 2-7 employees. According to the frontline salespeople, the yield rate per customer, i.e. the fraction of people who walk into the store making a purchase, is quite high for men’s clothing retail, at between 50 and 90 percent. Despite this high yield rate, employees and managers believe that sales amounts are quite sensitive to employee effort. In part this is because Firm A’s salespeople play a more active role than many U.S. salespeople: With very limited display space, most items must be retrieved from storage to be tried on. Customers also place a high value on fit and speed of service, both of which are sensitive to employee effort.

As retail sales also depend on customer traffic, daily sales fluctuate quite dramatically over time, as shown in Figure 1. This figure plots the average daily sales per store on every calendar day in 2015 and 2016. The smaller cycles in the figure represent output within a week, with higher daily sales on weekends. The largest spikes are labelled, and correspond to major holidays when people shop for menswear heavily, such as Chinese New Year or Father’s Day. Controlling for these large day-of-week and holiday effects will play a significant role in our econometric analysis. As shown in Table 1, the average daily sales output was \$582 in 2016, but daily sales were both ‘lumpy’ and unpredictable: only 11.3 relatively high-value items were sold per day, and the standard deviation of daily sales (\$1241) was over twice the mean and 4.5 times the median.

2.2 The Compensation Schedule

The employees in our sample are paid monthly, including a base salary and a commission. In 2016, the average monthly compensation was \$545, around 40% of which was commission. Within a store, the base salary varies with employees’ tenure, while commission rates are identical for all employees. Commissions are based on the *store*’s total monthly performance; they are set lower in

¹¹Our sample excludes 46 stores whose workers were not directly hired and paid by Firm A. Since the teams’ monthly targets, T , were given by their own sales in the previous year (2015), our sample also excludes stores that did not operate in 2015.

¹²Firm A’s salespeople had experience with similar types of incentive schemes in previous years, but the details varied somewhat from year to year. More specifically, by 2016, Firm A had used a similar nonlinear schedule –containing bonus(es) and accelerator(s)– for at least five years, but the targets were either determined by store employees themselves (subject to management approval), or by regional or higher-level managers.

larger stores, in order to yield similar daily wages across teams of different sizes, as indicated in Table 1.

As illustrated in Figure 2, Firm A's nonlinear incentive plan in 2016 includes a monthly sales target, T , at which a bonus is awarded and the commission rate increases, plus two pure accelerators where only the commission rate rises. If a store attains T , each team member receives a lump sum bonus of \$15.63 (100 CNY), and is exempt from a \$3.13 (20 CNY) penalty for not meeting the target, resulting in a jump equivalent to \$18.76 (120 CNY). This is worth 3.5% of employees' average monthly compensation. At the target T , the commission rate accelerates to 1.5 times the baseline rate. Next, there are two pure kinks in the schedule: The first accelerator is at $1.3T$, where the commission rate rises to 1.8 times the baseline rate, and the second is at $1.6T$, beyond which twice the baseline commission rate is paid.

At the beginning of each year, the store managers are informed of their team's monthly targets for the coming year. In 2016, a store's target T was equal to its sales in the same calendar month in 2015. In general, targets that are explicitly linked to past performance could lead to strategic output restriction, or ratchet effects (Charness et al., 2011), which would make each team's targets endogenous. This is not a concern in our setting, however, for three reasons. First, turnover at Firm A is high, which reduces workers' incentives to restrict output today in order to have a lower target one year from now. Second, 2016 was the first year in which Firm A tied the sales targets to a store's previous year's performance. Finally, the decision to link the 2016 targets to 2015 sales was announced after 2015 sales were realized. Related, Firm A had a policy of not adjusting its sales targets during the course of a year, even in response to unforeseen events such as the departure of an employee or construction in the host shopping mall.

In interviews, management described three main motivations for their use of this pay scheme. First, they believe that the target T is a simple and effective way to communicate their expected performance level to each store, in a context where customer traffic varies dramatically across stores and across months of the year (Ockenfels et al., 2015). Second, as argued in Kuhn and Yu (2021), Firm A's use of *team*-based incentives is part of an HR system that effectively delegates many key decisions, including recruiting and employee discipline, to the team members.¹³ Third, the firm's use of accelerating piece rates (without bonuses) at higher performance levels ($1.3T$ and $1.6T$) is based mostly on a stated desire to be consistent with other local retailers and retain high-performing salespeople. Indeed, Firm A's commission schedule—a target with a bonus attached, plus accelerators beyond that—was common among its competitors in the region. Thus we are studying a pay system that has withstood the test of time and competition, not an unfamiliar, experimental policy in a particular firm.

¹³Despite the theoretical possibility of free-rider problems, team-based incentives are used quite widely, including Continental Airlines (Knez and Simester, 2001), Microchip (Adamson et al., 2014), U.S. steel minimills (Boning et al., 2007), and apparel manufacturing (Berg et al., 1996; Hamilton et al., 2003). Examples from retail include Wal-Mart's profit-sharing plan, team-based bonuses in German retail establishments (Friebel et al., 2017), and tip pooling in restaurants (Scudder, 2017). Lawler and Mohrman (2003) report that the share of Fortune 1000 companies using work-group or team incentives for more than a fifth of their workers more than doubled, from 21 to 51 percent, between 1990 and 2002.

According to Firm A, aside from the bonus at T , no additional monetary rewards are attached to the nonlinear incentive plans. As noted, individual workers have no way to signal their promotion prospects by outperforming salient thresholds, since individual sales data are not collected. In the rare cases where store managers are promoted to head office positions, Firm A also states that those promotions do not use information on how frequently a manager’s team crosses these thresholds— education and interpersonal skills are the key criteria. This does not, of course, rule out the possibility that some managers *believe* that passing the thresholds raises their own promotion chances, and that those managers can motivate their team-mates to work hard to improve the manager’s perceived promotion opportunities. We assess this possibility in Section 6.3 of the paper.

With all these considerations noted, the presentation of pay information in workers’ monthly pay slips may attach symbolic rewards to this incentive scheme. Figure 3 presents a template of the pay slip distributed to salespeople. Specifically, the commissions are calculated and recorded separately for the four sales ranges that correspond to the available commission rates (i.e. $1\times$, $1.5\times$, $1.8\times$, or $2\times$ the baseline). On the pay slip, these ranges are labelled as “Standard”, “Good”, “Excellent”, and “Superb”, respectively. Moreover, at the bottom of the pay slip, there is an extra line called “Assessment”, which simply reports the highest of these categories attained by the team. We do not know precisely when this practice started, but salespeople recall instances from as early as 2014.

3 Theory– Predicted Output Distributions

In this section we compare the predictions of three simple models for the empirical distributions of output around the two types of thresholds in Firm A’s pay schedule— the target (T) and the kinks ($1.3T$ and $1.6T$). The *standard model* assumes that workers derive utility only from money and leisure, and have quasi-linear utility in money. Following [Saez \(2010\)](#) all our models assume that workers differ according to a single cost-of-effort parameter that is strictly positive and continuous on $(0, \infty)$. In the *symbolic rewards (SR)* model, in addition to the standard determinants of income and effort, teams perceive attaining the kink as a (lump sum) psychic reward that directly raises their utility. Finally, we consider a *loss aversion (LA)* model in which a kink creates a reference point, and induces loss aversion relative to it. Appendix A provides additional detail on the structure of the three models and derives the implications summarized below.

Turning first to the models’ predictions for the two ‘pure’ kinks ($1.3T$ and $1.6T$), the standard model predicts a gap, or missing density, in the distribution of output in an interval that (strictly) includes the kink, i.e both at the kink and on both sides of it. This result is well known in the literature on bunching estimators ([Kleven, 2016](#)). The symbolic rewards model, on the other hand, predicts missing density to the left of the kink, *bunching* at the kink, and excess density to the right of the kink. Intuitively, the symbolic reward is equivalent to a bonus that pulls some people up to the threshold, while the increase in slope pulls some people past it. Finally, the predictions of the loss-aversion model depend on whether the loss aversion around the kink is strong or weak

relative to the change in the commission rate. If loss aversion is not strong enough to outweigh the effects of the rising cash incentives at the kink, the model’s predictions are the same as the standard model– a gap in the output distribution around the kink. If loss aversion outweighs the acceleration in financial rewards, we should see bunching at the kink, and excess mass to its right.

Appendix A also derives the predictions of these three models for the distribution of sales output around the target, T . Now the predictions are the same for all three models, regardless of the level of loss aversion: bunching at the target, reduced mass to the left, and increased mass to the right. The intuition is that the cash bonus creates bunching at the target, regardless of the nature of the associated psychological rewards. For easy reference, all these predictions are summarized in Table 2, which also highlights the output patterns that we find –quite robustly– in our empirical analysis. Among the models considered, only the SR model, and the LA model *with strong loss aversion* can account for all the empirical patterns.

4 Descriptive Evidence

As a first look at sales patterns in Firm A’s stores, Figure 4(a) plots the distribution of sales across 1,143 store-month observations in 2016, where sales are measured relative to the team’s target T (i.e. its same-month sales in 2015). For every store, observations are grouped into bins of width $0.1T$ that are aligned with the target ($0.8T$ to $0.9T$, $0.9T$ to T , etc); consistent with Firm A’s target-setting formula, sales teams fail to reach their monthly target in 50% of all store-month observations. Consistent with the standard model described in Section 3, we see clear bunching at T ; this is, in fact where the bunching is most striking. Contrary to the standard model, however, we observe some bunching at $1.3T$. In contrast, there is no visual evidence of bunching, nor of the predicted gap, at the last kink point $1.6T$. Aside from these two mass points, the histogram elsewhere exhibits a smooth right-skewed bell shape.¹⁴

To quantify these visual impressions, Figure 4(b) applies the approach developed by Saez (2010), Chetty et al. (2011), and Kleven and Waseem (2013) to estimate the amount of excess density around T , $1.3T$, and $1.6T$. Using the number of observations that fall within each bin, we fit a high-degree polynomial (8th degree), excluding data in those three bins.¹⁵ Then, using the estimated coefficients from the polynomial, we extrapolate a fitted distribution to the excluded points to estimate the counterfactual distribution.¹⁶ The estimated bunching using this approach is 0.3030 at T , which is significant at 1% using the bootstrapped standard error. This implies that the

¹⁴The observations in the tails are mostly generated by the fact that the company’s targeting formula fails to account for holidays or sales promotions that occur in different months in different years. For example, the Chinese New Year arrived in February 2015 but in January 2016; as a result the teams’ 2016 sales targets were very easy in January, and quite unrealistic in February.

¹⁵In a typical application of this method, a narrow range of bins near the hypothesised bunching location will also be excluded, to allow the excess mass to diffuse around these points. We do not exclude neighboring bins for two main reasons. First, we would have to exclude all bins from $0.9T$ to $1.7T$, as the three kinks in our context are close to each other, and that would lose a major part of the distribution. Second, the bins used in our analysis already allow for considerable noise around the kink, because each bin of $0.1T$ is equivalent to three average days’ output.

¹⁶Details of this procedure are provided in Appendix B.

excess mass around the target is 30% of the average height of the counterfactual bin count. The estimated bunching at $1.3T$ is 0.2194, which is also significant at 1%. In contrast, at $1.6T$, we find negative bunching of -0.1002, but the estimate is not significant.¹⁷

While Figure 4(b) provides preliminary evidence of bunching around the target and the first kink, our ability to make sharp statements using this approach is limited by the fact that we have only 1,143 observations of monthly store performance, compared for example to the 11.6 million and 4 million observations used by Chetty et al. (2011) and Kleven and Waseem (2013) respectively. These limitations motivate our focus on daily sales, which allow us to focus our bunching tests only on situations where there is a close connection between the current day’s efforts and the probability of attaining a salient threshold in the current month. This connection is closest on the last days of months in which a threshold is less than three days’ worth of sales away, i.e. *within reach*.¹⁸ In monthly data, high levels of randomness in store traffic frequently make it impossible for teams to attain a threshold in some months, and push teams beyond a threshold well before the end of other months.¹⁹

To illustrate the relationship between teams’ output thresholds and the *daily* distribution of sales, we start by measuring each team’s ‘inherited’ output at the start of every day, t , as the sum of sales from the first day of the month up to and including day $t - 1$. Then, separately for each of the three thresholds, we restrict our sample to days in which the threshold is within reach, given the team’s inherited output. Next, in Figures 5(a), (b) and (c), we plot the distribution of each day’s *actual* output relative to the three thresholds (T , $1.3T$, and $1.6T$ respectively). We show these distributions separately for days that are the last of a calendar month versus all the other days on which T is within reach.²⁰ Sales are measured in \$100 bins relative to T , so a team that beat the target by \$350, for example, would fall in the middle of bin 3, and a team that beat the target by less than \$100 would fall into bin 0.

Again, and consistent with the standard model, Figure 5(a) shows a pronounced difference between the output distributions on the last day of the month –which is the team’s last chance to reach T – compared to all other days on which T is within reach. On the last day, a disproportionate share of teams sell just enough to beat the target by less than \$100. Contrary to the standard model, however, Figures 5(b) and (c) also show clear bunching at the two kinks, $1.3T$, and $1.6T$, but *only* on last days of the month. Together, these results suggest that teams who find themselves within reach of a salient threshold on the last day of the month tailor their efforts to just reach the threshold

¹⁷Interestingly, this figure also shows evidence of bunching at $2T$, which supports the possibility that round numbers also serve as symbolic rewards, or reference points. This phenomenon is also found in Benson (2015) and Allen et al. (2017).

¹⁸Table 6 considers the effect of alternative cutoffs for defining a threshold being within reach. Appendix C.2 uses an alternative definition of daily mean sales to calculate ‘reachability’.

¹⁹Specifically, the target, T , is within reach on only 11.5 percent of the last days of the month. The kinks $-1.3T$ and $1.6T$ – are within reach on 7.3 and 3.3 percent of all days. The analysis in Figure 4(b) includes not only these months, but all the other months on which prior output shocks make thresholds irrelevant to Firm A’s sales teams.

²⁰Note that the daily sales distributions in Figure 5 tend to be centered a little below T . This is a consequence of our three-day window for a threshold being within reach: With such a window, the threshold is on average about 1.5 days of sales away on the morning of the current day.

on that day. In Appendix C.1 we repeat this exercise for placebo thresholds at $0.9T$, $1.1T$, $1.2T$, $1.4T$, $1.5T$, and $1.7T$. Here, there is no visible difference between the two distributions (with the possible exception of $1.5T$, where a few teams may have aggressively pursued the highest kink at $1.6T$ and fallen just short).

5 Estimation Approach

Motivated by the descriptive evidence in Figures 4 and 5, our main econometric approach will compare a sales team’s output on last days on which a threshold (T , $1.3T$ or $1.6T$) is within reach to the same team’s output on all other days.²¹ As noted, the distinction between these two types of days is whether or not there is a clear connection between a sales team’s effort on that day and the team’s chances of attaining a salient sales threshold. Compared to the more familiar bunching methods discussed in the previous section, a key advantage of our approach is the large sample of control days, which allow us estimate the effect of the presence *and* location of a threshold on the entire daily sales distribution, without the need for smoothness assumptions. Instead, the sales distribution is represented by a finite number of bins, with an arbitrary baseline density attached to each of them. In this sense, our estimation approach is similar to [Abeler et al. \(2011\)](#), who study how the excess density in workers’ performance levels around an induced reference point *moves* when that reference point is manipulated by the experimenters.²²

To implement our approach, we create a set of \$100 daily sales bins for every store, which run from zero to the maximum daily sales we see for that store in our data period.²³ With 103 stores in our sample, this yields 34,863 daily store sales observations and 2,778,836 store \times day \times bin observations.²⁴

5.1 Estimation without Covariates

To illustrate how our approach works, we first consider an example where the reward schedule has a single threshold (for example, at the sales target, T), and no adjustment is made for covariates—for example, factors like weekends that affect the distribution of sales. To simplify the discussion, we will define a threshold as *attainable* on a given day if that threshold is within reach on that day, *and* that day is the last one in a month. In this case we would estimate the following linear

²¹Appendix C.3 shows that our results are robust to a variety of alternative control days.

²²Our approach is also similar in spirit to [Benson \(2015\)](#), who creates a ‘treatment bubble’ of managers whose cumulative performance near the end of a pay period makes their current actions decisive for attaining a bonus.

²³Since stores differ in size and performance, this yields a different number of bins for each store. While this might seem cumbersome, forcing every sales bin to be \$100 wide has the advantage that we can unambiguously interpret ‘hitting’ the bin that meets the team’s monthly target as being within \$100 of the target. More generally, our ‘binned’ approach accommodates noise in the production process: By classifying any level of sales in the same \$100 bin as being ‘at’ the target, it includes teams that ‘tried but just missed’ as ‘bunchers’.

²⁴To assess the robustness of our approach to constructing sales bins, Appendix C.4 uses a bin size that is proportional to each store’s target output, and shows that the main results are very similar.

probability model:

$$S_{it,b} = \alpha + \beta_0 \cdot K_0 + \sum_{j=-5}^{-1} \beta_j \cdot K_j + \sum_{j=1}^5 \beta_j \cdot K_j + I_i + \epsilon_{it,b}. \quad (1)$$

Our dependent variable, $S_{it,b}$ takes a value of one if store i 's actual sales on day t fall into bin b , and zero otherwise. Our main regressor of interest, K_0 is a dummy variable that equals one if T is attainable on that day *and* the current sales bin contains the target T . In addition, we include indicators for the five sales bins that are just above or below the target's bin on the same day. Thus, the coefficient β_0 estimates the additional likelihood that a store's sales output falls into bin b when that bin contains an attainable target. The ten coefficients β_{-5} through β_5 identify any excess or missing mass in the output bins surrounding an attainable target. Store fixed effects I_i are included to account for the fact that the number of output bins varies across stores. To measure the effects of the two 'pure' kinks affecting the density of sales, we define regressors analogously for $1.3T$ and $1.6T$, and we re-estimate equation (1) for the analogously defined regressors. Estimates of equation (1) for T , $1.3T$ and $1.6T$ are presented in columns 1-3 of Table 3.

5.2 Adding Covariates

One limitation of equation (1) is that it does not account for factors like holidays and the day of the week to affect the sales distribution. While we cannot think of an obvious reason why these factors will be correlated with the presence of an attainable sales target *in a particular bin* on the final day of a month, these factors are highly correlated with total daily sales, so it seems reasonable to control for them. Another limitation of equation (1) is the fact that it does not contain *bin* fixed effects. Thus it implicitly assumes that the baseline sales distribution (i.e. the distribution on days when no threshold is attainable) is uniform.

To address the above issues, we adopt a set of fixed effects that allow each store to have its own baseline sales distribution with unrestricted shape. In addition we allow each store's baseline distribution to have a different shape on each day of the week. Both of these goals are accomplished by giving each store its own set of DOW \times bin dummies. Similarly, to account for the substantial improvement in sales during holidays, we control for Holiday \times bin fixed effects, where Holiday is defined as a categorical variable that identifies 10 major shopping holidays.²⁵ This yields the

²⁵The holidays include New Year, Chinese New Year, May Day, National Day, Father's Day, November 11th, December 12th, and three other traditional Chinese holidays.

following regression: ²⁶

$$S_{it,b} = \alpha + \beta_0 \cdot K_0 + \sum_{j=-5}^{-1} \beta_j \cdot K_j + \sum_{j=1}^5 \beta_j \cdot K_j + \Phi_{iDOW,b} + \Theta_{H,b} + \epsilon_{it,b}, \quad (2)$$

where *DOW* indexes the days of the week and *H* indexes ten major holidays. All told, the control variables $\Phi_{iDOW,b}$ and $\Theta_{H,b}$ comprise 56,252 store \times DOW \times bin dummies and 3,971 holiday \times bin dummies. Standard errors are clustered by store. Clustering by stores accommodates a potential concern with estimating a complete distribution of sales for every store: Because sales must fall into exactly one of a store’s bins on every day, there will be a mechanical correlation of errors across the bins within every store \times day cell. Thus it seems essential to cluster at least at the store \times day level. Clustering at the store level should handle this issue, plus more general concerns such as autocorrelation of errors within stores over time. With heterogeneity across time and stores being accounted for in this way, equation (2) again tests whether sales are more likely to fall into a bin that exactly attains a threshold, compared to what we would expect that store \times day to produce otherwise. Analogously, we create regressors for the two kinks at $1.3T$ and $1.6T$. Estimates of equation (2) for T , $1.3T$ and $1.6T$ are presented in columns 4-6 of Table 3.

5.3 Threats to Identification

A natural question concerning our approach is whether the variation we exploit –the presence and location of an attainable threshold on the last day of a month– is ‘as good as random’. We offer three observations on this question. First, we note that the usual sorts of omitted variables that would affect an analysis of daily mean sales –for example a persistent shock to store output (perhaps due to a sales promotion) that could both bring a threshold within reach and raise the store’s chances of exceeding it– do not apply to our approach. Such shocks can raise mean sales on adjoining days, but they have no obvious effect on the probability that both the threshold and actual sales fall into the same \$100 sales *bin* on the last day of the month. A second and perhaps more plausible source of endogeneity actually works in our favor. Suppose that –despite the unpredictable nature of daily sales– some teams successfully adopted a forward-looking strategy of tailoring their sales on the last few days of the month so as to put a particular threshold within ‘reasonable’ reach on the final day.²⁷ If this type of forward-looking behavior accounts for our main results, it only strengthens our primary hypothesis, which is that teams tailor their sales efforts to ‘just reach’ salient thresholds in the monthly pay schedule. Third, Appendix C.3 presents a number of specification tests to show that our results are not sensitive to the use of a number of alternative control groups, including

²⁶Like equation 1, equation 2 focuses on the monthly target, T . Estimating equations for the two kinks, $1.3T$ and $1.6T$, are defined analogously. Notably, it is essentially never the case that more than one threshold is within reach on the same day. This is because the thresholds are about $0.3 \times 30 = 9$ days worth of sales apart, but a threshold must be within three days of sales to be within reach.

²⁷More specifically, suppose that teams try to get the threshold into a region we call the ‘sweet spot’ on the last day. A threshold is in the sweet spot if it is attainable with between one and two typical days’ worth of sales; in Section 6.3 we find that thresholds in this location account for much of the bunching we observe.

control groups drawn only from days closer to the start of the month, when strategic planning to target a precise output bin on the last day seems much less likely.

6 Results

6.1 Main Results

Estimates of equations (1) and (2) are presented in Table 3. Notably, across all six columns, we find substantial bunching at the target T , and also at the two pure kinks $1.3T$ and $1.6T$. In the absence of controls, column (1) shows that a \$100 sales bin that exactly contains an attainable target T is 17.1 percentage points more likely to be exactly achieved than the same bin when it does not contain an attainable target. In the presence of controls (column 4), this estimated coefficient is 9.19 percentage points. Column (4) also shows statistically significant missing mass in the bin just below target, equal to 4.7 percentage points. Columns 5 and 6 show estimates from the same regression for coefficients corresponding to the two kinks, $1.3T$ and $1.6T$. These show excess density of 11.4 and 14.9 percentage points respectively. Overall, in all three columns (4) through (6), we find sizable and significant bunching effects at targets and kinks, but –with the exception of the bin just below the target– no strong evidence of nearby missing masses.

To better interpret the excess densities in columns (4) - (6), it is helpful to have an estimate of the baseline density that we would expect to see in the bins that ‘tend to’ contain a target or kink, when no target or kink is present. To measure these densities, we use a two-step procedure detailed in Appendix B: First, we calculate for each store the probability the threshold falls into every one of the store’s output bins. Then we use these probabilities as weights to calculate the expected density in the output bins where an attainable threshold would ‘typically’ land on control days. These densities are 10.3, 7.7 and 5.9 percentage points for T , $1.3T$, and $1.6T$ respectively. (Because the higher thresholds are attained less frequently, their baseline densities are lower.) Relative to these baseline densities, the 9.19 percentage points of excess mass at the target T correspond to an 84 percent increase. The excess densities of 11.4 and 14.9 percentage points at $1.3T$ and $1.6T$ correspond to increases of 149 and 252 percent respectively. Thus the estimated effect sizes are substantial in magnitude.

To shed additional light on where the additional density at attainable thresholds in Table 3 is coming from, Table 4 estimates the following specification:

$$S_{it,b} = \alpha + \beta_0 \cdot K_0 + \beta_{-5,-1} \cdot K_{-5,-1} + \beta_{1,5} \cdot K_{1,5} + \beta_{\leq -6} \cdot K_{\leq -6} + \beta_{\geq 6} \cdot K_{\geq 6} \quad (3)$$

$$+ \Phi_{iDOW,b} + \Theta_{H,b} + \epsilon_{it,b},$$

Specifically, we create a dummy variable, $K_{-5,-1}$ equal to one in all five of the bins just below an attainable threshold, and a dummy variable $K_{\leq -6}$ for all the remaining bins that are further

below an attainable threshold. Similarly, we define $K_{1,5}$ and $K_{\geq 6}$ to identify the five bins just above an attainable threshold, and all remaining bins that are beyond the threshold. Notably, a dummy like $K_{\leq -6}$, which represents several underlying output bins, gives the effect of the presence of an attainable threshold on a given day on the probability that output will fall in *one* of those five bins. To estimate the effect of the presence of an attainable threshold on the total probability mass in a region represented by such a dummy, the coefficients reported in Table 4 scale up the estimated coefficient estimates by the number of \$100 bins each coefficient represents.²⁸ When scaled in this way, the coefficients in Table 4 summarize how the entire sales distribution shifts on days when an attainable threshold is present, controlling for $\text{Bin} \times \text{Store} \times \text{DOW}$ and $\text{Bin} \times \text{Holiday}$ fixed effects.

Column 1 of Table 4 shows that the presence of an attainable monthly target, T on a day creates an extra 9.2 percentage points of density in the exact \$100 bin that contains the target, and an extra 10.2 percentage points of density in the five bins immediately above the target. This extra density above the target is consistent both with the increase in the commission rate at the target, and with teams unintentionally exceeding the target due to the ‘lumpiness’ of sales. Interestingly, of this 19.4 percentage point increase, 6.9 percentage points come from the five bins just below the target, and 10.3 percent come from bins below that.²⁹ Thus, the target appears to be ‘pulling’ teams upwards from a substantial distance below it. Estimates for the two pure kinks exhibit similar properties, with much of the extra mass at and above the kinks coming from more than five output bins below. While lumpiness of sales and accelerating marginal incentives likely contribute to these patterns as well, Section 6.3 shows that most of the effects in Table 4 are driven by Firm A’s largest and highest-selling teams. For these teams, an extra \$500 per day is not an unusual distance to travel.

Finally, to quantify the change in mean sales on days when a threshold is attainable, we simply regress total daily sales on a dummy for whether the team’s target (T) or one of its two kinks ($1.3T$ or $1.6T$) were relevant on that day.³⁰ As shown in Table 5, sales output is \$240 (41 percent) higher on days where a target is attainable, and \$191 (33 percent) higher when $1.3T$ is attainable, than on comparable days with no attainable threshold. For the next kink at $1.6T$, the point estimate is \$258, but is statistically insignificant.

6.2 Robustness Checks

In our main analysis, we treated a threshold as attainable if it is within no more than three times the store’s mean daily sales during the previous days of the current month. Here, we focus on robustness to a different measure of mean daily sales output to calculate attainability. (Section 6.3 considers alternatives to the 3-day cutoff for attainability.) Specifically, motivated by the large day-of-week effects on sales at Firm A, Appendix C.2 defines a threshold to be attainable if it is

²⁸Please see Appendix B.3 for details.

²⁹An insignificant 1.9 percent come from more than \$500 above the target.

³⁰As noted earlier, we are less confident in these daily mean sales regressions than in our main estimates. This is because unmeasured shocks to team sales which last more than one day (such as unobserved sales and promotions) could cause the team’s target to be reached, creating bias from reverse causation. That said, Table 5 mitigates this concern by including controls for the store’s sales in each of the three days preceding the end of the month.

within three times the store’s mean daily sales during the previous same *days of the week* in the current month. Compared with the original definition, this definition includes some thresholds from high-sales days (like weekends) that were previously considered unattainable, and excludes some observations from weekdays since the attainable range is now defined more narrowly for these days. As shown in Appendix C.2, the bunching effects for T , $1.3T$, and $1.6T$ are robust to this change.

In our main analysis we maximized our sample by using all the non-last days in a month as controls for last days when a target or a kink is attainable. However, as noted earlier it is possible that forward-looking teams behave strategically in the days leading up to the end of the month. While this is not a source of concern for our main results –such strategic behavior would only provide additional support for our main hypothesis that teams tailor their effort levels to ‘just’ attain meaningful monthly sales thresholds– it is nevertheless of interest to assess whether this forward-looking type of behavior might be affecting our results. To address this issue, Table C.3.1 replicates Table 3, dropping the seven days preceding the final day of the month from the control group. The estimates are very similar, suggesting that team behavior in the week leading up to the final day is not qualitatively different from earlier days in the month. A second possible concern with our control group is that the last days of a month are so unique that no other days from the same month are good comparators. To address this possibility, Table C.3.2 replicates Table 3 using *only* last days when no threshold is attainable as controls for last days when such a threshold is attainable.³¹ Once again, the results are very similar to Table 3.

Table C.4.1 conducts a robustness check by re-defining the sales bins in our analysis. For simplicity and transparency, our main analysis used a fixed sales bin width of \$100 in all stores. In Table C.4.1, we replicate our main results (columns 3-6 of Table 3) using a bin width that is proportional to a store’s expected sales. Specifically, the width of a store’s sales bins is set at 0.01 of its monthly target T . Not only does this give us smaller bins in smaller stores, it also adjusts bin width from month to month to reflect the strong seasonality of sales. For example, bins are wider in months containing major holidays, when mean sales are higher. Despite this change, Table C.4.1’s results are very similar to Table 3’s.

6.3 Heterogeneity Analysis

Throughout our analysis so far, we have treated all thresholds within three days’ worth of sales as being within reach. However, if our sales teams truly treat the thresholds as goals, it is possible that the thresholds with the largest effects on sales are the ones that are neither too easy nor too hard to attain. To see if that is the case, Table 6 disaggregates Table 3’s thresholds into three types (using the full specification in columns 4-6):

- ‘easy’ goals (when the threshold is within 1 day of mean store sales),

³¹To economize on degrees of freedom, Table C.3.2 replaces the Bin×Store×DOW fixed effects in Table 3 by Bin×Store FEs plus Bin×Holiday fixed effects. (The Table 3 specification has essentially only 12 degrees of freedom and exhibits very high standard errors).

- ‘sweet spot’ goals (when the threshold is between 1 and 2 days of sales away), and
- ‘stretch’ goals (when the threshold is more than 2 days of sales away).

If targets and kinks function as implicit goals, then when a goal is very close, we would expect less bunching, since most teams would go beyond it even with ‘normal’ effort. Similarly, when a goal is very far away, we would also expect less bunching, since the goal may be impossible to reach for most teams. But if the goal is right in the ‘sweet spot’ (challenging but still attainable), it could be especially effective in motivating agents to reach it. Consistent with these expectations, Table 6 shows that bunching is strongest when a sales target *or* commission kink falls into the ‘sweet spot’ that is neither too easy nor too hard to attain.

Appendix D shows the results of four additional heterogeneity analyses, beginning with the effects of recent turnover events at the store. Recent turnover may affect the amount of bunching if less-intact teams find it harder to co-ordinate on a shared sales goal; this is especially relevant to Firm A because most departures occur near the end of the month (Kuhn and Yu, 2021). However, as shown in columns (1)-(2) of Table D.1.1, the likelihood of bunching at both the target and kinks is not significantly different in months when a turnover has occurred, compared to other months. In columns (3)-(4), we repeat this analysis for turnover in the *preceding* month, and we still find no significant difference in the amount of bunching at any of the thresholds.

Next, we examine whether the amount of bunching at both the target and the kinks varies with the size of the team. While we have modeled the behavior of Firm A’s sales teams as though they were individual workers, it is interesting to ask how issues like free riding and co-ordination problems might affect output bunching patterns. On the one hand, if larger teams find it hard to co-ordinate their activities to obtain a threshold public good –like the bonus at the target or a symbolic reward attached to a kink– we should see less bunching among larger teams. On the other hand teams might take advantage of these three salient thresholds as focal points to solve co-ordination and free rider problems. In this case, the utility derived from attaining a clearly defined, shared goal could create *more* bunching at all the thresholds in larger teams. As shown in Table D.2.1, bunching at both the target and the kinks is stronger in larger teams. For the target T , these differences are large in magnitude and statistically significant ($p=0.0121$ for teams of 2 versus 4 or more members); for the kinks, these estimates are large in magnitude and in the same direction, but not significant at conventional levels ($p=0.1213$ for the same comparison). Appendix Figure D.2.1 confirms these findings graphically: For all three thresholds, bunching is visibly stronger in teams of at least 4 workers. We conclude that all three thresholds might, indeed, be functioning as coordination devices for team production (Heath et al., 1999; Allen et al., 2017).

In Table D.2.2, we study effect heterogeneity with respect to a store’s typical daily sales volume, rather than the number of its employees. To do so, we first calculate the typical daily sales output of a store-month as $TDS = T/D$, where T is a store’s target in the current month, and D is the number of days in the month. Then we present separate estimates for store-months with $TDS \leq 500$ and $TDS > 500$. Consistent with the preceding finding, bunching is more pronounced in store-months

with higher levels of expected sales. Further, Table 4's finding of significant shifts in density from more than \$500 below the target and kinks is confined to these high-sales team-months. In these situations, improving sales by more than \$500 is not as 'big a lift' as in other situations.

Next, we ask whether *experienced* sales teams are more or less likely to bunch at kinks than other teams. If bunching is a behavioral bias that reduces experienced utility, experienced teams might learn *not* to bunch; this is especially the case for the two kinks because classical utility-maximizing behavior involves *avoiding* kinks. In contrast, if –as suggested above– bunching at kinks is an effective way for teams to motivate their members, experienced teams might bunch more than others. As shown in Table D.3.1, the amount of bunching at both the target and the kinks increases with average firm tenure of the sales team; in both cases the interaction terms are large and statistically significant. Once again, the evidence suggests that teams may use the kinks as goals to co-ordinate their efforts.³²

Finally, we examine whether the bunching we observe can be explained by a store manager's perceived promotion prospects: If a store manager believes that more frequently attaining a target or kink can improve her promotion chances, she may attempt to motivate her team to attain those thresholds. We offer four observations on this point. First, Firm A informs us that store manager promotions are infrequent, because most head office positions are filled via outside hires or via promotions within the head office. Second, Firm A reports that store manager promotion decisions do not consider the frequency with which a manager's team hits the sales thresholds we study. Instead, potential promotees are typically nominated by regional managers, who mainly consider a store manager's education background and communication skills. Third, even if store managers believed such a connection existed, they have very limited material leverage with which to induce individual team members to work hard to advance the *manager's* own career. Aside from fixed premia for tenure and location, all team members receive the same pay each month. Total weekly work hours are fixed, and constrained by the opening hours of the department stores and shopping malls in which Firm A's stores are located. Thus, any inducements that store managers might use would most likely be in the form of praise and peer pressure, i.e. non-tangible rewards and punishments of the sort we argue are at work in these teams.

Finally, we exploit the fact that managerial positions are only offered at the head office, and as such are not attractive to employees living far away.³³ Thus, if store managers' perceived promotion prospects account for the bunching we see in our data, that bunching should only occur in stores near the headquarters. To this end, Table D.4.1 replicates Table 4, interacting the target and kink bins with an indicator for whether the store is in the same city as Firm A's headquarters. No significant difference is found.

³²Motivated by these findings, we were able to converse with a small number of team members, who confirmed that some team leaders promote the thresholds as informal team goals for the month.

³³Among other impediments, relocation costs are not covered and wage increments are relatively small.

7 Conclusion

This paper has studied the performance of 103 small retail sales teams that face a nonlinear commission schedule that includes a target (with a bonus attached to it) plus two ‘pure’ concave kinks. We have shown that the presence of an attainable target *or* kink on a workday shifts the distribution of daily sales from below the target or kink towards higher levels. Thus, accelerating commission rates appear to be effective motivators, relative to a linear commission scheme at the pre-kink commission rate. While we cannot formally show that adding such accelerators raises Firm A’s profits, we note three factors that make this more likely. First, while adding accelerating commissions will raise compensation costs for workers who qualify for them, they are much less costly than raising the value of a linear commission rate because they only apply to sales above the kink. Second, accelerating commission rates may also have attractive selection properties for firms: they provide a simple, rules-based way to raise the pay of high-ability workers without paying other workers more. Third, at least in our context, most of the negative side-effects of nonlinear incentives, such as timing gaming and price distortion, do not apply.

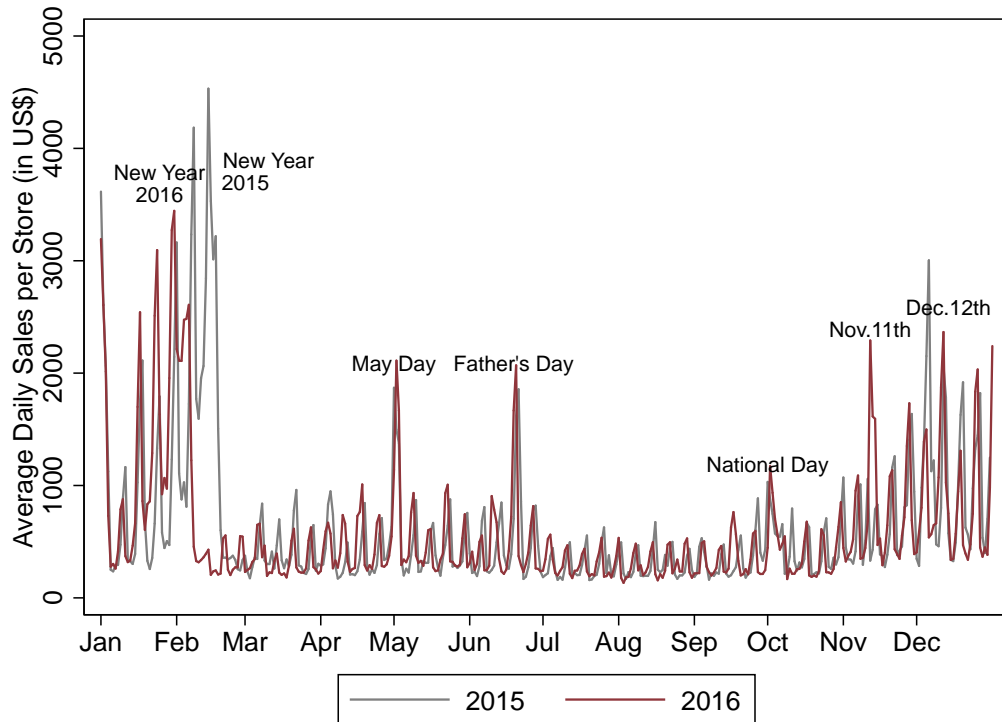
A surprising finding of our paper is the bunching of performance around the two pure concave kinks. In the paper we show that these patterns are consistent with a loss-aversion model, but only if the loss aversion associated with the kinks is sufficiently strong relative to the increase in marginal commission rates at the kink. The patterns are consistent with a symbolic rewards model regardless of the size of the symbolic reward associated with reaching the kink. In either case, we note that these psychological incentives appear to be strong enough to override the purely economic incentives that motivate teams to avoid, rather than congregate at concave kinks in any reward schedule. We also find that bunching at all three of the thresholds in teams’ pay schedules is stronger in larger, experienced teams than smaller, inexperienced ones. This suggests that the symbolic rewards or loss aversion associated with attaining different ranges of the output distribution may be best interpreted as reflecting shared goals that are used by teams to motivate and co-ordinate their members.

Finally, we note that Firm A’s nonlinear pay scheme functions as part of system of personnel policies (Boning et al., 2007) that allow Firm A to operate in the very particular environment it faces. In this environment, the sales employees are dispersed across 103 remote locations (stores), making it costly for Firm A to manage and monitor employees’ daily activities. Management does, however, observe each store’s total daily sales. Given these monitoring and informational constraints, Firm A has adopted a flexible, team-based policy that relies on peer monitoring and internal coordination to improve team production. Team-based commissions not only incentivize teams to sell more; they also reward team members for monitoring each other, for training and co-operating with each other, for selecting able new employees, and for encouraging under-performing employees to leave (Kuhn and Yu, 2021). The *target* in this commission scheme serves as an effective way to communicate management’s month-and-store-specific expected performance levels to the teams. The concave kinks, or *accelerators*, provide additional financial incentives at

relatively low costs to the firm and appear to function as symbolic rewards as well. We hope that these findings will encourage further research on team-based accelerators, and on symbolic rewards for *absolute, team-based* performance as effective motivators that lack the negative spillover effects of other reward schemes.

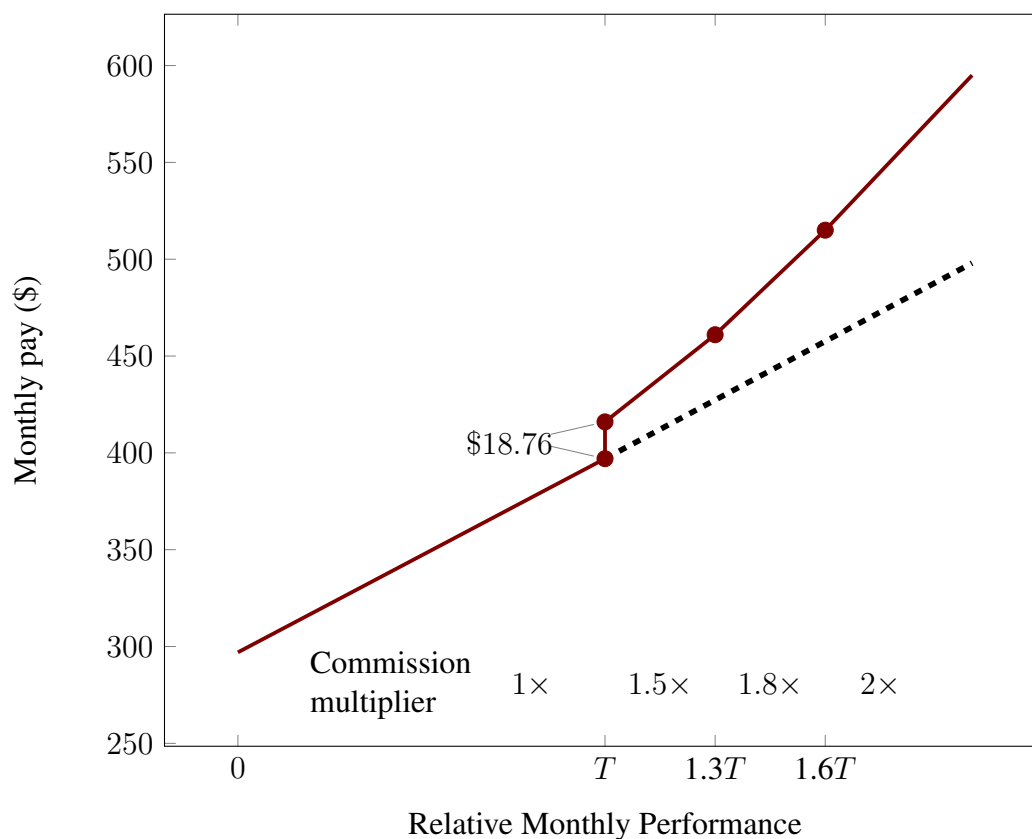
Figures and Tables

Figure 1: Average Daily Sales per Store by Calendar Day



Notes: This figure plots average daily sales per store on every calendar day in 2015 and 2016. The labelled spikes correspond to holidays or major shopping events. New Year denotes the Chinese Lunar New Year. November 11th and December 12th are the major shopping events in China, similar to Black Friday and Cyber Monday in the United States.

Figure 2: Pay Scheme



Notes:

This figure plots Firm A's pay scheme in 2016, based on the average amount of base payment and commission rate. Relative monthly performance is defined as the monthly output in 2016 relative to the monthly target T , i.e. the same store's monthly output during the same calendar month in 2015. Employees' monthly pay includes a base payment and a commission payment. Commission payment is based on the store's total monthly output, multiplied by a pre-determined baseline commission rate ($1\times$).

If the store has met its monthly goal, then each team member is rewarded with a lump sum bonus of \$15.63 (100 CNY), and is exempt from a \$3.13 (20 CNY) penalty, resulting a jump equivalent to \$18.76 (120 CNY) in the compensation schedule. On sales above T , commission rate is 50% higher than the baseline rate ($1.5\times$).

There are another two *pure kinks* in the pay schedule: on sales between $1.3T$ and $1.6T$, the commission rate is 80% higher than the baseline rate ($1.8\times$); on sales exceeding $1.6T$, the commission rate is 100% higher than the baseline rate ($2\times$).

For reference, the dashed line shows the linear schedule under the baseline commission rate.

Figure 3: A Template of Pay Slip

工资条													
		年 月 日			编号								
序号	姓名	应发			应扣		实发数						
		基本工资	绩效工资	奖金	电话补贴	加班工资	绩效扣款	缺勤天数	代扣	小计	小计		
XX	XXX	2000	1000	450	18	100				3568	-200	-200	3368

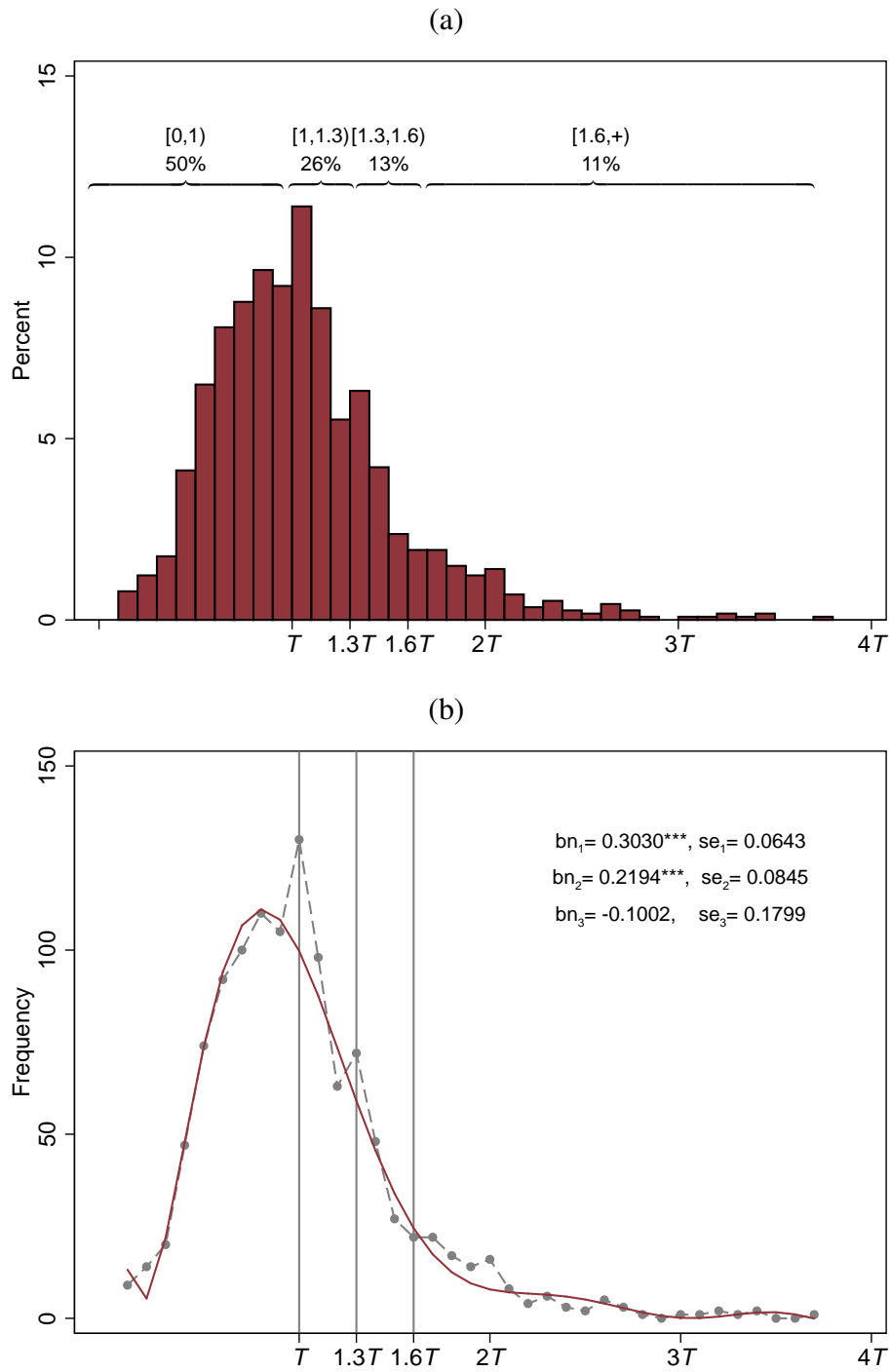
月度绩效: 优秀

Pay Slip												
		Date						No.				
#	Name	Earnings				Deductions				Net Payment		
		Base	Commission	Bonus	Cell Phone Allowance	Overtime Pay	Total	Penalty	Absenteeism		Other Deductions	Total
XX	XXX	2000	450	18	100	3568				-200	-200	3368

Monthly Assessment: Excellent

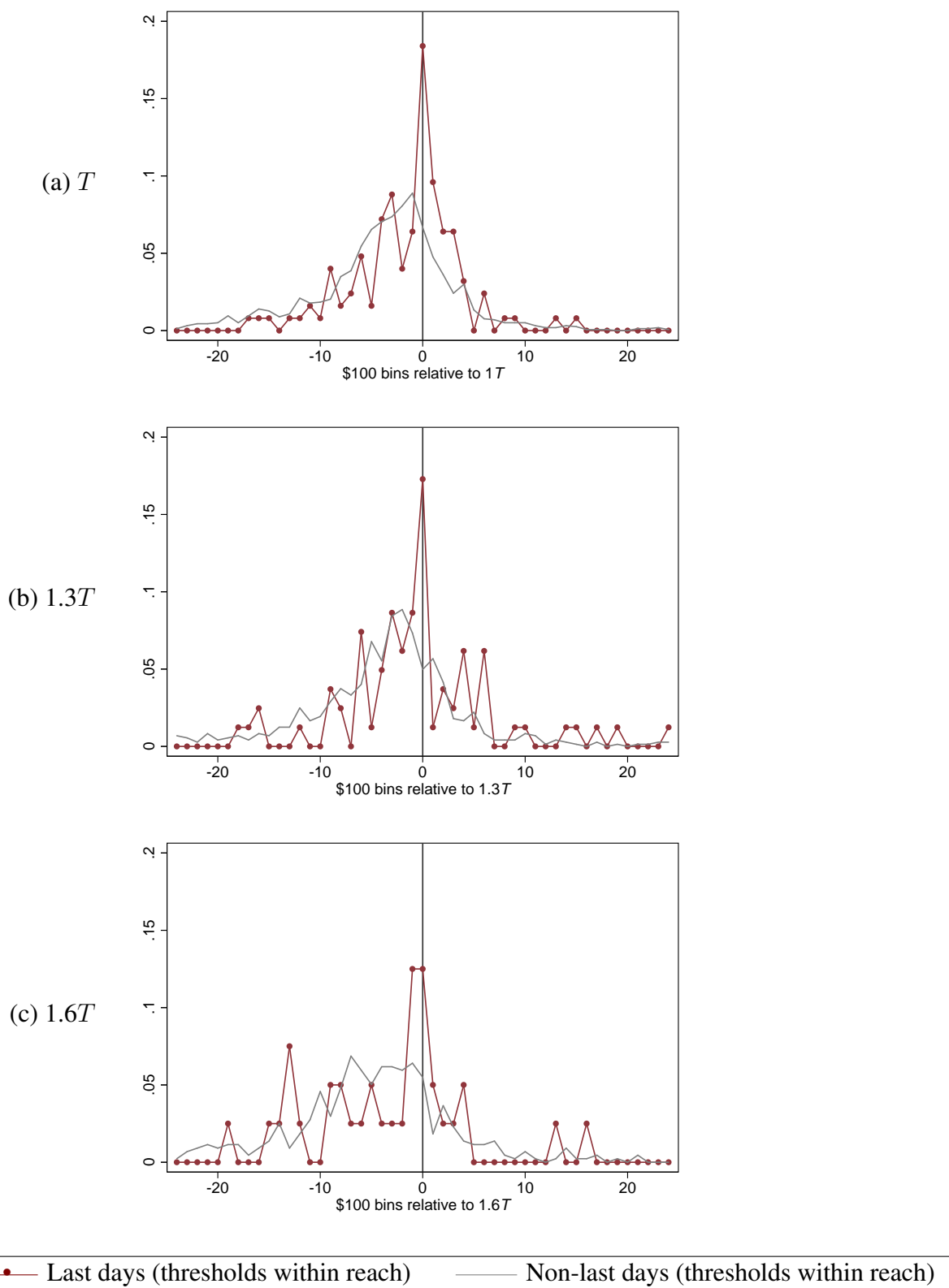
Notes: This figure presents a template of the pay slip to salespeople. The commissions are presented separately with the applicable commission rates, and the ranges are titled as "Standard", "Good", "Excellent", and "Superb". On the bottom of the pay slip, the monthly assessment takes the highest nonempty title.

Figure 4: Distribution of Monthly Sales, Relative to the Thresholds



Notes: Figure 4(a) plots the histogram of relative monthly performance, i.e. the monthly output in 2016 relative to the same store's output in 2015 during the same calendar month (i.e. T). Figure 4(b) plots the actual number of observation in each bin (connected dotted line), and the estimated number of observations in each bin (dashed curve) from a high-degree polynomial.

Figure 5: Distribution of Daily Sales, Relative to the Thresholds



Notes: These figures plot the histogram of excess (or deficient) sales relative to the corresponding kinks, measured in bins of \$100, separately for last days and non-last days. For better visual display, only the range $[-\$2,500, \$2,500]$ is presented. Figures (a), (b), and (c) in the center column are plotted for kinks T , $1.3T$, and $1.6T$, respectively. Figures for placebo kinks are presented in the Appendix Figure D.2.1.

Table 1: Descriptive Statistics

	Mean	SD	Median	N
Panel A: Product price (in US \$)	51.72	49.79	30.94	437
Accessories	16.41	9.77	20.03	20
Shirts and Polos	27.66	10.05	27.81	142
Pants	27.81	10.84	24.69	129
Sweaters	32.66	16.06	28.59	19
Jackets	101.09	34.23	82.50	91
Suits	180.28	53.94	180.94	36
Panel B: Daily sales (in US \$)	582	1241	273	34,863
Target size=2	274	368	172	6,376
Target size=3	634	1446	266	18,728
Target size=4	680	1150	382	9,759
Panel C: Relative Monthly Performance	1.07	0.56	1.00	1,143
Target size=2	0.97	0.45	0.91	209
Target size=3	1.10	0.62	1.01	614
Target size=4	1.07	0.49	1.00	320
Panel D: Monthly compensation (in US \$)	545	156	512	4,176
Target size=2	507	96	488	446
Target size=3	549	171	513	2,126
Target size \geq 4	550	145	525	1,604

Notes: Product prices presented in Panel (A) are from a sample of items sold in September, 2016; prices reflect original tag prices and discounted prices.

Target size in Panels (B) through (D) is observed from the annual sales plan, at store-year level. For newly-opened stores whose target size is not available from the annual sales plan, we use their team size 30 days after the opening as the target size for the current year. (The month of store openings are excluded from the analysis, so the first 30 days do not apply.)

In Panel (C), relative monthly performance is a store's monthly output in 2016 relative to the same store's output in 2015 in the same calendar month (i.e. T), for each store-month observations in 2016.

Monthly compensation presented in Panel (D) includes a base salary and a commission component based on team performance, along with the social security payments. Monthly compensations are missing for 6% of employee-month observations.

Table 2: Model Predictions

	Change in Mass		
	Below the Threshold	At the Threshold	Above the Threshold
A: Predictions at the ‘pure’ kinks ($1.3T$ and $1.6T$)			
1. Standard Model	–	–	+
2. Symbolic Rewards (SR) Model	–	+	+
3. Loss Aversion (LA) Model (weak loss aversion)	–	–	+
4. Loss Aversion (LA) Model (strong loss aversion)	–	+	+
B: Predictions at the target (T)			
1. Standard model	–	+	+
2. Symbolic Rewards (SR) Model	–	+	+
3. Loss Aversion (LA) Model (weak loss aversion)	–	+	+
4. Loss Aversion (LA) Model (strong loss aversion)	–	+	+

Notes:

- Changes in mass refer to the difference in total density between two situations: (a) when a target or kink is present, versus (b) when the worker faces a linear commission at the (lower) rate that prevails to the left of the threshold.
- “At the threshold” refers to the bin containing the threshold (and to immediately neighboring bins if teams can’t perfectly target). “Below” (above) the threshold refer to the regions of output space outside this interval. Thus, for example, (–,+,+) means bunching at the threshold (and possibly neighboring bins), missing mass below this bunching region, and excess mass above it.
- The shaded rows indicate that the prediction is consistent with our empirical estimates.
- Predictions are derived in Appendix A.

Table 3: Estimates of Excess Density in and around Attainable Output Thresholds

	Dependent variable: Indicator for team sales falling in a given bin on the last day of the month					
	Specification 2: FEs Controlled					
	Specification 1: Raw Estimates					
	T	$1.3T$	$1.6T$	T	$1.3T$	$1.6T$
K_{-5}	0.0481 (0.0298)	0.0340 (0.0326)	-0.0125*** (0.0012)	-0.0211 (0.0261)	-0.0362 (0.0337)	-0.0536*** (0.0125)
K_{-4}	0.0375 (0.0240)	0.0463 (0.0328)	0.0986 (0.0597)	-0.0355 (0.0239)	-0.0285 (0.0250)	0.0672 (0.0589)
K_{-3}	0.1303*** (0.0396)	0.0401 (0.0295)	0.0232 (0.0341)	0.0396 (0.0380)	-0.0331 (0.0324)	0.0010 (0.0264)
K_{-2}	0.0568** (0.0285)	0.0935** (0.0424)	0.0178 (0.0301)	-0.0290 (0.0262)	0.0077 (0.0382)	-0.0522 (0.0402)
K_{-1}	0.0321 (0.0329)	0.0675** (0.0424)	0.0415 (0.0740)	-0.0469** (0.0184)	0.0096 (0.0291)	-0.0229 (0.0274)
K_0	0.1707*** (0.0329)	0.1923*** (0.0424)	0.1980*** (0.0740)	0.0919*** (0.0328)	0.1143** (0.0439)	0.1494** (0.0665)
K_1	0.0944*** (0.0259)	0.0357 (0.0237)	0.0138 (0.0255)	0.0394 (0.0270)	-0.0069 (0.0228)	-0.0202 (0.0255)
K_2	0.0638** (0.0262)	0.0116 (0.0167)	0.0401 (0.0370)	0.0274 (0.0250)	-0.0239 (0.0160)	0.0134 (0.0381)
K_3	0.0333* (0.0178)	0.0116 (0.0167)	0.0138 (0.0254)	0.0067 (0.0158)	-0.0116 (0.0164)	0.0008 (0.0235)
K_4	0.0486** (0.0208)	0.0357 (0.0228)	0.0401 (0.0349)	0.0310 (0.0203)	0.0208 (0.0228)	0.0352 (0.0349)
K_5	0.0028 (0.0109)	0.0116 (0.0165)	-0.0125*** (0.0012)	-0.0023 (0.0107)	0.0051 (0.0172)	-0.0187*** (0.0048)
N	2697816	2693899	2691619	2697816	2693899	2691619

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

Notes: Columns (1)-(3) report estimates of equation (1), which controls for store fixed effects. Columns (4)-(6) report estimates of equation (2), which controls for Bin \times Store \times DOW and Bin \times Holiday fixed effects. Standard errors reported in parentheses are clustered at the store level. Each column represents a different regression, and the sample in each column comprises the control sample plus the treated sample for the threshold in question (T , $1.3T$, or $1.6T$). An observation is a bin \times store \times day cell. K_0 is an indicator variable, identifying the bin that exactly contains the threshold; the remaining ' K ' variables indicate the five bins immediately below and above the threshold.

Table 4: Estimates of Excess Density below, at, and above Attainable Thresholds

	Dependent variable: Indicator for team sales falling in a given bin on the last day of the month		
	T	$1.3T$	$1.6T$
$K_{\leq -6}$	-0.1033*** (0.0276)	-0.1413*** (0.0294)	-0.1613*** (0.0602)
$K_{-5,-1}$	-0.0690** (0.0344)	-0.0447 (0.0426)	-0.0548 (0.0624)
K_0	0.0919*** (0.0328)	0.1143** (0.0439)	0.1494** (0.0665)
$K_{1,5}$	0.1020*** (0.0385)	-0.0170 (0.0415)	0.0099 (0.0537)
$K_{\geq 6}$	-0.0192 (0.0192)	0.0844** (0.0422)	0.0576 (0.0740)
Bin×Store×DOW	Yes	Yes	Yes
Bin×Holiday	Yes	Yes	Yes
N	2697816	2693899	2691619

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

Notes: This Table reports estimates of equation (3), with controls for Bin×Store×DOW and Bin×Holiday fixed effects. Standard errors reported in parentheses are clustered at the store level. Each column represents a different regression, and the sample in each column comprises the control sample plus the treated sample for the threshold in question (T , $1.3T$, or $1.6T$). An observation is a bin × store × day cell. K_0 is an indicator variable, identifying the bin that exactly contains the target or the kink. $K_{-5,-1}$ identifies the five bins *just* below the threshold, and $K_{1,5}$ identifies the five bins *just* above it. $K_{\leq -6}$ identifies all the bins that are below the threshold and more than five bins away, while $K_{\geq 6}$ identifies all the bins more than five bins above the threshold.

Table 5: Daily Mean Sales Regressions

	Dependent variable: Daily Sales (in US\$)		
	T	$1.3T$	$1.6T$
i. (Last day w/ attainable threshold)	239.7** (100.9)	191.3** (91.0)	257.7 (336.4)
Store×DOW	Yes	Yes	Yes
Holiday	Yes	Yes	Yes
N	34863	34863	34863

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

Notes: The dependent variable is the daily sales output; an observation is a store × day cell. The regressor of interest is an indicator variable, taking a value of 1 if the current day is the last day of a month and contains an attainable threshold of the type indicated (T , $1.3T$, or $1.6T$). All regressions control for the store's sales in each of the three days preceding the end of the month, Store×DOW fixed effects, and Holiday fixed effects. Standard errors reported in parentheses are clustered at the store level.

Table 6: Heterogeneity Examination – Easy Goals, ‘Stretch’ Goals, and Goals in the ‘Sweet Spot’

	Dependent variable: Indicator for team sales falling in a given bin on the last day of the month								
	Easy [0, 1)		Sweet [1, 2)		Stretch [2, 3)				
	T	1.3 T	1.6 T	T	1.3 T	1.6 T			
K_{-5}	-0.0660*** (0.0247)	-0.1516** (0.0699)	-0.1366*** (0.0277)	-0.0404 (0.0356)	-0.0224 (0.0625)	-0.0509** (0.0222)	-0.0008 (0.0415)	-0.0256 (0.0435)	-0.0314** (0.0122)
K_{-4}	-0.0651** (0.0258)	-0.0759** (0.0332)	-0.0999** (0.0445)	-0.0721** (0.0277)	-0.0249 (0.0569)	0.0610 (0.0906)	0.0013 (0.0436)	-0.0202 (0.0295)	0.1189 (0.0962)
K_{-3}	0.0049 (0.0716)	-0.0878*** (0.0271)	0.1425 (0.1737)	0.0216 (0.0647)	-0.0876*** (0.0200)	-0.0318*** (0.0101)	0.0663 (0.0486)	0.0234 (0.0650)	-0.0161* (0.0095)
K_{-2}	-0.0819* (0.0467)	-0.1145*** (0.0258)	-0.2128*** (0.0569)	-0.0082 (0.0552)	-0.0055 (0.0536)	0.0224 (0.0976)	-0.0253 (0.0222)	0.0736 (0.0585)	-0.0191*** (0.0060)
K_{-1}	-0.0741* (0.0434)	0.0031 (0.0682)	0.0148 (0.0847)	-0.0678** (0.0275)	0.0200 (0.0532)	-0.0603** (0.0270)	-0.0066 (0.0260)	0.0059 (0.0336)	-0.0258** (0.0118)
K_0	0.0735 (0.0653)	0.0939 (0.0853)	0.1824 (0.1127)	0.1188** (0.0597)	0.1882** (0.0843)	0.1410* (0.0829)	0.0848* (0.0447)	0.0740 (0.0582)	0.1255 (0.0920)
K_1	-0.0067 (0.0492)	-0.0391 (0.0465)	-0.0099 (0.0704)	0.1403** (0.0611)	0.0076 (0.0414)	-0.0372*** (0.0136)	-0.0137*** (0.0026)	0.0144 (0.0345)	-0.0164* (0.0093)
K_2	0.0455 (0.0611)	-0.0364 (0.0442)	0.0200 (0.0764)	0.0189 (0.0332)	-0.0331*** (0.0066)	0.0504 (0.0915)	0.0156 (0.0239)	-0.0035** (0.0017)	-0.0217** (0.0106)
K_3	0.0270 (0.0369)	-0.0038 (0.0444)	-0.0451*** (0.0104)	0.0027 (0.0235)	-0.0231*** (0.0068)	0.0690 (0.0806)	-0.0122*** (0.0028)	-0.0100*** (0.0030)	-0.0103* (0.0059)
K_4	0.0231 (0.0370)	0.0743 (0.0627)	0.1167 (0.1023)	0.0550 (0.0393)	-0.0141*** (0.0044)	-0.0154** (0.0076)	0.0151 (0.0237)	-0.0059** (0.0027)	-0.0007 (0.0014)
K_5	-0.0122 (0.0228)	-0.0046 (0.0337)	-0.0337*** (0.0105)	-0.0117*** (0.0025)	0.0265 (0.0405)	-0.0115** (0.0046)	0.0187 (0.0232)	-0.0025 (0.0016)	-0.0098 (0.0063)
N	2691267	2689892	2689126	2690917	2689934	2689018	2691086	2689527	2688929

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

Notes: This Table replicates columns (4)-(6) of Table 3 for three different types of thresholds—easy, ‘sweet spot’, and ‘stretch’. Easy goals refer to thresholds that are no more than 1 day of mean store sales; ‘sweet spot’ goals are when the threshold is between 1 and 2 days of mean store sales; and ‘stretch’ goals are when the threshold is more than 2 days of sales away. Standard errors reported in parentheses are clustered at the store level. Please see notes to Table 3 for additional details.

References

- Abeler, J., Falk, A., Goette, L., and Huffman, D. (2011). Reference points and effort provision. *American Economic Review*, 101(2):470–92.
- Adamson, B., Dixon, M., and Toman, N. (2014). Why individuals no longer rule on sales teams. *Harvard Business Review*.
- Agarwal, S., Fang, X., Goette, L., Sing, T. F. S., Schoeb, S. S., Tiefenbeck, V., Staake, T. S., and Wang, D. (2017). The role of goals in motivating behavior: Evidence from a large-scale field experiment on resource conservation. *unpublished manuscript, University of Bonn*.
- Allen, E. J., Dechow, P. M., Pope, D. G., and Wu, G. (2017). Reference-dependent preferences: Evidence from marathon runners. *Management Science*, 63(6):1657–1672.
- Ashraf, N., Bandiera, O., and Lee, S. S. (2014). Awards unbundled: Evidence from a natural field experiment. *Journal of Economic Behavior and Organization*, 100:44–63.
- Basu, A. K., Lal, R., Srinivasan, V., and Staelin, R. (1985). Salesforce compensation plans: An agency theoretic perspective. *Marketing Science*, 4(4):267–291.
- Benson, A. (2015). Do agents game their agents' behavior? Evidence from sales managers. *Journal of Labor Economics*, 33(4):863–890.
- Berg, P., Appelbaum, E., Bailey, T., and Kalleberg, A. L. (1996). The performance effects of modular production in the apparel industry. *Industrial Relations: A Journal of Economy and Society*, 35(3):356–373.
- Boning, B., Ichniowski, C., and Shaw, K. (2007). Opportunity counts: Teams and the effectiveness of production incentives. *Journal of Labor Economics*, 25(4):613–650.
- Cengiz, D., Dube, A., Lindner, A., and Zipperer, B. (2019). The effect of minimum wages on low-wage jobs. *Quarterly Journal of Economics*, 134(3):1405–1454.
- Charness, G., Kuhn, P., and Villeval, M. C. (2011). Competition and the ratchet effect. *Journal of Labor Economics*, 29(3):513–547.
- Chetty, R., Friedman, J. N., Olsen, T., and Pistaferri, L. (2011). Adjustment costs, firm responses, and micro vs. macro labor supply elasticities: Evidence from Danish tax records. *Quarterly Journal of Economics*, 126(2):749–804.
- Cohen, A. (2008). Package size and price discrimination in the paper towel market. *International Journal of Industrial Organization*, 26(2):502–516.

- Dobronyi, C. R., Oreopoulos, P., and Petronijevic, U. (2019). Goal setting, academic reminders, and college success: A large-scale field experiment. *Journal of Research on Educational Effectiveness*, 12(1):38–66.
- Douthitt (1976). The use of sales quotas by industrial firms. *Journal of the Academy of Marketing Science*, 4:467–472.
- Freeman, R. B., Huang, W., and Li, T. (2019). Non-linear incentives and worker productivity and earnings: Evidence from a quasi-experiment. *National Bureau of Economic Research, working paper no 25507*.
- Friebel, G., Heinz, M., Krüger, M., and Zubanov, N. (2017). Team incentives and performance: Evidence from a retail chain. *American Economic Review*, 107(8):2168–2203.
- Gu, Z. and Yang, S. (2010). Quantity-discount-dependent consumer preferences and competitive nonlinear pricing. *Marketing Science*, 47(6):1100–1113.
- Gubler, T., Larkin, I., and Pierce, L. (2016). Motivational spillovers from awards: Crowding out in a multitasking environment. *Organization Science*, 27(2):286–303.
- Hamilton, B. H., Nickerson, J. A., and Owan, H. (2003). Team incentives and worker heterogeneity: An empirical analysis of the impact of teams on productivity and participation. *Journal of Political Economy*, 111(3):465–497.
- Heath, C., Larrick, R. P., and Wu, G. (1999). Goals as reference points. *Cognitive Psychology*, 38(1):79–109.
- Holmstrom, B. and Milgrom, P. (1987). Aggregation and linearity in the provision of intertemporal incentives. *Econometrica*, pages 303–328.
- Joseph, K. and Kalwani, M. (1998). The role of bonus pay in salesforce compensation plans. *Industrial Marketing Management*, 27:147–159.
- Kaur, S., Kremer, M., and Mullainathan, S. (2015). Self-control at work. *Journal of Political Economy*, 123(6):1227–1277.
- Kleven, H. J. (2016). Bunching. *Annu. Rev. Econ*, 8:435–64.
- Kleven, H. J. and Waseem, M. (2013). Using notches to uncover optimization frictions and structural elasticities: Theory and evidence from Pakistan. *Quarterly Journal of Economics*, 128(2):669–723.
- Knez, M. and Simester, D. (2001). Firm-wide incentives and mutual monitoring at Continental Airlines. *Journal of Labor Economics*, 19(4):743–772.

- Kosfeld, M. and Neckermann, S. (2011). Getting more work for nothing? Symbolic awards and worker performance. *American Economic Journal: Microeconomics*, 3:1–16.
- Kőszegi, B. and Rabin, M. (2006). A model of reference-dependent preferences. *The Quarterly Journal of Economics*, 121(4):1133–1165.
- Kuhn, P. J. and Yu, L. (2021). How costly is turnover? evidence from retail. *Journal of Labor Economics*.
- Lambrecht, A., Seim, K., and Skiera, B. (2007). Does uncertainty matter? Consumer behavior under three part tariffs. *Journal of Marketing Research*, 26(5):698–710.
- Larkin, I. (2014). The cost of high-powered incentives: Employee gaming in enterprise software sales. *Journal of Labor Economics*, 32(2):199–227.
- Lawler, E. and Mohrman, S. A. (2003). Pay practices in Fortune 1000 corporations. *WorldatWork Journal*, 12(4):45–54.
- Lazear, E. P. (2000). Performance pay and productivity. *American Economic Review*, 90(5):1346–1361.
- Mortenson, J. A. and Whitten, A. (2019). Bunching to maximize tax credits: Evidence from kinks in the US tax schedule. *American Economic Journal: Economic Policy*, 12(3):402–432.
- MyFirstSalesJob (2020). How to Structure Sales Commission Plans.
- Ockenfels, A., Sliwka, D., and Werner, P. (2015). Bonus payments and reference point violations. *Management Science*, 61(7):1496–1513.
- O’Connell, W. A. (1989). *Salesforce Compensation, Dartnell’s 25th Survey*. Dartnell Corporation.
- Oettinger, G. S. (2002). The effect of nonlinear incentives on performance: Evidence from Econ 101. *Review of Economics and Statistics*, 84(3):509–517.
- Owan, H., Tsuru, T., and Uehara, K. (2015). Incentives and gaming in a nonlinear compensation scheme. In *Evidence-based HRM: A Global Forum for Empirical Scholarship*. Emerald Group Publishing Limited.
- Oyer, P. (2000). A theory of sales quotas with limited liability and rent sharing. *Journal of Labor Economics*, 18:405–426.
- Parrinello, R. (2017). Using accelerators to drive winning sales teams. *Workspan*, (August):32–39.
- Pierce, L., Rees-Jones, A., and Blank, C. (2020). The negative consequences of loss-framed performance incentives. *NBER working paper*, (26619).

- Royer, H., Stehr, M., and Sydnor, J. (2015). Incentives, commitments, and habit formation in exercise: Evidence from a field experiment with workers at a Fortune 500 company. *American Economic Journal: Applied Economics*, 7(3):51–84.
- Saez, E. (2010). Do taxpayers bunch at kink points? *American Economic Journal: Economic Policy*, 2(3):180–212.
- Scudder, M. D. (2017). Will the DOL rescind the tip pool rule? *The National Law Review*.
- Seibold, A. (2021). Reference points for retirement behavior: Evidence from German pension discontinuities. *American Economic Review*, page forthcoming.

Appendices

A Appendix A: Theory

This Appendix derives the predictions of three models for the effects of adding two forms of non-linearity to a previously-linear cash incentive scheme. The first form of non-linearity is a concave *kink* at which only the commission rate increases, like those at 1.3T and 1.6T in Firm A. The second form of non-linearity occurs at Firm A's *target* (T) where an increase in the commission rate is combined with a small bonus. The **standard model** assumes workers derive utility only from money and leisure, and have quasi-linear utility in money. In the **symbolic rewards (SR)** model, workers derive a lump sum psychic reward from attaining the threshold. In the **loss aversion (LA)** model (based on [Kőszegi and Rabin \(2006\)](#)), the threshold creates a downward discontinuity in the marginal utility of money. In all cases our models treat teams as unitary decisionmakers, maximizing the utility of a representative worker. Thus, while we empirically estimate the effects of factors that might exacerbate free-riding and co-ordination failures –like team size and experience– our theoretical analyses abstract from these issues.

We begin by formally stating the assumptions of each model, then we derive the three models' predictions for the *kinks*. We conclude by deriving the predictions for the *target*, T. All the predictions are summarized in Table 2 of the paper.

A.1 Assumptions

All three models are special cases of the following framework. The utility of a team's representative worker is:

$$u(y, q) = v(y) - \frac{c(q)}{\psi}, \quad (\text{A.1})$$

where y is income and q is output. Marginal costs of producing output, $\frac{c'(q)}{\psi}$ are increasing, convex and differentiable everywhere; $\psi > 0$ represents heterogeneity in these costs. While teams at Firm A face two important sources of cost heterogeneity –the mean ability of their members and the current day's level of customer traffic– to simplify the discussion throughout this Appendix we adopt the more familiar framing where higher levels of ψ represent higher-ability 'workers'. Utility derived from income, or *effective income*, is given by $v(y)$. Working in terms of effective income allows us to express symbolic rewards and loss aversion as equivalent to the following changes to the commission rate, relative to a threshold output level, y^t :

- the lump-sum symbolic award is equivalent to a cash bonus at the threshold.
- loss aversion is equivalent to a convex kink in the commission schedule at the threshold.

In more detail, the **standard model** assumes that:

$$v(y) = y \quad (\text{A.2})$$

The **symbolic rewards** model assumes that:

$$v(y) = \begin{cases} y & \text{if } y < y^t \\ y + B & \text{if } y \geq y^t \end{cases} \quad (\text{A.3})$$

where y^t is the worker's income at the threshold and $B > 0$.

Finally, the **loss aversion** model assumes that:

$$v(y) = \begin{cases} y + \eta\lambda \cdot (y - y^t) & \text{if } y < y^t \\ y + \eta \cdot (y - y^t) & \text{if } y \geq y^t \end{cases} \quad (\text{A.4})$$

where $\eta \geq 0$ represents the strength of *comparison* utility (money relative to the threshold level) and $\lambda \geq 1$ represents loss aversion– the additional marginal utility of money when the team is below the threshold.³⁴

Notably, all our models ignore the ‘lumpy’ and uncertain nature of sales at Firm A, which significantly constrain a team’s ability to exactly hit the sales bin that contains a salient threshold. To accommodate this uncertainty, Table 2 focuses on our models’ predictions for three fairly coarse outcomes: the amount of probability mass in a small interval around the threshold (“bunching”), and the total amount of probability masses below and above that interval.

A.2 Predictions for the Kinks (1.3T and 1.6T)

For all three models, this Section compares the predicted behavior of workers under two conditions. In the first condition, workers face a linear baseline commission rate:

$$y = y^K + w_1(q - q^K), \forall q \quad (\text{A.5})$$

In the second, we add an accelerator to this schedule, as follows:

$$y = \begin{cases} y^K + w_1(q - q^K) & \text{if } q < q^K \\ y^K + w_2(q - q^K) & \text{if } q \geq q^K \end{cases} \quad (\text{A.6})$$

³⁴When η is strictly greater than zero, [Kőszegi and Rabin \(2006\)](#)’s model has the property that introducing a reference point such as a label of “good” performance raises the *marginal* utility of money above that reference point. Since it is not obvious to us that this should be the case –and for simplicity– our formal analysis will focus mainly on the $\eta = 0$ case. The $\eta > 0$ case is then a straightforward extension.

where $w_1 < w_2$, and y^K and q^K are the income and output levels at the kink in the budget constraint.

For the baseline (unkinked) budget constraint in the standard model, there is a one-to-one mapping between the ability parameter, ψ , and the worker's optimal sales output, q . If the density of ability, $f(\psi)$ is strictly positive and continuous on $(0, \infty)$, this yields a continuous density of observed output levels. Following Saez (2010), we can then derive predictions for the effect of introducing a concave kink at q^k on the distribution of output levels.

A.2.1 The Standard Model

In the standard model, $v(y) = y$. In effective income terms, the budget constraint and the indifference curve of a (who is indifferent between the two segments of the budget constraint) are shown below:

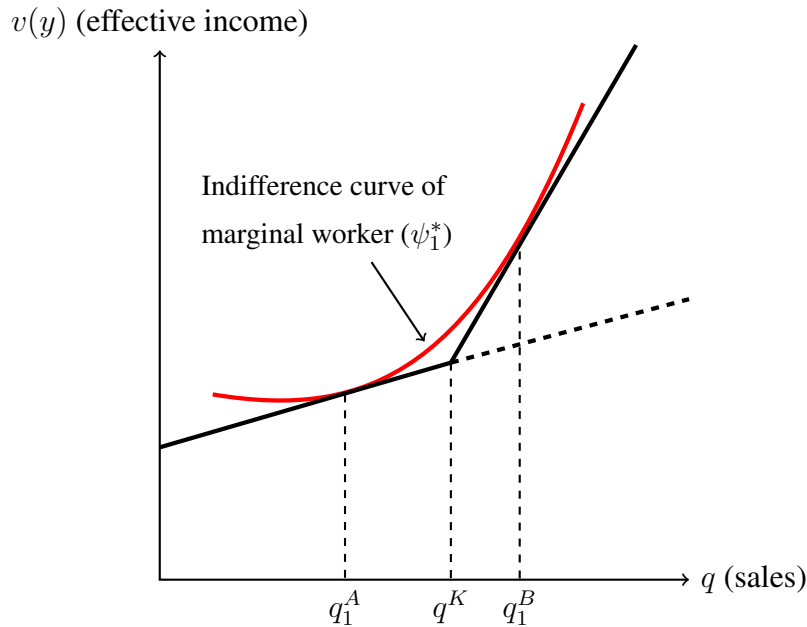


Figure A.1: The Standard Model

In this model, workers with ability (ψ), below ψ_1^* will sell less than q_1^A in the absence of an accelerator, and will not change their output (q) when the accelerator is introduced. Workers with ability above ψ_1^* will raise their (q) to a level above q_1^B . Finally, no workers will locate between q_1^A and q_1^B .

Adding an accelerator therefore:

- leaves the density of sales to the left of q_1^A unchanged;
- reduces the density of sales (to zero) in a region that (strictly) includes the kink (between q_1^A and q_1^B);

- raises the number of workers producing strictly more than the kink (q^K);
- no mass point (bunching) is induced.

A.2.2 The Symbolic Rewards Model

In the symbolic rewards model, $v(y) = y$ if $y < y^K$, and $v(y) = y + B$ if $y \geq y^K$. Now, the effective-income budget constraint and marginal indifference curve take the following form:

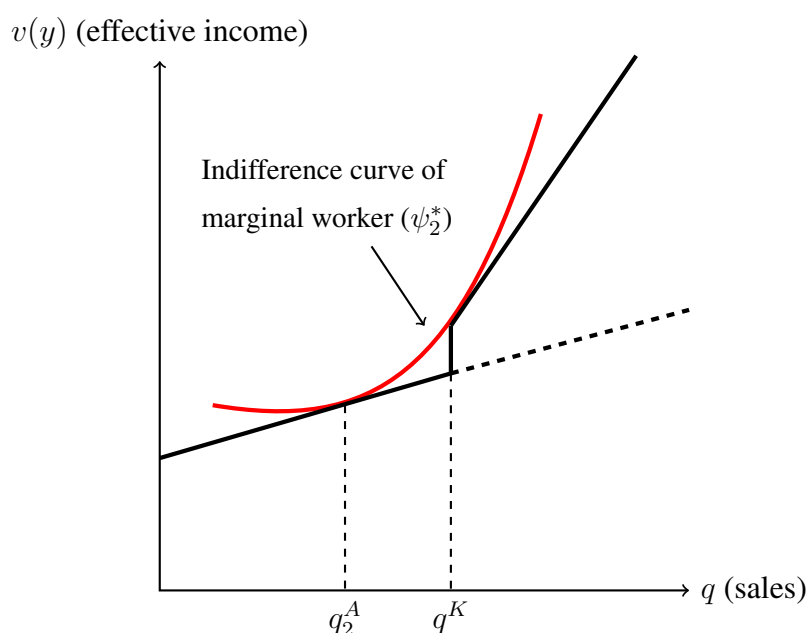


Figure A.2: The Symbolic Rewards Model

Now, workers with ability, ψ , below ψ_2^* will sell less than q_2^A in the absence of an accelerator, and will not change their output (q) when the accelerator is introduced. Because the symbolic reward makes it more attractive to reach the kink, $q_2^A < q^A$. Some of the workers with ability, ψ , above ψ_2^* will raise their output (q) to exactly q^K . These are the workers with ability in some interval (ψ_2^*, ψ_2') , where ψ_2' corresponds to tangency (from the right) at q^K .³⁵

The remaining workers (with $\psi > \psi_2'$) will raise their output levels to a point strictly above q^K . No workers will locate between q_2^A and q^K .

Adding an accelerator therefore:

- leaves the density of sales to the left of q_2^A unchanged;
- reduces the density of sales (to zero) *below* the kink (between q_2^A and q^K);
- raises the number of workers producing strictly more than the kink (q^K);

³⁵The indifference curve of the marginal worker will in general be steeper than the budget constraint at q^K ; with probability measure zero it will be tangent there, in which case there is no bunching.

- creates a mass point (bunching) at q^K .

A.2.3 The Loss Aversion Model

In this model, $v(y) = y + \eta\lambda(y - y^K)$ if $y < y^K$, and $v(y) = y + \eta(y - y^K)$ if $y \geq y^K$. Here, the effects of introducing an accelerator depend on the relative size of the loss-aversion parameter (λ) and the jump in the commission rate ($\frac{w_1}{w_2}$) at q^K . There are two cases:

1. *Weak Loss Aversion* ($\lambda < \frac{w_1}{w_2}$):

When loss aversion is weak relative to the marginal wage increase at the kink, the marginal effective-income gradient to the left of the kink ($\frac{dv}{dq}$) remains lower than the gradient to the right of the kink. Thus, in effective income terms, the worker still faces a concave kink at q^K . Therefore, as in the standard model, the accelerator should create a *gap* in the density of sales around (i.e. on both sides of) the kink point. No bunching of sales should be observed.

2. *Strong Loss Aversion* ($\lambda > \frac{w_1}{w_2}$):

Now the effective-income gradient falls at the kink, creating a convex kink (in effective income terms):

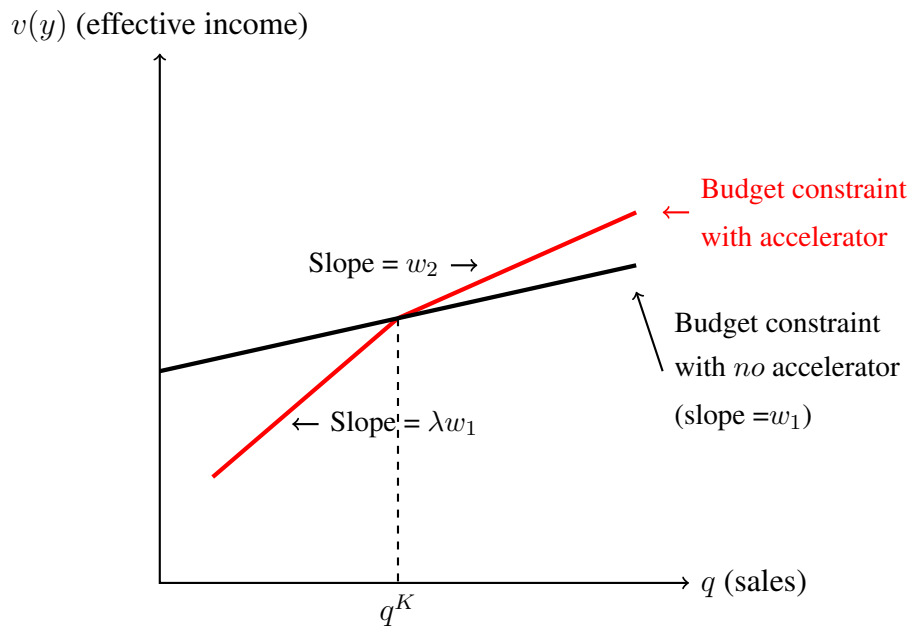


Figure A.3: The Loss-Aversion Model with Strong Loss Aversion and $\eta = 0$

Figure A.3 illustrates the strong loss-aversion case when $\eta = 0$. When $\eta > 0$, the budget constraint in the presence of the kink will continue to exhibit the same downward jump in slope at q^K (because $\lambda > \frac{w_1}{w_2}$), but the entire constraint rotates counter-clockwise around the kink. Thus, regardless of η there will be bunching at the kink. Relative to the case where no accelerator is

present, there will be a decline in density to the left of the kink (because some teams are ‘pulled up’ to the kink), and (b) an increase in total density to its right (because marginal cash incentives are higher than before).

Adding an accelerator therefore:

- creates a mass point (bunching) at q^K ;
- reduces the number of teams producing strictly less than q^K ;
- increases the number of teams producing strictly more than q^K .

Combining cases 1 and 2, loss aversion can cause bunching at the kink **only if** it is strong enough, relative to the increase in marginal cash incentives at the kink. If that is the case, adding an accelerator has similar predicted effect as in the symbolic rewards model.

A.3 Predictions for the Target (T)

With the preceding tools in hand, it is straightforward to derive the three models’ predictions for the output distribution around the target, T , where an accelerator is combined with a cash bonus. To see this, note first that the standard model is now formally equivalent to the symbolic rewards (SR) model, with the cash bonus now playing the same role as the SR model’s *psychic* bonus. Thus, we expect bunching at the target and excess mass to its right. By the same reasoning, the qualitative predictions of the symbolic rewards model are the same at the target as they were at the kinks– the only difference between the target and kinks is that the bonus at the target includes a cash component in addition to the symbolic one. Finally, the presence of a cash bonus means that, in general, the loss aversion (LA) model –like the standard and SR models– *also* predicts some bunching at the target. Extending previous reasoning, this bunching will be accompanied by excess mass to the right of the target when loss aversion is weak, and missing mass if loss aversion is strong.

B Appendix B: Methods

B.1 Statistical Significance of Bunching in the Monthly Sales Distributions

To estimate the significance of bunching in Figure 4(b), we estimate a regression of the following form:

$$c_j = \sum_{i=0}^p \beta_i \cdot (z_j)^i + \sum_{i=1,1.3,1.6} \gamma_i \cdot \mathbb{1}[z_j = i] + \epsilon_j, \quad (\text{B.1})$$

where c_j is the number of store-month observations in bin j , z_j is the level of relative performance in bin j ($z_j = 0.1, 0.2, \dots, 4$), and p is the order of polynomial. Then we use the predicted values from the above regression to construct the counterfactual bin counts, i.e. $\hat{c}_j = \sum_{i=0}^p \hat{\beta}_i \cdot (z_j)^i$. The bunching is then estimated as the excess mass relative to the counterfactual bin counts, i.e. $b_j = c_j - \hat{c}_j = \hat{\gamma}_j$, where $j=1, 1.3$ and 1.6 in our context. The solid curve in Figure 4(b) shows the counterfactual distribution with an eighth-degree polynomial.

B.2 Calculating Baseline Densities in the Output Bins that ‘Typically’ Contain the Target and Kinks

Here we describe how we calculate the baseline densities in the output that ‘typically’ contain the target and two kinks on control days. These are used in Section 5.1 to assess the magnitude of our main estimated coefficients.

We start by computing the mean density in a target’s or kink’s bin at each store i , separately for T , $1.3T$, and $1.6T$. For instance, the vector for T at store i , ω_i^T , reflects the likelihood that each bin contains T , i.e. $\omega_i^T = (\omega_{i,1}^T, \dots, \omega_{i,N_i}^T)$, where N_i is the number of bins for store i . Then, we restrict to the control sample to compute the mean density of every bin in which store i ’s sales output falls in, and call it $\mathbf{f}_i = (f_{i,1}, \dots, f_{i,N_i})$. To compute the likelihood that store i ’s sales output falls in the bin that contains T , we compute the mean of $f_{i,b}$ across its bins, weighted by $\omega_{i,b}$, i.e. $F_i^T = \omega_i^T \times \mathbf{f}_i'$. Thus, F_i^T reflects the likelihood that store i ’s output falls in the bins where T typically occurs. Finally, we take the sum of F_i^T across all stores to get the baseline density, weighted by the number of observations contributed by each store in the overall sample. This baseline density is estimated to be 10.3 percentage points for the target T . In other words on days when *no* target or kink is attainable, the mix of bins in which the target would typically fall contain 10.3% of the density of store sales. The baseline densities for the two pure kinks $1.3T$ and $1.6T$ are defined analogously and equal 7.7 percentage points for $1.3T$, and 5.9 percentage points for $1.6T$.

B.3 Rescaling Estimates for Indicator Variables that Include Multiple Bins

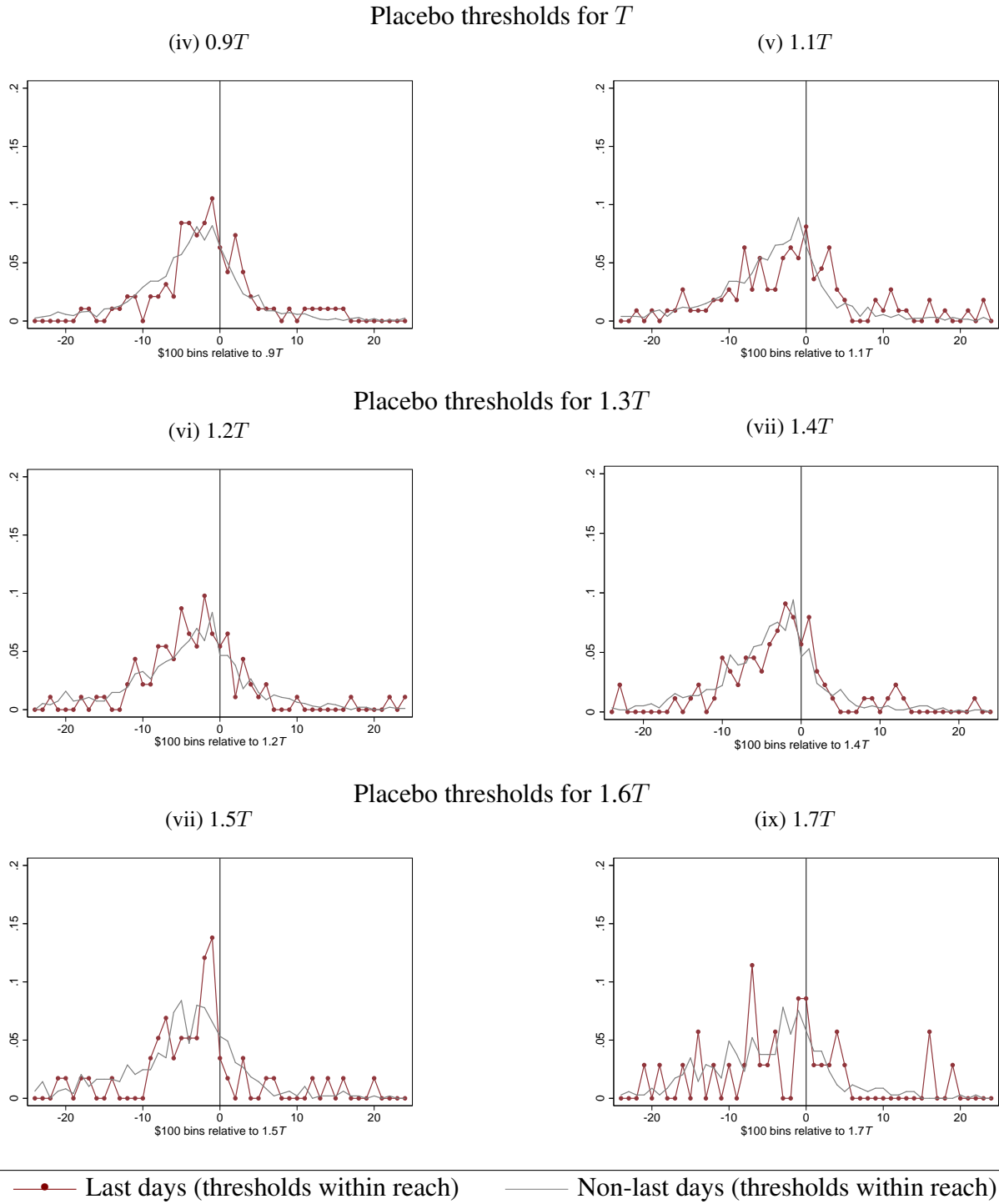
Here we describe how we rescale estimates for indicator coefficients that include multiple bins. These are used in Section 6.1 to identify where the additional density at thresholds is coming from.

The method is illustrated using the threshold T , and we repeat this practice analogously for the thresholds $1.3T$ and $1.6T$. First, we count the total number of bins identified by T , and the number of bins identified by each coefficient. For $K_{<-6}$, this summarizes the number of bins that are more than 5 bins away below the threshold's bin. Using the total number of bins included in T , we calculate the share of bins in the specified range for each coefficient. Then we adjust these shares by comparing them with the share of bins specified by K_0 . This gives us the representation scale for each coefficient, adjusted by the representation of the threshold's bin. We then use these scales to multiple the corresponding coefficients and standard errors, to retain the significance levels.

C Appendix C: Robustness

C.1 Bunching Relative to Placebo Thresholds

Figure C.1:1 Excess (or Deficient) Sales Relative to the threshold (Placebo Thresholds)



Notes: The sample for these figures is restricted to days on which a placebo threshold (e.g. $0.9T$) is within reach. Sales densities are shown separately for last days of the month (in red) and non-last days (in gray). The horizontal axis shows \$100 bins relative to the placebo threshold. For better visual display, only the range $[-\$2,500, \$2,500]$ is presented. Corresponding Figures for the *actual* thresholds are shown in Figure 5 of the paper.

C.2 Alternative Definition of Attainability

Our main definition of attainability uses a store's daily mean sales on all previous days in the current month to determine whether a threshold is attainable on a given day. Here we use the store's daily mean sales on the previous same *days of the week* in the current month to calculate attainability. The results are very similar. Please see Table 5 for the effects of alternative attainability *cutoffs* on the results.

Table C.2.1: Robustness Check – Defining Attainability

	Dependent variable: Indicator for team sales falling in a given bin on the last day of the month		
	<i>T</i>	1.3 <i>T</i>	1.6 <i>T</i>
K_{-5}	-0.0266 (0.0205)	-0.0017 (0.0459)	-0.0497*** (0.0120)
K_{-4}	-0.0245 (0.0233)	-0.0410** (0.0182)	0.0242 (0.0388)
K_{-3}	0.0293 (0.0366)	-0.0247 (0.0319)	-0.0022 (0.0226)
K_{-2}	-0.0223 (0.0257)	0.0137 (0.0341)	-0.0191 (0.0448)
K_{-1}	-0.0306 (0.0208)	-0.0028 (0.0257)	-0.0231 (0.0245)
K_0	0.0811** (0.0330)	0.1028** (0.0421)	0.1312** (0.0580)
K_1	0.0331 (0.0273)	-0.0194 (0.0192)	-0.0192 (0.0226)
K_2	0.0286 (0.0259)	-0.0238 (0.0157)	0.0094 (0.0347)
K_3	-0.0009 (0.0143)	-0.0112 (0.0161)	0.0003 (0.0213)
K_4	0.0327 (0.0206)	0.0202 (0.0225)	0.0310 (0.0319)
K_5	-0.0023 (0.0111)	0.0051 (0.0170)	-0.0185*** (0.0043)
N	2697816	2693899	2691619

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

Notes: This Table replicates columns (4)-(6) of Table 3 using a different definition of attainability. Standard errors reported in parentheses are clustered at the store level. Please see notes to Table 3 for additional details.

C.3 Alternative Control Groups

In our main analysis, the control group included all non-last days in a month, plus non-last days on which no threshold was attainable. However, it is possible that a forward-looking team can strategically manipulate the presence or the position of an attainable threshold on the last day of a month. To alleviate this concern, Table C.3.1 excludes the week preceding the month's last day from the control group, since it becomes easier to manipulate the position to a threshold when it proceeds near the end of a month.

Table C.3.1: Robustness Check – Excluding the Week Preceding the Month's Last Day

	Dependent variable: Indicator for team sales falling in a given bin on the last day of the month		
	T	$1.3T$	$1.6T$
K_{-5}	-0.0158 (0.0283)	-0.0356 (0.0344)	-0.0551*** (0.0109)
K_{-4}	-0.0449* (0.0249)	-0.0315 (0.0260)	0.0542 (0.0519)
K_{-3}	0.0505 (0.0399)	-0.0255 (0.0343)	-0.0074 (0.0251)
K_{-2}	-0.0284 (0.0282)	0.0054 (0.0340)	-0.0580 (0.0387)
K_{-1}	-0.0431** (0.0191)	0.0329 (0.0313)	-0.0263 (0.0263)
K_0	0.0940*** (0.0337)	0.0896** (0.0427)	0.1416** (0.0622)
K_1	0.0328 (0.0264)	-0.0204 (0.0210)	-0.0200 (0.0250)
K_2	0.0260 (0.0262)	-0.0245 (0.0162)	0.0140 (0.0363)
K_3	0.0003 (0.0146)	-0.0079 (0.0171)	0.0041 (0.0217)
K_4	0.0326 (0.0212)	0.0224 (0.0230)	0.0318 (0.0338)
K_5	-0.0024 (0.0114)	0.0053 (0.0181)	-0.0143*** (0.0045)
N	2150998	2146959	2145790

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

Notes: This Table replicates columns (4)-(6) of Table 3, excluding the week preceding the month's last day from the control group. Standard errors reported in parentheses are clustered at the store level. Please see notes to Table 3 for additional details.

Another possibility is that last days are so unique that we cannot use non-last days to control for their sales cycle. In Table C.3.2, we use only last days when no target or kink is attainable as controls. Due to the much smaller sample size, we control for Bin×Store and Bin×Holiday fixed effects in columns (4)-(6). In both cases, the estimated effects are also very similar to our main results.

Table C.3.2: Robustness Check – Using Only Last Days as Control Days

	Dependent variable: Indicator for team sales falling in a given bin on the last day of the month		
	<i>T</i>	<i>1.3T</i>	<i>1.6T</i>
K_{-5}	-0.0056 (0.0303)	-0.0263 (0.0352)	-0.0295*** (0.0108)
K_{-4}	-0.0530* (0.0272)	-0.0044 (0.0300)	0.0176 (0.0608)
K_{-3}	0.0721* (0.0399)	-0.0289 (0.0361)	-0.0052 (0.0306)
K_{-2}	-0.0047 (0.0271)	0.0208 (0.0387)	-0.0379 (0.0359)
K_{-1}	-0.0185 (0.0196)	0.0297 (0.0323)	-0.0334 (0.0291)
K_0	0.1002*** (0.0373)	0.1309*** (0.0423)	0.1343** (0.0665)
K_1	0.0271 (0.0273)	-0.0306 (0.0226)	-0.0026 (0.0231)
K_2	0.0405 (0.0279)	-0.0183 (0.0156)	0.0233 (0.0339)
K_3	-0.0053 (0.0166)	-0.0204 (0.0201)	0.0101 (0.0242)
K_4	0.0365 (0.0221)	0.0304 (0.0217)	0.0294 (0.0330)
K_5	-0.0082 (0.0128)	-0.0001 (0.0189)	-0.0277*** (0.0085)
N	80506	76467	75298

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

Notes: This Table replicates columns (4) - (6) of Table 3, using only last days of the month as the control group. Standard errors reported in parentheses are clustered at the store level. Please see notes to Table 3 for additional details.

C.4 Setting Sales Bin Widths Relative to a Store's Target

In this section, instead of a common bin size of \$100 across all stores, sales bins are set equal to 0.01 of each store's monthly target T . Thus we now use smaller bins in smaller stores. The table replicates columns 3-6 of Table 3, and shows very similar results.

Table C.4.1: Robustness Check – Defining Distance and Sales by Share of the Target

	Dependent variable: Indicator for team sales falling in a given bin on the last day of the month		
	T	$1.3T$	$1.6T$
K_{-5}	-0.0630 (0.0407)	-0.0787*** (0.0280)	-0.0074 (0.0571)
K_{-4}	-0.0589 (0.0379)	-0.0688** (.0306)	-0.0164 (0.0405)
K_{-3}	-0.0620** (0.0306)	-0.0064 (0.0402)	-0.0363 (0.0254)
K_{-2}	0.0180 (0.0384)	-0.0286 (0.0257)	-0.0208 (0.0436)
K_{-1}	-0.0559*** (0.0210)	-0.0214 (0.0329)	-0.012 (0.0344)
K_0	0.0670* (0.0342)	0.0791** (0.0382)	0.1116* (0.0653)
K_1	0.0407 (0.0302)	0.0049 (0.0281)	0.0308 (0.0383)
K_2	0.0321 (0.0255)	0.0101 (0.0232)	-0.0343*** (0.0068)
K_3	0.0312 (0.0236)	-0.0137 (0.0182)	0.0252 (0.0338)
K_4	-0.012 (0.0108)	0.021 (0.0243)	0.0015 (0.0254)
K_5	-0.0031 (0.0110)	0.0222 (0.0212)	-0.0128*** (0.0035)
N	2292967	2289744	2288239

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

Notes: This Table replicates columns (4) - (6) of Table 3. Distance and sales are defined into bins according to share of the monthly target T . Each bin is defined as 0.01 of monthly target. Standard errors reported in parentheses are clustered at the store level. Please see notes to Table 3 for additional details.

D Appendix D: Heterogeneity

D.1 Heterogeneity Examination – Turnover Events

Table D.1.1 studies the bunching effect when there are recent turnover events at the store. We find that the likelihood of bunching at both the target and kinks is not significantly different in months when a turnover has occurred, or in months following a turnover event, compared to other months. But for the two kinks, the presence of turnover events is estimated to reduce the likelihood of shifting sales up to the bins that are more than five bins above the kink's bin.

Table D.1.1: Heterogeneity Examination – Turnover Events

	Dependent variable: Indicator for team sales falling in a given bin on the last day of the month			
	T	$1.3T$ & $1.6T$	T	$1.3T$ & $1.6T$
$K_{\leq -6} \times \text{Turnover}_{im}$	-0.0230 (0.0631)	0.0966 (0.1054)		
$K_{-5,-1} \times \text{Turnover}_{im}$	-0.0725 (0.0969)	-0.0773 (0.1010)		
$K_0 \times \text{Turnover}_{im}$	0.0861 (0.1032)	0.1267 (0.1086)		
$K_{1,5} \times \text{Turnover}_{im}$	-0.1310 (0.0935)	0.0190 (0.1063)		
$K_{\geq 6} \times \text{Turnover}_{im}$	0.0960 (0.0832)	-0.1815*** (0.0672)		
$K_{\leq -6} \times \text{Turnover}_{im2}$			-0.0079 (0.0628)	0.1181 (0.0890)
$K_{-5,-1} \times \text{Turnover}_{im2}$			-0.1219 (0.0989)	-0.0835 (0.0904)
$K_0 \times \text{Turnover}_{im2}$			0.0843 (0.0887)	0.0782 (0.0905)
$K_{1,5} \times \text{Turnover}_{im2}$			-0.0430 (0.0930)	0.0185 (0.0884)
$K_{\geq 6} \times \text{Turnover}_{im2}$			0.0512 (0.0576)	-0.1546** (0.0672)
N	2697816	2697791	2697816	2697791

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

Notes: This Table replicates Table 4. In addition to variables identifying bins on either side of the threshold as in Table 4, each regression adds regressors that interact the corresponding bins with indicators for recent turnover on the sales team. $\text{Turnover}_{i,m}$ is an indicator variable, taking a value of 1 if there is turnover event during the current month m at store i . $\text{Turnover}_{i,m2}$ is an indicator variable, identifying month m if there is turnover event at store i , and identifying the month $m + 1$ following the turnover. To economize on degrees of freedom, the regressions for $1.3T$ and $1.6T$ are combined. Coefficients are rescaled as in Table 4. Standard errors reported in parentheses are clustered at the store level. Please see notes to Table 4 for additional details.

D.2 Heterogeneity Examination – Effects of Team Size and Expected Sales

In this section, we examine whether the amount of bunching varies with the size of the team. As shown in Table D.2.1, we find bunching at both the target and the kinks to be statistically significant and generate larger effects in teams of more workers. For the target T , the bunching effects in teams of two workers are significantly smaller than in teams of three or more workers (p -values are 0.04 and 0.01, respectively). But the bunching effects at the kinks are not significantly different across team sizes. Figure D.2.1, we disaggregate stores by the number of workers to replicate Figure 5. In the top, middle, and bottom panel, we plot excess (or deficient) sales at the thresholds T , $1.3T$, and $1.6T$. For all thresholds, we find the bunching to be more prominent in stores of at least 4 workers. Empirical analysis and graphical evidence both suggest that thresholds might be functioning as coordination devices for team production.

In Table D.2.2, we disaggregate stores by the size of their typical daily sales. To do so, we first calculate the typical daily sales output of a store-month by $TDS = T/D$, where T is a store's target in the current month, and D is the number of days in the month. Consistent with the prior finding, bunching is more pronounced in store-months with higher levels of expected sales. In addition, in the larger stores, the excess masses at bunching are from bins that are more than five bins below the threshold's bin. Since typical daily sales output is more than \$500 in these stores, to improve sales output by more than \$500 is not overly challenging for them.

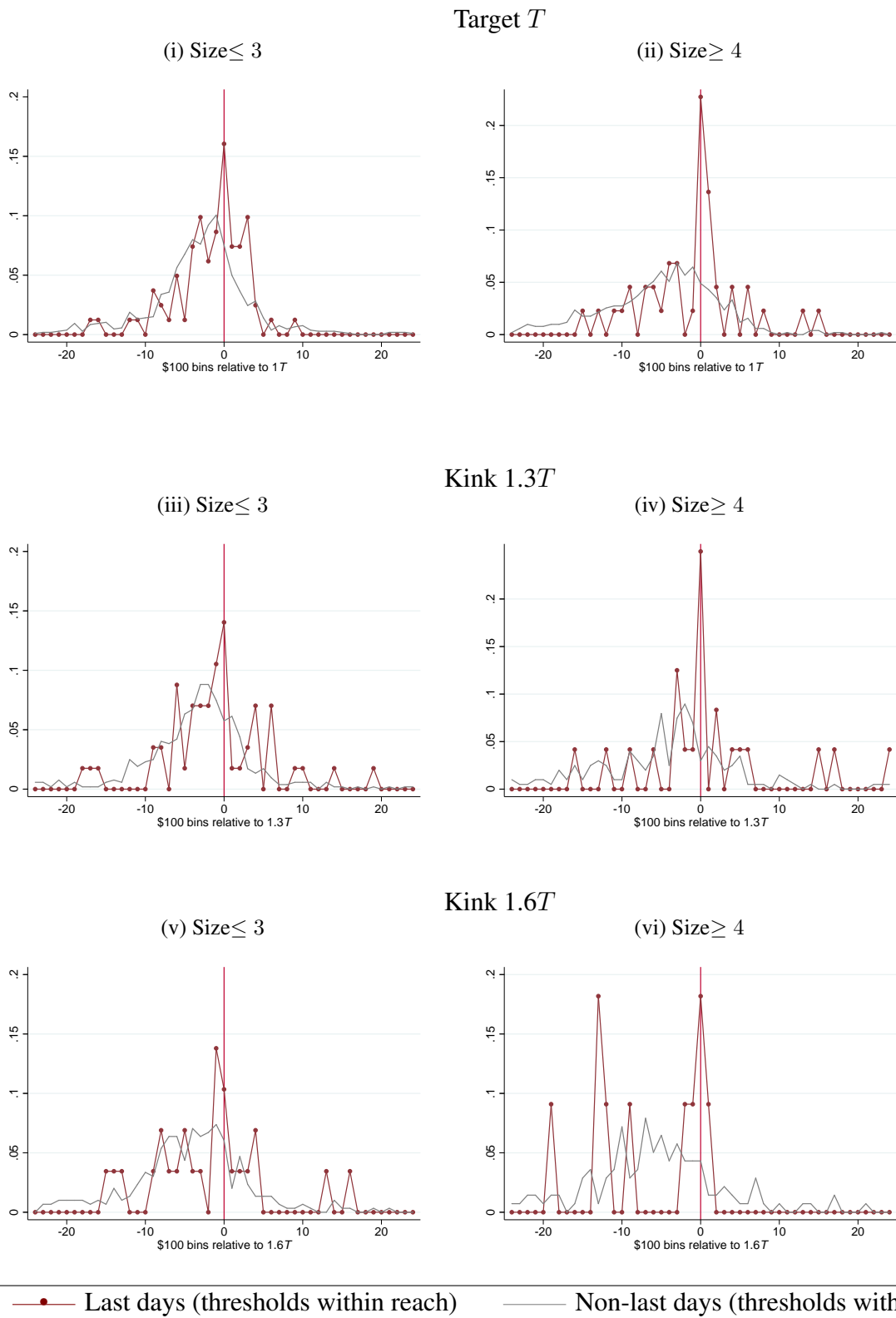
Table D.2.1: Heterogeneity Examination – Team Size Effects

Dependent variable: Indicator for team sales falling in a given bin on the last day of the month		
	T	$1.3T$ & $1.6T$
Size=2:		
$K_{\leq -6} \times (\text{Size}=2)$	0.0725 (0.1109)	-0.1730 (0.3643)
$K_{-5,-1} \times (\text{Size}=2)$	-0.0127 (0.1408)	0.0275 (0.1402)
$K_0 \times (\text{Size}=2)$	-0.0866 (0.0758)	0.0213 (0.0861)
$K_{1,5} \times (\text{Size}=2)$	0.0955 (0.0960)	-0.0734 (0.0559)
$K_{\geq 6} \times (\text{Size}=2)$	-0.0640 (0.0448)	0.1546 (0.2285)
Size=3:		
$K_{\leq -6} \times (\text{Size}=3)$	-0.1141*** (0.0427)	-0.1421*** (0.0341)
$K_{-5,-1} \times (\text{Size}=3)$	-0.0622 (0.0491)	-0.0619 (0.0417)
$K_0 \times (\text{Size}=3)$	0.0965** (0.0454)	0.1067** (0.0456)
$K_{1,5} \times (\text{Size}=3)$	0.0955* (0.0545)	0.0175 (0.0459)
$K_{\geq 6} \times (\text{Size}=3)$	-0.0384 (0.0256)	0.0941* (0.0538)
Size≥4:		
$K_{\leq -6} \times (\text{Size} \geq 4)$	-0.1033*** (0.0362)	-0.1553*** (0.0530)
$K_{-5,-1} \times (\text{Size} \geq 4)$	-0.0862* (0.0502)	-0.0461 (0.0608)
$K_0 \times (\text{Size} \geq 4)$	0.1442*** (0.0492)	0.2050** (0.0801)
$K_{1,5} \times (\text{Size} \geq 4)$	0.1155* (0.0655)	-0.0339 (0.0604)
$K_{\geq 6} \times (\text{Size} \geq 4)$	0.0062 (0.0448)	0.0470 (0.0470)
N	2697816	2697791

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

Notes: This Table replicates Table 4, interacting the main regressors with indicators for team size. To economize on degrees of freedom, the regressions for $1.3T$ and $1.6T$ are combined. Standard errors reported in parentheses are clustered at the store level. Coefficients are rescaled as in Table 4. Please see notes to Table 4 for additional details. For the target T , p -values for the differences between $K_0 \times (\text{Size}=2)$ and $K_0 \times (\text{Size}=3)$, and $K_0 \times (\text{Size}=2)$ and $K_0 \times (\text{Size}=4)$ are 0.0408 and 0.0121 respectively. For the kinks, those p -values are 0.3824 and 0.1213 respectively.

Figure D.2.1: Excess (or Deficient) Sales Relative to the threshold (By Store Size)



Notes: These figures plot the histogram of excess (or deficient) sales relative to the corresponding kinks, measured in bins of \$100, separately for last days and non-last days. For better visual display, only the range [-\$2,500, \$2,500] is presented. The figures are plotted for T , $1.3T$, and $1.6T$, separately. On the left-hand side, the sample is restricted to stores which have no more than 3 workers; and on the right-hand side, the sample is restricted to stores which have at least 4 workers. Please see Figure 5 for figures of the full sample.

Table D.2.2: Heterogeneity Examination – By Size of a Store’s Typical Daily Sales

	Dependent variable: Indicator for team sales falling in a given bin on the last day of the month	
	<i>T</i>	1.3 <i>T</i> & 1.6 <i>T</i>
Low sales team months (TDS ≤ 500):		
$K_{\geq -6} \times i.(TDS \leq 500)$	-0.1112 (0.1069)	0.0051 (0.1181)
$K_{-5,-1} \times i.(TDS \leq 500)$	-0.0550 (0.0598)	-0.0494 (0.0531)
$K_0 \times i.(TDS \leq 500)$	0.0704 (0.0473)	0.0987** (0.0467)
$K_{1,5} \times i.(TDS \leq 500)$	0.0760 (0.0515)	-0.0110 (0.0454)
$K_{\geq 6} \times i.(TDS \leq 500)$	-0.0960 (0.0896)	-0.0538 (0.0336)
High sales team months (TDS > 500):		
$K_{\leq -6} \times i.(TDS > 500)$	-0.1069*** (0.0283)	-0.1275*** (0.0265)
$K_{-5,-1} \times i.(TDS > 500)$	-0.0553 (0.0354)	-0.0223 (0.0384)
$K_0 \times i.(TDS > 500)$	0.1321*** (0.0482)	0.1283** (0.0560)
$K_{1,5} \times i.(TDS > 500)$	0.1115* (0.0605)	-0.0180 (0.0414)
$K_{\geq 6} \times i.(TDS > 500)$	0.0576* (0.0384)	0.2083** (0.0874)
N	2697816	2697791

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

Notes: This Table replicates Table 4, interacting regressors with indicators of high or low typical store sales. TDS is calculated for each store-month by $TDS = T/D$, where T is the store’s monthly target, and D is the number of days in the current month. To economize on degrees of freedom, the regressions for 1.3*T* and 1.6*T* are combined. Coefficients are rescaled as in Table 4. Standard errors reported in parentheses are clustered at the store level. Please see notes to Table 4 for additional details.

D.3 Heterogeneity Examination – Firm Tenure Effects

Table D.3.1 studies whether experienced sales teams are more or less likely to bunch at thresholds. As shown in Table D.3.1, the amount of bunching at both the target and the kinks increases with average firm tenure of the sales teams, suggesting that teams may use thresholds as goals to coordinate team efforts.

Table D.3.1: Heterogeneity Examination – Firm Tenure Effects

	Dependent variable: Indicator for team sales falling in a given bin on the last day of the month	
	<i>T</i>	1.3 <i>T</i> & 1.6 <i>T</i>
$K_0 \times \text{c.Tenure}$	0.0128** (0.0058)	0.0190** (0.0090)
N	2697816	2697791

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

Notes: This Table replicates Table 4. In addition to variables identifying bins on either side of the threshold as in Table 4, each regression adds regressors that interact the threshold-containing bin with a continuous measure of the team's mean tenure with Firm A. To economize on degrees of freedom, the regressions for 1.3*T* and 1.6*T* are combined. Like Table 4, all regressions include variables identifying bins on either side of the threshold. Standard errors reported in parentheses are clustered at the store level. Please see notes to Table 4 for additional details.

D.4 Heterogeneity Examination – Promotion Prospects

Table D.4.1 examines whether bunching is more pronounced in teams where their managers can possibly receive promotions into managerial positions in the head office. If a store manager believes that attaining a higher assessment could increase her promotion chances into managerial positions, she might try to motivate her teammates to reach a higher threshold. As managerial positions are only offered at the head office, we use the location of the retail store to proxy for the manager’s promotion chances. As shown in Table D.4.1, promotion prospects do not significantly affect the likelihood of bunching in our context.

Table D.4.1: Heterogeneity Examination – Promotion Prospects

	Dependent variable: Indicator for team sales falling in a given bin on the last day of the month	
	<i>T</i>	1.3 <i>T</i> & 1.6 <i>T</i>
$K_{-5,-1} \times \text{Near HQ}$	0.1728** (0.1041)	0.0044 (0.0674)
$K_0 \times \text{Near HQ}$	-0.0804 (0.0667)	0.0142 (0.0681)
$K_{1,5} \times \text{Near HQ}$	0.0385 (0.0800)	0.0160 (0.0689)
N	2697816	2697791

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

Notes: This Table replicates Table 4. In addition to variables identifying bins on either side of the threshold as in Table 4, each regression adds regressors that interact the corresponding bins with an indicator for a store’s location. Near HQ is an indicator variable, taking a value of 1 if the store locates in the same city where the head office locates. As managerial positions are only offered at the head office, it is a proxy for higher promotion chances. To economize on degrees of freedom, the regressions for 1.3*T* and 1.6*T* are combined. Coefficients are rescaled as in Table 4. Standard errors reported in parentheses are clustered at the store level. Please see notes to Table 4 for additional details.