The Negative Consequences of Loss-Framed Performance Incentives

Lamar Pierce, Alex Rees-Jones, and Charlotte Blank*

Abstract

Behavioral economists have proposed that incentive contracts result in higher productivity when bonuses are "loss framed"—prepaid then clawed back if targets are unmet. We test this claim by randomizing the pre- or post-payment of sales bonuses at 294 car dealerships. Although somewhat statistically imprecise, our analysis provides strong indications that the random assignment of loss framing had quantitatively important *negative* effects. We document that negative effects of loss framing can arise due to an increase in incentives for "gaming" behaviors. Based on these claims, we reassess the common wisdom regarding the desirability of loss framing.

Keywords: Loss Aversion, Field Experiments, Worker Incentives, Franchise Contracts.

JEL Codes: D9, D81, J22, J31.

*Date: Feb 11, 2024. Pierce: Olin Business School, Washington University in St. Louis, pierce@wustl.edu. Rees-Jones: The Wharton School, University of Pennsylvania and NBER, alre@wharton.upenn.edu. Blank: Jaguar Land Rover NA, cblank1@jaguarlandrover.com. The authors are grateful to Maritz, LLC for making possible the natural field experiment in this paper, and to Meghan Busse and Florian Zettelmeyer for help with design of the experiment. Pierce and Rees-Jones accepted no compensation from any party for this research. Blank served as Chief Behavioral Officer of Maritz, LLC during the conduct of this research. For especially helpful comments, we thank Greg Besharov, Stefano DellaVigna, Jon de Quidt, Alex Imas, Ian Larkin, and Ted O'Donoghue. We also thank seminar participants at Cornell University, George Washington University, Goethe University Frankfurt, INSEAD, the University of Mannheim, Purdue University, the University of Chicago, the University of Maryland, the University of South Florida, and Yale University, as well as conference participants at the AEA Annual Meetings, the Society for Judgment and Decision Making Annual Meetings, the Stanford Institute for Theoretical Economics Psychology and Economics Conference, and the Institutions and Innovation Conference.

Worker incentives commonly feature bonuses for exceeding a production target. At least since the work of Hossain and List (2012), behavioral economists have often advised that the benefits of such incentives can be raised by prepaying the bonus with a threat of clawback if the target is not met. If workers are loss averse and if prepayment leads workers to perceive failure to retain the bonus as a loss, then this intervention could induce an endowment effect (Kahneman, Knetsch and Thaler, 1991) that makes hitting the target seem more important. Greater effort and productivity could then follow.

In this paper, we report the results of a field experiment where a car manufacturer applied this advice to its dealer sales incentive program. In this field experiment, random assignment of the prepayment of bonuses did not result in the predicted increase in sales. Instead, we find a marginally significant and quantitatively meaningful *negative* effect. This finding is surprising when viewed through the lens of the conventional wisdom on the effects of loss aversion, and when considered in the context of a purely rational model.[1] Our analysis suggests that this negative effect is at least partly influenced by the interaction between loss framing and the perverse incentives that arise in multitasking environments. Motivated by these empirical results, we theoretically explore the positive and negative consequences of loss framing and establish that an important negative consequence is the exacerbation of broadly defined "gaming" responses to incentives.

This project began when the car manufacturer became aware of the behavioral economics literature on loss-framed incentives. This literature, and its prescriptions, appeared directly relevant: the manufacturer provides car dealers with separate monthly targets for the sales of two groups of cars (which we will refer to as *model groups*) and pays a large bonus for exceeding each target. Based on the line of logic presented above, the manufacturer believed they might induce greater dealer effort by changing their existing post-payment scheme to a prepaid one. Toward that end, they recruited the authors of this paper to help implement the new bonus scheme.

Due to concerns about both the efficacy and desirability of loss framing, the authors recommended that this policy be tested in a randomized controlled trial (RCT) prior to im-

[1] In a purely rational model, from the perspective of car dealers, the effect of early payment is simply the provision of additional liquidity. If liquidity constraints do not bind, this should be irrelevant for productivity, and if liquidity constraints do bind then prepayment could enable productivity-enhancing investment.

plementation. While many "nudge" interventions are extremely low cost and may therefore be productively adopted even if their impact is low (Benartzi et al., 2017), this intervention differs: prepayment of incentives is in effect an interest-free loan, the costs of which are substantial given the scale of the bonus program. To directly test if the benefits of prepayment justify these costs, 294 car dealerships—with monthly sales of approximately 15,000 vehicles and monthly revenue of $600 million—had $66 million in bonuses randomized to pre- or post-payment conditions. Except for the difference in timing, financial incentives were identical across treatment groups.

In this RCT, the loss-framed prepayment scheme resulted in lower total sales. Our initial analyses are based on a difference-in-differences design, in which we compare the sales path of pre- and post-paid dealers in the four months before and after randomization. During the treatment window, dealers assigned to prepayment sell 2.31 fewer cars, on average ($s.e. = 1.34; p = 0.086$), or 4% less in a Poisson specification ($s.e. = 2\%; p = 0.081$). These estimates are marginally significant but economically large. To illustrate, a 2.31 car sales decline per dealer per month would imply that the treated dealers lost $45 million in revenue during the 4 month treatment window. Forecasting further, this estimate would imply that adopting this policy network-wide for a year would reduce revenue by over $1 billion.

Our estimates suggest that the negative total-sales effect of prepayment was driven by a decline in sales in one of the two model groups. This model group had substantially lower ex ante average sales volume, commensurate with a smaller bonus and thus a smaller potential loss. Moving forward, we will refer to this model group as the *small-bonus model group.* The model group with higher sales volume, which has a larger bonus and thus a larger potential loss, is estimated to have had a near-zero and statistically insignificant change in monthly sales. Moving forward, we will refer to this model group as the *large-bonus model group.*

Models of loss aversion make additional predictions on *when* these contracts would most change behavioral responses. Most importantly, the behavioral response is naturally predicted to depend on proximity to the performance target. If loss framing leads a dealer to care more about exceeding the target, then a loss framing intervention should have its strongest effects when a dealer is at risk of falling just short of its target. In contrast, a dealer who would otherwise substantially over- or under-shoot its target should be minimally influenced by this intervention.

To assess these predictions, we present a novel modification of the standard difference-in-differences approach that estimates the impact of treatment on a distribution rather than a mean. This approach, which we call difference-in-kernel-density-differences (DKDD), estimates the location of missing or excess mass in a distribution by comparing the pre- and post-treatment distributions in the treatment group to the pre- and post-treatment distributions in the control group.

Applying this technique, we examine the effect of the switch to the prepaid incentives on the distribution of monthly sales of each model group. Consistent with our earlier estimate of an average reduction in the small-bonus model group's sales, we find that this model group's sales are generally shifted from higher to lower values. More interestingly, we find that the previously reported null effect on average sales for the large-bonus model group was masking substantial and statistically detectable changes to the distribution. Dealers became significantly less likely to end the month with large-bonus model group sales narrowly below the target. Instead of ending the month having just barely triggered a loss, some dealers increased their sales over the target. Other dealers decreased their sales, with some evidence that they attended to selling the small-bonus model group instead. Taken together, these two groups lead to offsetting average effects, demonstrating why the initial difference-in-differences estimates yielded a null effect on average large-bonus model group sales despite statistically detectable changes in their behavior.

The patterns of results that we document resemble common findings on the downsides of multitasking incentives. An organizing idea from this literature is that, when multiple dimensions are separately incentivized, agents may benefit—perhaps at the expense of a principal—by neglecting one dimension to attend to another dimension with higher private returns. This tension is present in the foundational model of Holmstrom and Milgrom (1991), and has been extensively studied in research in both policy and firm settings (see, e.g., Baker and Hubbard, 2003; Dumont et al., 2008; Chen, Li and Lu, 2018; Kim, Sudhir and Uetake, 2021). However, to our knowledge, the manner in which these incentives interact with the loss-framing interventions suggested by Hossain and List (2012) has not been previously explored.[2] Our belief that the logic of this multitasking literature might explain the otherwise surprising findings of our experiment led us to conduct a theoretical reexamination of the predicted role of loss framing.

To study this issue theoretically, we model a principal incentivizing an agent's acceptance of private costs of production through bonuses based on ex-post performance, potentially across multiple measured dimensions. We model agents as having a set of possible production strategies available, with each strategy characterized by its induced joint distribution of dimension-specific production and private costs. In this environment, we study how the desirability of different strategies changes when loss framing is applied.

Strategies that are incentivized by loss framing are straightforward to characterize. Loss framing increases the perceived incentive to avoid strategies that are high in a particular measure of exposure to loss. This means that, in fully general environments, loss framing does not have an unambiguously signed effect on productivity. If there are strategies available that reduce this measure of exposure to loss while also decreasing expected productivity, the incentive to pursue these strategies increases. These strategies include, but are not limited to, standard ways of gaming a multidimensional contract.

Our model differs from those previously used to study loss-framed incentives in more ways than just including multiple dimensions. To facilitate comparisons to prior research, we next reevaluate the nested unidimensional case to see if the prevailing wisdom holds there. In general, it does not: we again find that the performance impact of loss framing remains ambiguous and determined by the measure of exposure to loss. However, we illustrate a special case of broad relevance in which loss framing can be understood to be unambiguously helpful: the case where the agent's choice of distributions of output induced by his strategy can be ordered by first-order stochastic dominance. When this property holds, loss framing is predicted to weakly increase expected productivity. When this property does not hold, effects are ambiguous. The key insight driving this distinction is that loss framing not only steepens incentives, but also modifies risk tolerance. If actions can be taken that reduce the

[2] A partial exception to this claim is Balmaceda (2018), which presents results on multitasking incentives in the expectations-based reference-dependence model of Kőszegi and Rabin (2006). This model features loss aversion, but does not have scope for a framing effect induced by prepayment because reference points are endogenously determined by expectations.

risk of a loss at the cost of expected production, loss framing incentivizes such actions. When outcome distributions are ordered by first-order stochastic dominance, there is no tradeoff between loss probability and expected productivity, and thus this potential for perverse incentives is eliminated. When outcome distributions are not so ordered, this tradeoff can become important.

In both the uni- and multi-dimension cases, a simple intuition drives our results: inducing loss aversion through loss framing motivates loss avoidance. While this insight borders on tautology, notice that it is different than the typical claim that inducing loss aversion motivates effort. In cases where agents' only means to protect themselves from losses is to increase the distribution of productivity (whether through effort or other means), then loss framing is unambiguously predicted to steepen the incentives to take those actions. If, however, more elements of the agent's approach (such as risk tolerance or multitasking strategy) can be modified to influence the probability of losses, then steepening incentives will also influence those decisions. Gaming of incentive schemes through such modifications is typically viewed to be undesirable from the perspective of a principal, and this influence on incentives to game is the negative consequence of loss-framed performance incentives.

Our findings contribute to several lines of research. First, and most directly, our findings directly inform the literature on loss-framed performance incentives. This literature was largely initiated by Hossain and List (2012), who made a conceptual argument that loss framing should enhance the productivity of loss-averse workers and tested that prediction in a comparatively small field experiment. This paper motivated a substantial body of follow-up research from both the lab (see, e.g., Brooks, Stremitzer and Tontrup, 2012; Imas, Sadoff and Samek, 2016; de Quidt et al., 2017; de Quidt, 2017; DellaVigna and Pope, 2018; Ahrens, Bitter and Bosch-Rosa, 2023) and the field (see, e.g., Hong, Hossain and List, 2015; Chung and Narayandas, 2017; Brownback and Sadoff, 2020; Bulte, List and van Soest, 2020, 2021; Fryer et al., 2022). As emphasized in the recent metaanalysis of Ferraro and Tracy (2022), a notable contrast exists between this lab and field literature: field tests on average suggest near-zero effects, whereas lab tests on average suggest large positive effects. Ferraro and Tracy document that this contrast may be at least partly attributed to publication bias and underpowered designs leading to an exaggerated average published effect within the lab literature.

Our paper provides two critical inputs to this literature. First, we provide an unusually large-scale and unusually highly incentivized test of the impact of loss-framed performance incentives. This test yields perhaps surprising results, demonstrating previously undocumented potential for reliance on loss framing to be counterproductive. Second, we establish that the central prediction that is stated and tested within this literature—that loss framing increases productivity among loss-averse agents—is not unambiguously a prediction in relatively general models without further conditions. We provide more refined conditions that provide guidance on when such productivity gains are guaranteed, and point to factors that could lead the common statement of the prediction to fail. Our theoretical findings may help further organize the different effects documented across settings. The lab settings that have most reproducibly found a positive effect of loss framing feature relatively uncomplicated effort decisions that seem particularly amenable to the assumption that strategies are ordered by first order stochastic dominance. In contrast, the field settings studied examine more complex decisions, which likely involve more elaborate tradeoffs between risk of loss and productivity. While we do not believe that this is the only explanation for the contrast, we do believe that it contributes.

Our paper also contributes to the broader literature on applications of behavioral economics in policy settings. In recent years, there has been a great deal of interest in harnessing behavioral economics in general, and loss framing in specific, as a policy tool. As examples, loss framing interventions have been deployed with the goal of influencing tax compliance (Hallsworth et al., 2017), teaching effort (Fryer et al., 2022), or disposable bag use (Homonoff, 2018). Despite some success stories, studies of large-scale policy interventions of this type are still relatively rare, and results on the efficacy of these interventions are mixed. A reasonable reader of this literature can still question whether these interventions can be deployed "at scale" with expectations of meaningful behavioral response. We provide two inputs to this debate. First, we document the deployment of a loss-framing intervention at a scale rarely attempted in the behavioral economics literature, and find that it did lead to meaningful behavioral response. While ultimately the induced behavioral response was undesirable, the fact that behavioral-economic factors were activated suggests hope for harnessing them for policy purposes. Second, our study provides a rare demonstration of such interventions having a behavioral-economic impact when the targeted units are *firms* and not *individuals*. We return to further discussion of these contributions in the conclusion.

The remainder of this article proceeds as follows. In Section 1 we describe the setting, design, and results of our field experiment. In Section 2 we present a theoretical examination of the positive and negative consequences of loss-framed incentives. Section 3 concludes, and the online appendix contains a variety of additional analyses.

1 An RCT on Loss Framing in Automobile Sales

In this section we present our field experiment. We begin by providing institutional details necessary for understanding the market, followed by a detailed description of the experiment that was deployed. We then analyze the effect of prepayment on average sales before proceeding to examine its effect on the distribution of sales.

1.1 Setting

Our empirical setting is the new automobile market—a market governed by extensive contracting between automobile manufacturers and automobile dealers.

State laws in the United States significantly restrict manufacturers from directly selling vehicles to consumers, effectively forbidding vertical integration. As a result, manufacturers sell vehicles to independent dealers, who then sell or lease vehicles to consumers. Dealers are organized into designated market areas (DMAs) such as St. Louis, Missouri or Philadelphia, Pennsylvania, within which multiple independent dealers typically compete.[3] The sales activity of these dealers is our topic of focus in this paper.

Independent dealers operate under franchise agreements that significantly shape the automobile market. Existing franchise laws restrict manufacturers' ability to open or close dealers, heavily limiting manufacturers' ability to control interbrand competition or react to changes in demand. Such laws also highly limit the ability of manufacturers to treat dealers

[3] See Lafontaine and Scott Morton (2010) or Murry and Schneider (2016) for detailed information on the history and function of car dealership franchises.

differently, limiting the means by which they may present tailored sales incentives through dealer-specific prices (and more broadly limiting ability to price discriminate).

Although many of the approximately 17,000 new vehicle dealers in the United States are now being consolidated by private equity and publicly traded corporations such as Autonation and Lithia, the industry has historically been fragmented for several reasons (Roberts, 2018). First, manufacturers have typically blocked dealer sales that put multiple competing dealers under common ownership. Second, many franchise laws prohibit owning more than a certain number of dealerships per state. With the lack of entry and exit and the limits on consolidation, the dealers for any given manufacturer represent a diverse mix of entities organized as public and private national corporations, private equity, family-owned dealers and groups, and other independently owned dealers. Across these disparate players, there are substantially varying managerial approaches and degrees of efficiency.

As a consequence of this market structure, manufacturers must use somewhat indirect means to influence dealers' sales incentives. Franchise laws discussed above constrain manufacturers to sell vehicles to dealers for a fixed invoice price. Dealers then have discretion to price the vehicles independently before selling to consumers at a negotiated price (Busse and Silva-Risso, 2010; Bennett, 2013). Since manufacturers rarely change published retail and invoice prices within a given year, other mechanisms are used to affect prices and quantities within their dealer network. Manufacturers provide cash incentives to both dealers and customers in order to reduce vehicles' effective price if there is excess inventory (Busse, Silva-Risso and Zettelmeyer, 2006). Similarly, they subsidize loans and leases through captive finance arms (Pierce, 2012). To motivate sales volume, manufacturers typically use direct incentive programs that reward units sold. These programs are intended to partially address misaligned incentives for dealers to sell lower volume at higher prices, since the manufacturer benefits from volume but cannot capture value from higher prices in a given year (Busse, Silva-Risso and Zettelmeyer, 2006).[4] Volume-based incentives also directly promote increasing the manufacturers' market share, which is valuable for preserving the manufacturer's competitive position within the industry and which is directly rewarded in executive

[4]This incentive misalignment is partially mitigated by the profits from servicing cars they have sold, which is ultimately influenced by sales volume.

compensation (Ritz, 2008; Pierce, 2012). These volume-based incentive programs are the policies of interest in this paper.

The car manufacturer that we study has multiple vehicle models sold at over a thousand dealers across all fifty states. Since our access to data is governed by a non-disclosure agreement, we will refer to this manufacturer as CarCo.[5] Like many manufacturers, CarCo uses incentive programs for specific vehicle models to motivate sales volume at the majority of its dealers. The specific program considered in this paper has two separate incentive schemes, each focused on a particular set of models. We refer to these sets of models as the large-bonus model group and the small-bonus model group, as described in the introduction. Both groups are typically, but not always, retailed together at the same dealerships: 88% of dealers in the 2017 program sold both model groups. Dealers in the focal incentive program sell over 90% of all vehicles in the two model groups.

Like most dealer incentive programs, CarCo's program rewards dealers for reaching monthly targets.[6] These targets are set for each model group by taking average sales in that calendar month for the previous four years. All vehicles in a specific model-group that are sold to consumers count toward the target equally.[7] For example, a given dealer's April 2018 target would be the average of their April sales from 2014 through 2017 rounded up to the nearest integer. Quota-based systems like these are common across industries, with variation in the number of targets and the time period allowed to reach those targets (Chung, Narayandas and Chang, 2019; Misra and Nair, 2011; Chung, Steenburgh and Sudhir, 2013).

Figure 1 illustrates CarCo's incentive program. Dealers selling below their monthly target earn no bonus. Dealers selling between 100% and 110% of their monthly target receive \$100 for each vehicle sold in that month. Dealers selling 110% or more receive \$700 per vehicle for the large-bonus model group or \$800 per vehicle for the small-bonus model group. The higher per-vehicle bonus in the small-bonus model group primarily reflects slowing sales for those

[5]Our legal agreement allowed CarCo to correct the paper for factual accuracy and anonymity, but we had full rights to publish regardless of our results and conclusions.

[6]For an illustrative example of how these shape policies and behavior at one dealer (not necessarily a CarCo franchisee), listen to Episode 513 of This American Life at https://www.thisamericanlife.org/513/129-cars. We note that supply and demand disruptions due to the COVID-19 pandemic led to the suspension of many of these quota-based systems.

[7]Historically, certain models have received "double points" toward the target, but this did not happen during our time period.

models, but also serves to counteract the group's smaller discrete bonuses at 110% based on lower average sales volume. Bonus qualification operates separately for each model group, such that a dealer could qualify for one or both group-specific bonuses. Because the per-vehicle bonuses apply to *all* vehicles from that group sold in that month—not just marginal vehicles over the target—exceeding the 110% threshold conveys a large fixed bonus.[8] To illustrate, applying the average targets reported in Table 1, selling the single vehicle on the margin of the 110% threshold would yield $\$22,300$ or $\$12,000$ in a given month in the large-bonus and small-bonus model groups, respectively. For the largest dealers, this marginal car can be worth over $\$200,000$.[9]

Participation in this program is voluntary and requires payment of a fixed monthly fee, but the majority of dealers choose to participate.[10] The almost universal participation in this program is largely due to the often intense price competition between dealers in the same DMA. As one dealer explained, a competitor will price below invoice, hoping to make all their profit off the monthly incentive program. A non-participating dealer thus cannot compete on price without losing money.

Although exogenous variation in demand plays a large role in target achievement, car dealers have multiple ways in which they can increase sales that can be broadly categorized as "effort." First, they can increase managerial oversight or incentives for salespeople. Second, they can increase their advertising spending. Third, they can price more aggressively, accepting lower (and at times negative) margins to close deals. Fourth, they can work more aggressively to qualify buyers for financing, which effectively lowers the vehicle price.[11]

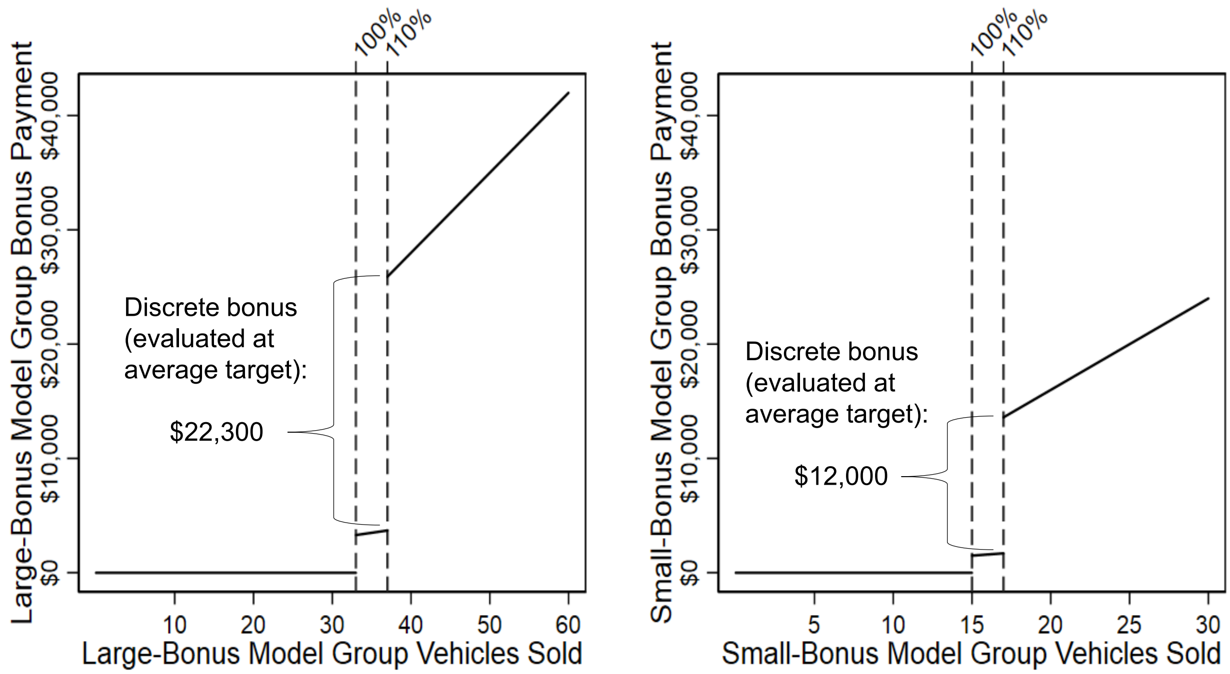
[8]Interviews with dealers consistently indicated that they view the 100% goal as mostly irrelevant, or a small consolation prize for missing the key 110% level.

[9]Such dramatic discontinuities are rarely features of optimal incentives in common related models, and as a result many economists view these contracts as surprising. We note, however, that both Herweg, Müller and Weinschenk (2010) and Gill and Stone (2010) offer cases where highly discontinuous or even binary incentive contracts can be rationalized among reference-dependent agents. In practice, the popularity of contracts like these are often driven by the relative ease of assessing incentives. In order to perceive strong incentives for effort, an agent merely needs to assess whether they have reached a salient sales benchmark. In contrast, qualitative incentives are substantially more complex to evaluate in the non-linear contracts commonly arising in economic theory. Some practitioners view avoiding this complexity to be a major desiderata of contract design.

[10]Most non-participants are very small dealers. The program contract operates on an annual basis, such that incentive structure or features cannot be changed within a calendar year. CarCo will occasionally offer optional features (e.g., model-specific bonuses) within the year that some dealers will choose to accept.

[11]Note that this final option is not desirable if it exposes a captive lender to undue default risk because of misrepresentation of creditworthiness, or if it ruins the credit of customers with strong brand loyalty, thereby

Figure 1: Dealer Incentive Plan for the Two Model Groups



Notes: This figure shows the bonus structure for both model groups. Each monthly model-group bonus is earned independently. Targets are set based on the dealer's average sales over the previous four years in that calendar month, determining the "100%" target indicated in the figure. If sales exceed 110% of that target, a large fixed bonus is conveyed along with additional marginal incentives. To help illustrate the typical application of this bonus system, we have plotted this figure applying the the average targets taken from Table 1.

Dealers also have several ways in which they can react to monthly sales targets that involve "gaming" the incentive system (Larkin, 2014; Oyer, 1998; Courty and Marschke, 2004). First, dealers can attempt to move customers across calendar months (or direct their salespeople to do so). If the dealer is near the crucial 110% threshold, they can attempt to accelerate the purchase decision of a customer by offering a lower price or prioritizing the paperwork and financing. If a given sale is irrelevant for target attainment, dealers can attempt to delay its closing until after the month-end so it helps towards next-month's target attainment. Second, dealers can potentially influence the types of cars they sell by either directing customers to particular models or prioritizing customers with a preference for particular models. Dealers may attempt to sell specific cars within a model group based on expected ease of sale (to earn volume bonuses) versus their mark-up (to yield direct profits).

barring them from future new car purchases (Jansen et al., 2023).

They may also seek to advance one entire model group over the other based on the relative likelihood of reaching their discrete 110% bonuses in that month. Dealers rarely focus all attention on selling one model group from the beginning of the month because the relative likelihoods (and thus expected values) of reaching the targets are not immediately clear. Customers may strongly prefer one model over others, so pushing them hard to purchase from the other model group risks losing a sale to other dealers. Dealers also are sensitive to CarCo's ability to reallocate inventory away from dealers that ignore one particular model group.

Although these gaming behaviors might mitigate the risk of missing the large payoffs from hitting thresholds, they can be costly to the manufacturer due to their risk of driving customers to other manufacturers' vehicles. More directly, increased gaming is undesirable for the manufacturer because it is aimed to increase bonus payments without necessarily increasing sales volume. Like Oyer (1998) and Larkin (2014), we note that the marginal cost of increased gaming does not imply that the incentive structure is necessarily suboptimal, since the motivational benefits from the convex structure may justify gaming costs.

1.2 Experimental Design

The experiment in this paper originated with a CarCo brand manager's interest in using loss framing to increase the efficacy of their dealer incentive program. Several CarCo executives had been discussing launching a version of their incentive program in which they manipulated the timing of bonus payments to induce loss framing. Monthly bonus payments had historically been paid in the month following the incentivized sales, allowing the program manager time to validate the qualification of each vehicle reported as sold. As an alternative, CarCo considered giving dealers a large up-front bonus each month, then clawing back any unearned money at the end of the month. CarCo leaders expected this front-loaded payment to be both highly appealing and motivational for dealers, who would appreciate the advanced cash flow and work hard to retain it. The brand managers thus proposed to change the incentive plan to universally implement advanced payments.

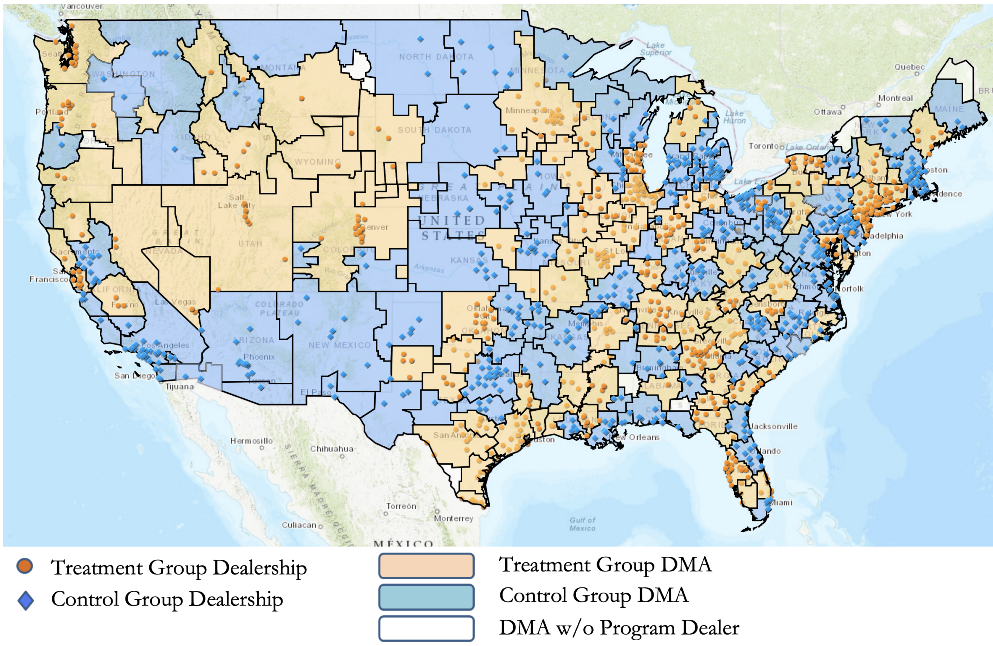
The researchers, working in conjunction with the brand managers, proposed an RCT that would guide whether this modification should be deployed. Control dealers would keep their existing post-payment system. Treated dealers would be prepaid the bonus associated with exactly meeting their 110% sales target. If this sales target were not achieved by the end of the month, the difference between their earned and prepaid bonus would be clawed back. If the sales target were instead exceeded, the dealer would receive an additional bonus payment to transfer the excess earned incentives.

Although there were reasons to believe loss-framing might improve performance, several considerations raised doubts. First, the goal of advanced payment is to motivate additional marginal effort to hit the monthly target. However, given the existing extreme motivations to hit the target, it was possible that most cost-effective marginal activities were already incentivized. Second, there were concerns that increased incentive gaming might substitute for effort at the margin. Incentive gaming is widely recognized as a concern in contracts of this nature (Oyer, 1998; Steenburgh, 2008; Larkin, 2014; Benson, 2015; Ederer, Holden and Meyer, 2018), particularly those with non-linear returns to performance. Finally, given the heterogeneity of organizational forms, management style, and sophistication, it was unclear how many dealers would respond "behaviorally." If enough firms viewed prepayment as free capital with no associated motivational effects, and if that free capital did not enable productivity-enhancing investments, then the loss-framed contracts would be inherently costly to CarCo.

In agreeing to the field experiment, CarCo granted us access to data detailing dealer-level incentives, individual vehicle sales, and location and contact information (Maritz, 2018). Due to fairness concerns, CarCo wished to avoid experimental participants in direct competition facing different treatment assignments. This led us to block-assign treatment by designated market area (DMA)—a classification of geographic areas created by Nielsen which classifies local media markets (and thus regions for advertising). Figure 2 presents a map of DMAs in the lower 48 states, their treatment assignments, and the dealers within them.

Anticipating our use of a difference-in-differences design, we randomized the treatment status of DMAs in a manner that guarantees similar pre-trends in sales. We first calculated DMA-specific sales trends using logged unit sales data between January, 2014 and February, 2017. We followed Athey and Imbens (2017) in stratifying by region and DMA size and then using a bipartite matching procedure to generate the set of matched pairs that minimized

Figure 2: Block Assignment of Dealers by DMA



Notes: This figure shows all program dealers in the lower 48 states with their block-assigned treatment group. Clear spaces reflect DMAs without program dealers.

the sum of distances between mate sales trends (Lu et al., 2011). Mates were then randomly assigned to the two conditions.

Several institutional features influenced our research design. First, again guided by fairness considerations, we were required to ensure that each participating dealer spend equal time in each experimental condition. Consequently, participants were randomly assigned to treatment or control for an initial four-month treatment window, and then conditions were flipped for a second four-month treatment window. We believe that the comparison between these two groups in the initial four months provides a clean estimate of treatment effects. In contrast, the comparison between groups after treatments are flipped may be influenced by their prior experimental assignment. In the text of this paper, we focus attention on treatment effects estimated in the initial treatment window, and when we refer to the “treatment” or “control” group we refer to assignment in that period. Analysis of the second window is relegated to the online appendix (but provides similar evidence of negative treatment effects).

Second, the franchise agreement required dealer approval for any change to the incentive program. We therefore invited dealers to opt-in to participation after assigning DMAs to condition, but without revealing that assignment to the dealers. Dealers were initially invited to participate via postings on the program tracking website. The postings described the program and asked them to either opt-in or opt-out of the pilot program. The posting explained that participating dealers would be randomly assigned to receive prepayment for four months starting either in May, 2017 or September, 2017. Dealers opting out would remain in the existing post-payment system. Multiple reminders were posted before CarCo regional sales directors attempted to directly contact the dealers in person or by phone. Of the 1,227 dealers in the incentive program, only the 294 (24%) that chose to participate were included in the experiment and subsequent analysis. Additionally, 336 explicitly opted out of the program, while 597 failed to respond. Interviews with CarCo managers and dealerships indicated that many non-participants did not want the accounting hassle of prepayment.

We present extensive analysis of selection into the study in Appendix A.3, comparing these groups with both CarCo data from before the experiment and dealer-level data from the National Establishment Time Series (NETS) Dataset (Walls & Associates, 2015). NETS data show no statistically identifiable differences among the three groups in dealer age, employee count, creditworthiness or liquidity, ownership structure, or membership in a dealer group. Participants and opt-out dealers were statistically indistinguishable in average monthly sales (48.5 vs. 43.2, $p = 0.16$), but non-respondents were considerably smaller than both (31.1, $p < 0.01$). Non-respondents were less likely to carry both model groups (86%) than participants (90%) and opt-out dealers (93%). Participation was markedly different across regions ($\chi^2 = 101.89$, $p = 0.00$), which CarCo attributed largely to regional sales representatives using different follow-up strategies to encourage participation after the standardized initial email.[12] Overall, selection into participation based on observables is modest, and due to our experimental design selection does not threaten internal validity. However, the potential for selection on unobservables remains—for example, based on anticipation of susceptibility to loss-framing or on self-assessments of elasticity to incentives—and thus caution in warranted when extrapolating our findings to new populations.

Finally, CarCo required us to change the group assignment of two DMAs due to concerns

[12]There are no differences in treatment group assignment by region ($\chi^2 = 2.46$, $p = 0.65$).

about competition spreading between neighboring DMAs. Recall that fairness concerns motivated the DMA-level blocking of treatment to avoid close competitors being assigned to different treatments. Upon observing treatment assignments, CarCo representatives identified two pairs of neighboring DMAs where they believed cross-DMA competition was a concern and where treatment assignments were different. To address this concern, one of the DMAs within each pair of neighbors had its treatment assignment exchanged with its DMA mate from the bipartite matching procedure. The net result is four DMAs having their treatment assignment modified.

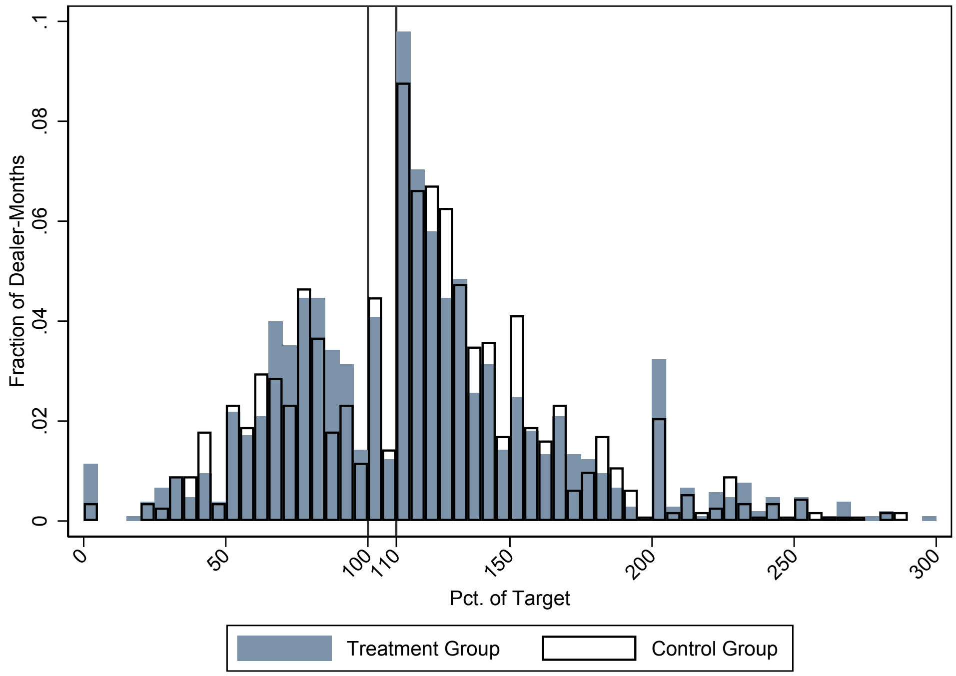
While treatment reassignments have the potential to confound experiments, several considerations suggest that this instance is comparatively benign. First, note that CarCo's intervention was not to ensure or prohibit treatment from being deployed to specific DMAs or dealers. All experimental participants experience four months under prepaid treatment and postpaid control with only the timing randomized, and CarCo's intervention was to ensure that the *timing* of treatment was the same across these particular neighbors. Second, because we flip assignment of the target DMAs and their mates in the bipartite match, the potential of this flip to proxy for some feature of the DMA is mitigated to the extent our bipartite match mates are similar. Given these considerations, we believe these flips should be treated analogously to a constraint in the randomization procedure that guarantees the same treatment assignments for these neighbors, as compared to a more selection-inducing form of experimental noncompliance.[13]

Based on this construction, our experimental sample consists of 294 dealers represented 116 DMAs in 47 states. Of the 294 dealers, 140 from 55 DMAs were assigned to the treatment group, while 154 from 61 DMAs were assigned to the control group.[14] All but 29 of these dealers sold both model groups: 18 sold only the large-bonus model group and 11 sold only the small-bonus model group. Collectively, these dealers represent monthly sales of over 15,000 vehicles and $600 million in revenue.

[13]For completeness, in Appendix Section A.5 we reproduced our main analyses while dropping the reassigned DMAs (i.e., the 2 identified by CarCo and their match mates). Results are broadly similar, although the statistical significance of our pooled estimates is weakened. Comparatively statistically strong results remain for the small-bonus model group (with p-values for the OLS and PPML specifications of 0.059 and 0.035, respectively). And the pattern of results found in the DKDD analysis is essentially identical (as would be expected given that so few observations are being removed from the distributions used in this procedure).

[14]See Appendix A.1 for the density of dealers within DMAs across conditions.

Figure 3: Pre-treatment Monthly Sales (Relative to Target) by Treatment Assignment



Notes: This figure shows the distribution of monthly sales in the four months prior to the experiment, expressed as a percentage of the assigned sales target. Model group results are included separately for each dealer-month. The two vertical lines represent the discrete bonus thresholds at 100% and 110%.

The experiment initiated on May 1, 2017, with all treatment group dealers receiving their first-month bonus prepayment. The treatment group received an average total monthly prepaid bonus of \$37,395, with averages of \$25,077 and \$13,330 for the large- and small-bonus model groups, respectively.[15] For those dealers who failed to meet the 110% target and therefore suffered clawbacks, those clawbacks averaged \$28,306, with averages of \$26,290 and \$13,564 for the large- and small-bonus model groups, respectively.

Despite researcher concerns, no dealer during our study was unable to repay unearned prepayments due to overspending and insufficient cash-on-hand. In addition, no dealer asked to be removed from the program.

[15]The two model group averages do not sum to the total average because some dealers sell only one model group.

1.3 Response to Pre-treatment Incentives

Before proceeding to analysis of the effects of our program, we document dealers' behavior in the presence of the baseline, post-paid incentive scheme.

Figure 3 shows the distribution of each monthly dealer sales outcome as a percentage of its respective target in the four months prior to the experiment. Dealers that sell both model-groups have each represented separately. This figure demonstrates the importance of the 110% target in driving sales. The incentive contracts shown in Figure 1 feature a large discontinuity in both levels (a "notch") and slopes (a "kink") occurring at the target, with both features leading to a prediction of excess mass in the vicinity of the target.[16] Such excess mass is starkly visible in this figure, with substantial asymmetries in mass observed around both the 100% and 110% target implying elastic response to incentives. Note that, in the pre-period data examined in this figure, there is no statistically detectable difference in either the average or the distribution of the plotted variable across treatment and control groups (T-test: $p = 0.702$, Kolmogorov-Smirnov: $p = 0.122$). This suggests pre-treatment balance on sales relative to targets (despite some imbalance on absolute sales that we discuss in the next section).

1.4 Summary Statistics

We now begin analysis of the sales activities in both the pre-treatment and post-treatment windows. Table 1 provides descriptive statistics on the monthly number of vehicles sold, target levels, and the rate of attaining given targets. While our ultimate analyses will be somewhat more sophisticated than the mere comparisons of means presented in this table, our results are foreshadowed by such comparisons.

Turning attention first to the large-bonus model group data for monthly vehicles sold, we note that the control group saw a modest increase in sales when comparing the pre- and post- treatment windows (39.39 vs. 41.06, an increase of 1.67). These control group increases primarily reflect seasonal demand fluctuations that are also reflected in the higher targets. For comparison, the treatment group saw a slightly larger but roughly similar

[16] See Kleven (2016) for a thorough review of bunching-based identification strategies.

Table 1: Summary Statistics

	LARGE-BONUS MODEL GROUP								
	Treatment Group			Control Group			By Participation Status		
	Pre	Post	Total	Pre	Post	Total	In	Out	All
Vehicles	29.52	31.41	30.46	39.39	41.06	40.23	35.49	27.23	29.24
Sold	(21.44)	(23.72)	(22.61)	(52.61)	(52.13)	(52.35)	(41.03)	(29.18)	(32.65)
Target	26.98	32.15	29.57	31.45	38.59	35.03	32.38	27.24	28.49
Sales	(19.36)	(22.03)	(20.89)	(33.37)	(41.69)	(37.92)	(30.98)	(24.64)	(26.41)
Hit 110%	0.56	0.43	0.50	0.65	0.50	0.57	0.54	0.46	0.48
Target	(0.50)	(0.50)	(0.50)	(0.48)	(0.50)	(0.49)	(0.50)	(0.50)	(0.50)
Hit 100%	0.62	0.48	0.55	0.70	0.55	0.62	0.59	0.53	0.54
Target	(0.49)	(0.50)	(0.50)	(0.46)	(0.50)	(0.49)	(0.49)	(0.50)	(0.50)
	SMALL-BONUS MODEL GROUP								
	Treatment Group			Control Group			By Participation Status		
	Pre	Post	Total	Pre	Post	Total	In	Out	All
Vehicles	15.14	14.16	14.65	17.22	18.80	18.01	16.37	11.07	12.35
Sold	(14.07)	(13.47)	(13.78)	(20.64)	(25.19)	(23.03)	(19.16)	(11.68)	(14.04)
Target	12.39	14.69	13.54	13.81	16.86	15.34	14.46	10.88	11.74
Sales	(9.91)	(11.92)	(11.02)	(13.37)	(17.00)	(15.36)	(13.45)	(9.90)	(10.96)
Hit 110%	0.64	0.44	0.54	0.64	0.48	0.56	0.55	0.48	0.50
Target	(0.48)	(0.50)	(0.50)	(0.48)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)
Hit 100%	0.70	0.50	0.60	0.70	0.53	0.62	0.61	0.56	0.57
Target	(0.46)	(0.50)	(0.49)	(0.46)	(0.50)	(0.49)	(0.49)	(0.50)	(0.49)

Notes: Summary statistics of monthly sales performance by model group, treatment, and participation status. Means and standard deviations presented in all cells. The first two rows present monthly sales and monthly target thresholds. The rows below summarize the probability of hitting earning the larger fixed bonus for exceeding the 110% threshold, the smaller fixed bonus for exceeding the standard target threshold. The seemingly identical pre-period small-bonus model group percentages for the two participant groups is due to rounding, and they are slightly different at the third decimal point.

increase (29.52 vs. 31.41, an increase of 1.89; difference-in-differences: 0.22). An analogous comparison of the rate of achieving the 110% target suggests that treatment increases the chance of earning the large fixed component of the bonus by 2 percentage points.[17] In short, simple comparisons of means suggest that the treatment was associated with quite modest increases in the large-bonus model group's sales and bonus-condition attainment.

Turning attention next to the small-bonus model group data for monthly vehicles sold, we note that the control group again saw a modest increase in sales when comparing the pre- and post- treatment windows (17.22 vs. 18.80, an increase of 1.58). For comparison, the treatment group saw a decrease in sales (15.14 vs. 14.16, a decrease of 0.98; difference-in-differences: -2.56). An analogous comparison of the rate of achieving the 110% target suggests that treatment decreases the chance of earning the large fixed component of the bonus by 4 percentage points.[18] In short, a simple comparison of means suggests a comparatively large negative effect of loss framing for the small-bonus model group's sales and bonus-condition attainment.

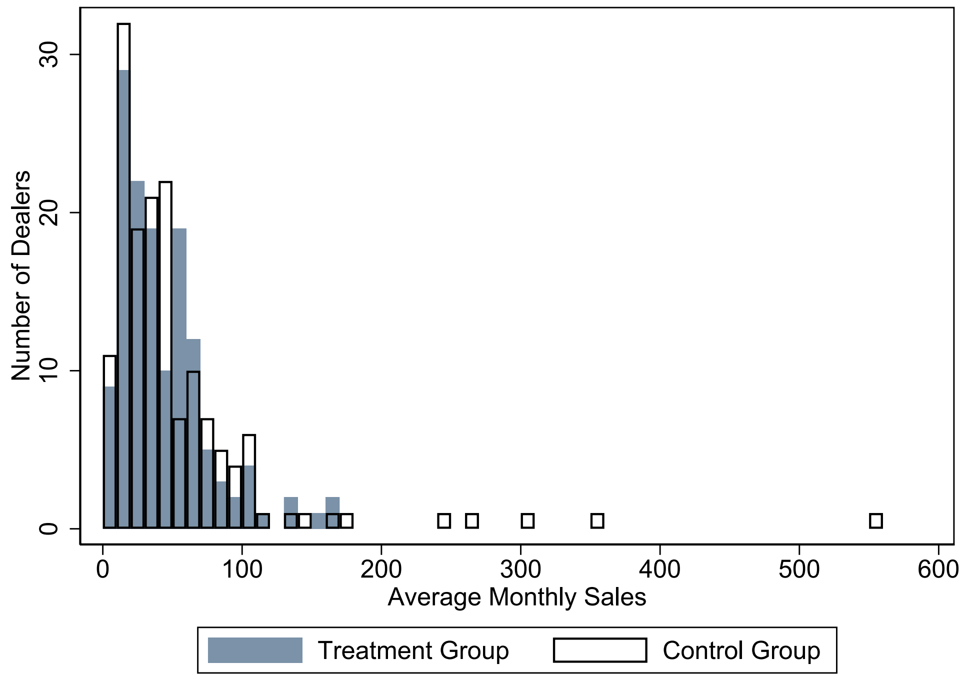
Examination of Table 1 also illustrates two key issues that will prompt robustness analyses.

First, even prior to treatment, the means of key variables (e.g., sales) are significantly different between the treatment and control groups. These differences may generate some concern about systematic problems with random assignment, but they can be attributed entirely to the presence of a small number of influential outliers in the control group. Five of the six largest dealers, who sell over five times the monthly volume of the average dealer, are in one DMA and thus block-assigned. The exclusion of this one DMA renders the two groups statistically indistinguishable in pre-period monthly sales. The largest dealer, which has ten times the monthly sales of the average dealer, was by chance also assigned to the control group. Figure 4 shows the distribution of average dealer monthly sales in the window prior to experimental intervention, illustrating broadly similar distributions with the exception of these outliers. These issues partially motivate the inclusion of dealer fixed effects in our

[17]Control: 0.65 vs. 0.50, a decrease of 0.15. Treatment: 0.56 vs. 0.43, a decrease of 0.13. Difference-in-differences: 0.02.

[18]Control: 0.64 vs. 0.48, a decrease of 0.16. Treatment: 0.64 vs. 0.44, a decrease of 0.20. Difference-in-differences: -0.04.

Figure 4: Pre-period Average Monthly Car Sales by Treatment Group



Notes: This figure shows histogram of average monthly sales for each dealer (summing across both model groups) in the four-months prior to the experiment. Five of the six largest dealers are in one DMA (and thus are block-assigned). This block assignment accounts for the unbalanced right-tail of control group observations.

main analyses. Additionally, in Appendix Section A.4, we reproduce our main results while excluding this outlier DMA and its matched mate and show that this exclusion minimally affects our reported results.

Second, sales and target attainment show clear trends in the control group. This non-stationarity is expected due to the significant seasonality in automobile sales, and illustrates the importance of our difference-in-differences design.

1.5 Impact of Loss Framing on Average Sales

In this section we examine the average effect of loss framing during the first four months of the experiment, when the 140 treatment group dealers received the prepayment treatment while 154 control group dealers retained the pre-existing post-payment scheme. Sales behavior in the four months prior to experimental assignment serves as the pre-period in our estimation data.

We consider two regression specifications:

$$E[Y|d, m, \text{Treated}_d, \text{Treatment period}_m] = \beta \times \text{Treated}_d \times \text{Treatment period}_m + \mu_m + \nu_d \tag{1}$$

$$ln(E[Y|d, m, \text{Treated}_d, \text{Treatment period}_m]) = \beta \times \text{Treated}_d \times \text{Treatment period}_m + \mu_m + \nu_d \tag{2}$$

Regression specification 1 is a standard OLS implementation of a difference-in-differences design.[19] The conditional expectation of the total number of cars sold by dealer d in month m is determined by the interaction between an indicator of whether the dealer was assigned to the treatment group (Treated_d) and an indicator for whether the current month m is in the window when treatment is active ($\text{Treatment period}_m$), month-specific fixed effects (μ_m), and dealer-specific fixed effects (ν_d). In this framework, β measures the treatment effect of interest in units of additional cars sold.

Regression specification 2 is a standard Poisson regression implementation of a difference-in-differences approach.[20] As illustrated in the equation, the same set of predictors and fixed effects are included. The key distinction is that, in this framework, coefficient β measures the treatment effect in log units of sales. When discussing these results, we will rely on the approximation that $ln(1 + \beta) \approx \beta$ for small β, and will thus convert the log-point estimates into effects expressed in percentages.

Throughout our analyses, all standard errors are clustered at the DMA level.

Table 2 presents these results. Columns 1–2 present results for total sales across both model groups, where the first column reports the OLS specification and the second column reports the Poisson specification. Columns 3–4 and 5–6 present these same regressions restricted to the large-bonus and small-bonus model groups, respectively.

The average treatment effect in column 1 implies that 2.31 fewer cars were sold per month under loss framing. Note that, due to the comparatively few clusters in our analysis arising

[19] Note that the necessary parallel trends assumption for this model is supported by the stratified matching randomization in our experimental design, and is visually presented in Appendix A.2.

[20] We apply Poisson regression rather than OLS with a logged dependent variable because some dealerships have months with 0 sales.

Table 2: Difference-in-Differences Estimates of Treatment Effect on Sales

	(1) OLS	(2) PPML	(3) OLS	(4) PPML	(5) OLS	(6) PPML
$\text{Treated}_d \times \text{Treatment period}_m$	-2.31	-0.038	0.14	0.019	-2.55	-0.154
	(1.34)	(0.022)	(0.94)	(0.025)	(1.40)	(0.060)
Model Group	Pooled	Pooled	L-B	L-B	S-B	S-B
Month Fixed Effects	X	X	X	X	X	X
Dealer Fixed Effects	X	X	X	X	X	X
Observations	2352	2352	2261	2261	2216	2216

Notes: This table presents regressions predicting the dealer/month-specific number of vehicles sold. Odd numbered columns present OLS estimates of equation 1. Even numbered columns present Poisson Pseudo-Maximum Likelihood (PPML) estimates of equation 2. We first present analysis pooling the two model groups together (cols 1 and 2), followed by analysis of large-bonus model group (cols 3 and 4) and small-bonus model group (cols 5 and 6). Standard errors are clustered at the DMA level.

from DMA-level block assignment, these estimates are not extremely precise. Despite this imprecision, the null hypothesis of no effect can be rejected at the 10%, but not the 5% α-level ($p = 0.086$). The Poisson model in column 2 produces similar results, with a 4% decrease in sales and similar statistical significance ($p = 0.081$).

We next examine the differential effect this policy had on the two model groups under consideration. Note that in columns 3–4 of Table 2, we find no strong indications of effect of the policy on sales for the large-bonus model group. Point estimates suggest an impact of 0.14 car sales per month, or 2%, but effects are far from any common significance thresholds ($p = 0.882$ and $p = 0.463$, respectively). In contrast, the estimated treatment effect for the small-bonus model groups are statistically stronger and quantitatively large: -2.55 cars sold per month ($p = 0.071$) or -15% ($p = 0.010$).

In light of the standard prediction that loss-framed contracts motivate additional productivity, these findings may be viewed as surprising. There is little indication of this positive effect, and indeed all but extremely modest positive effects fall outside of the confidence intervals of our pooled estimates.[21] Instead, our estimates tentatively point towards overall *negative* impacts of loss framing, although with substantial imprecision in our estimates and

[21] Based on our estimates from column 1 and 2, we may reject any positive treatment effect in excess of 0.33 cars a month or 0.47% of sales at a 5% α-level.

significance only at the lower-than-usual 10% α-level. Our estimates additionally suggest that the overall negative effect on joint sales is driven by a large reduction in small-bonus model group sales. Large-bonus model group sales, in contrast, saw a small and not meaningfully offsetting gain. This pattern will be more fully explored in the following section.

1.6 The Impact of Loss Framing on Distribution of Sales

Having assessed the impact of loss framing on average sales, we now turn to assessing its impact on distribution of monthly sales. The pre-experiment baseline distribution represented in Figure 3 shows that under the original post-payment scheme, dealer attention toward the 110% target was reflected in missing sales density just below the target and excess density above. Importantly, it also showed indistinguishable distributions between the two experimental groups. Our goal now is to evaluate how these distributions were affected by loss framing. By analyzing the full distribution, we may assess more nuanced predictions regarding the impact of loss framing: most centrally, that changes in behavior induced by the treatment should occur among dealers "close" to hitting the monthly target.

1.6.1 Difference in Kernel Density Differences Estimation (DKDD)

In order to infer the patterns of excess and missing mass induced by loss framing, we develop a methodology for density estimation closely related to standard difference-in-differences approaches. Conceptually, this approach may be thought of in two steps: (1) estimating the evolution of the distribution of final sales between the pre-period and the treatment window, and (2) estimating the difference in such evolutions between the treatment and control group. Under a generalization of the typical parallel-trend assumption—now requiring parallel evolution of full distributions rather than means—this provides an estimate of the causal impact of treatment on the distribution of final sales achieved.

Formally, consider a univariate, independent sample $\{x_n^{(g,t)}\}_{n\in N}$. In this notation, subscript n denotes the observation (out of a total set of N), the superscript g denotes two groups, and the superscript t denotes two time periods. This variable is distributed according to group-and-time-specific unknown densities $f^{(g,t)}$. Let the group-and-time-specific number of observations be denoted by $N^{(g,t)}$.

In the absence of group-specific intervention, densities are assumed to evolve in a manner that satisfies the assumption $f^{(g,2)} - f^{(g,1)} = f^{\Delta}$ for all groups g. f^{Δ} thus denotes the manner in which mass is shifted in the density function over time. However, if treatment is applied to one group (denoted T, with the other denoted C for control), the distribution that arises is represented as $f^{(T,2)} = f^{(T,1)} + f^{\Delta} + f^{T}$. The term f^{T} denotes an additional redistribution of mass induced by the treatment, and estimation of this term is the goal of this exercise.

Given these assumptions, a close analog to the common difference-in-differences estimator for means immediately arises:

$$f^{T} = (f^{(T,2)} - f^{(T,1)}) - (f^{(C,2)} - f^{(C,1)}). \tag{3}$$

Conceptually, we may examine the impact of treatment by examining how the distribution in the treatment group changed over time, differencing out the changes that may be attributable to the mere passage of time as inferred by the changes occurring in the control group.

Our formal estimate of f^{T} arises through the application of the analog principle (Goldberger, 1968; Manski, 1986), substituting finite-sample estimates of these densities for the population densities themselves. Given a kernel function K and a bandwidth h, define the kernel density estimator to be $\hat{f}_h^{(g,t)}(x) = \frac{1}{hN^{(g,t)}} \sum_{n=1}^{N^{(g,t)}} K(\frac{x - x_n^{(g,t)}}{h})$.[22] Given these definitions, the Difference in Kernel Density Differences (DKDD) estimator of f^{T}, evaluated at point x with bandwidth h, is given by

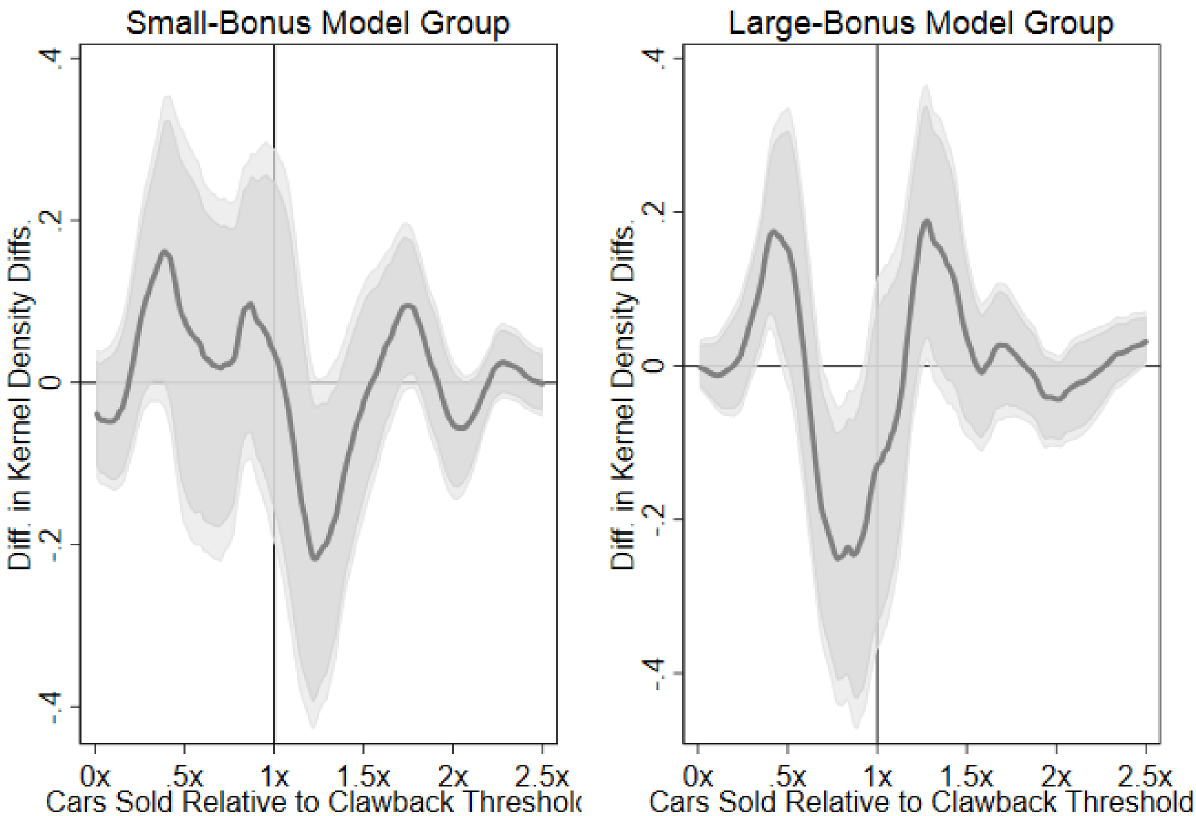
$$\hat{f}_h^{T}(x) = (\hat{f}_h^{(T,2)}(x) - \hat{f}_h^{(T,1)}(x)) - (\hat{f}_h^{(C,2)}(x) - \hat{f}_h^{(C,1)}(x)). \tag{4}$$

The consistency of equation 4 as an estimator for f^{T} follows immediately from the well-known consistency of the individual kernel estimators and Slutsky's theorem.

In the next section, we apply this method to infer f^{T} as induced by loss framing. When-

[22] While we have chosen to use kernel-based methods, the approach outlined above can be applied with a variety of alternative approaches to density estimation. Our DKDD estimator may be considered a special case in a broader class of "difference in density differences" estimators. See Kuhn and Yu (2021) for an application of this approach to histogram bins rather than kernel density estimates.

Figure 5: The Impact of Loss Framing on the Distribution of Sales



Notes: This figure shows difference-in-kernel-density-differences estimates of the impact of loss framing on the distribution of sales achieved. The x-axes show the amount of sales relative to clawback threshold. The two subfigures present estimates derived from each model group. Confidence regions, based on 10,000 bootstrap iterations resampled by DMA, are shaded. The 90% confidence region is shaded darkly and the 95% confidence region is shaded lightly. Kernel: Epanechnikov; Bandwidth: .1x; Sample Sizes: 2,175 (left panel) and 2,234 (right panel).

ever presented, confidence intervals will be generated through a bootstrap procedure, re-calculating $\hat{f}^{(g,t)}(x)$ from 10,000 simulated samples generated by sampling by DMA cluster (with replacement).

1.6.2 DKDD Estimates of the Distributional Impact of Loss Framing

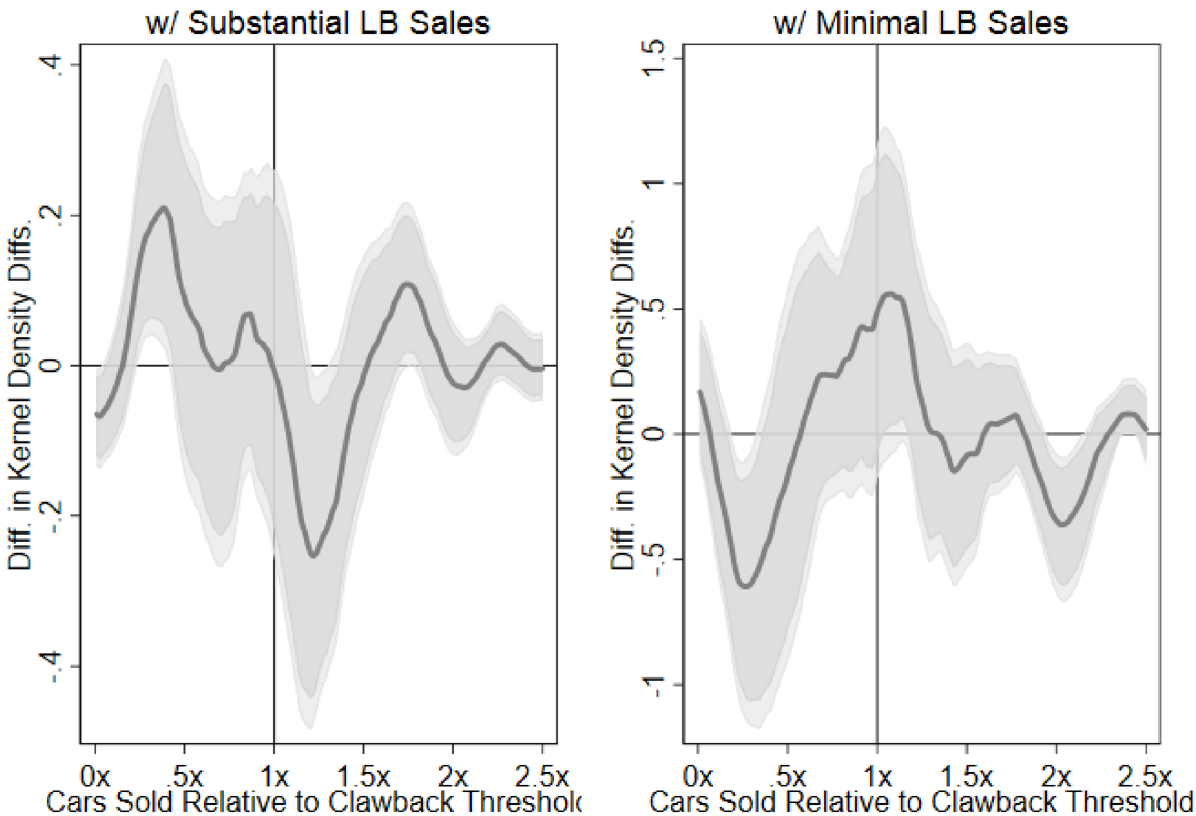
Figure 5 presents our estimates of the treatment effect of loss framing on the distribution of final sales achieved, measured relative to the clawback threshold (i.e., the sales necessary to reach the 110% sales target). In this figure, positive (negative) values indicate that treatment increases (decreases) the probability of ending the month with sales at the x-axis location.

The left panel presents the DKDD estimate for the small-bonus model group. Results are clearly in line with the earlier estimate of a negative impact of loss framing on average sales. Two patterns are particularly notable. First, the positive estimates in the region to the left of the clawback threshold indicate that there is an increase in the probability that dealers fail to achieve their 110% sales target (although, as indicated by the plotted confidence regions, at any individual point these positive estimates are statistically relatively weak). Second, the negative estimates just to the right of the clawback threshold indicate that dealers are less like to achieve sales just past their 110% sales target. Other than this sharp decline in the probability of ending the month with sales just over the threshold, this figure does not display sharp evidence of excess or missing mass at particular points, and instead might better be characterized as general and non-localized shift from higher to lower values. This shift downward is consistent with the 4% decline in 110%-sales-target attainment reported in Section 1.4.

The right panel presents results for the large-bonus model group. Recall that we reported a null effect of loss framing on average sales for this model group. In this figure, we see that the previous null finding was masking statistically significant effects across the distribution of sales. Most notably, there is a large decline in the probability that the treated dealers end the month in the region just under the clawback threshold, relative the control group. The redistribution of mass indicates that dealers who would otherwise have just triggered a clawback changed their behavior in one of two ways. Some dealers proceeded in a manner that resulted in more cars being sold, resulting in the observed excess mass of sales amounts just over the bonus threshold. This excess mass drives the 2% increase in target attainment reported in Section 1.4, and is consistent with these dealers directing more attention, effort, or resources to these sales to avoid triggering the loss of the large bonus. However, some dealers proceeded in a manner that resulted in fewer cars being sold, resulting in the observed excess mass of sales amounts substantially below the bonus threshold. In this figure, these two groups are approximately equally sized, demonstrating why they average to a null effect when combined in the earlier difference-in-differences analysis.

We have suggested that some of our results might be explained by loss-framing leading

Figure 6: The Impact of Loss Framing on Small-Bonus Model Group Sales, Conditional on Large-Bonus Model Group Sales



Notes: This figure shows difference-in-kernel-density-differences estimates of the impact of loss framing on the distribution of small-bonus model group sales achieved. The x-axes show the amount of sales relative to clawback threshold. The two subfigures present estimates conditioning on whether that month's sales of the large-bonus model group were comparatively high (over 50% of the clawback threshold) or low (under 50% of the clawback threshold). Confidence regions, based on 10,000 bootstrap iterations resampled by DMA, are shaded. The 90% confidence region is shaded darkly and the 95% confidence region is shaded lightly. Kernel: Epanechnikov; Bandwidth: .1x; Sample Sizes: 1,931 (left panel) and 244 (right panel).

to a reassessment of multitasking behavior. In essence, the loss-framing intervention could lead a dealer to redeploy effort or resources from the sale of the small-bonus model group to the sale of the large-bonus model group. Such a story might be natural if dealers become more attentive to, or concerned about, reaching the necessary sales target for large-bonus model group when loss framing is active. Such a story could explain the overall negative results for the small-bonus model group through cannibalization of its sales.

To partially probe this possibility, we assess how estimated treatment effects change when these potential incentives for cannibalization of the small-bonus model group are in effect "shut down." To do so, we isolate a minority of dealer-months in which the dealer fell far short of the large-bonus model group target.[23] We will assume that, in these cases, the dealers understood that diverting resources from small-bonus model group sales would not help push their large-bonus model group sales over the clawback threshold. If this is the case, then the effect of loss-framing would not operate through motivating that diversion of resources. We operationalize the notion of "falling far short" of the large-bonus model group target by flagging dealers who sold less than or equal to 50% of their large-bonus model group clawback threshold—a mere 11% of dealer-months fall in this group.

Figure 6 presents the relevant analyses. As a baseline, the left panel presents our DKDD estimate of the impact of loss framing on small-bonus model group sales when dealers with minimal large-bonus model group sales are excluded. Compared to results from the full sample (previously presented in Figure 5), the pattern of effects is essentially unchanged. However, among the minority of responses with minimal large-bonus model group sales, a substantially different pattern emerges. The right panel presents our DKDD estimate of the impact of loss framing on small-bonus model group sales among dealers who demonstrably were not in contention for bonus attainment for the large-bonus model group. Here, we see that loss-framing was associated with a higher probability of ending the month with small-bonus model group sales in the vicinity of the clawback threshold, and a lower probability of ending the month with sales substantially below that threshold. This pattern of results is consistent with a shift from lower to higher sales amounts, with an associated increase in target attainment. In short, when incentives to redirect resources to large-bonus model group sales are "shut down," the impact of loss-framing on small-bonus model group sales appears straightforwardly positive and in line with standard intuitions.

1.7 Summary and Interpretation of Empirical Results

Taking stock, we have shown that the overall effect of introducing loss framing was a marginally significant *reduction* of average monthly sales. This negative effect was driven by

[23] We include the small minority of dealers who sell only the small-bonus model group in this group.

a reduction of average sales for the small-bonus model group, whereas no effect was detected for the large-bonus model group. However, examining the manner in which distributions of sales changed, we observe that dealers facing loss framing were significantly less likely to end the month having narrowly missed avoiding clawback of the bonus for the large-bonus model group. Some of these dealers are estimated to have instead sold more vehicles from this model group, consistent with increasing or diverting effort and resources towards this model group's sale. Other dealers are estimated to have sold fewer vehicles from this model group. Furthermore, when isolating dealers who appear to not be pursuing target attainment for the large-bonus model group, treatment effects of loss-framing on the small bonus model group are straightforwardly desirable.

These results suggest that one consequence of loss framing was a rebalancing of effort and resources across the two model groups. On average, the model group with the smaller potential loss was neglected, potentially due to effort or resources being directed to help avoid the larger potential loss, and with this pattern reversing when the large potential loss is unavoidable. In Section 2, we present and analyze a model that illustrates this possibility.

1.8 Robustness Considerations and Additional Results

In the Appendix, we assess a variety of robustness considerations and present additional supporting results. These include assessment of the correlates of opting-in to the experiment (see Appendix A.3); the demonstration that our primary estimates are robust to the exclusion of the influential outliers documented in Figure 4 (see Appendix A.4); the demonstration that our primary estimates are robust to the exclusion of the four DMAs whose treatments were reassigned (see Appendix A.5) the estimation of treatment effects in the second treatment window based on a synthetic control approach utilizing experimental non-participants (see Appendix A.6); a summary of results from post-experiment interviews and surveys (see Appendix A.7); and an assessment of the timing of sales (see Appendix A.8). Ultimately, none of these analyses contradict the summary of results presented above, but at times they offer additional context and illustration of underlying dealer behavior.

2 A Theoretical Examination of Loss Framing

In this section we present a model that clarifies how inducing loss-averse evaluation of incentives through loss framing can help or hinder productivity. We use this model to demonstrate the ways in which past literature has incompletely characterized the predicted impact of loss-framing interventions, and use these results to consider situations in which prevailing wisdom is expected to hold or is at risk to fail.

In our model, an agent sells items on behalf of a principal. The principal benefits from the agent's sales, and the agent faces private costs (or benefits) from sales activities. To mitigate the divergence in incentives, and in particular the incentive to under-invest in sales activities due to private costs, the principal has provided the agent with a bonus contract that yields direct payments for the amount of sales completed. There are potentially multiple dimensions of sales activity, each with its own incentive contract.

Using this model, we consider the wisdom of applying loss framing, which we assume operates by inducing loss aversion relative to a target level of sales. We characterize sufficient conditions for cases in which alternative sales approaches will be made desirable as compared to the choice made in the absence of loss aversion. As we will illustrate, the alternative strategies that are incentivized can reduce expected sales under relatively common conditions.

While we use the terminology of "sales" to be explicit about the link to worker activities in the empirical section of our paper, note that this labeling is in no way essential. This theory can be understood to apply to broad notions of (observable) performance that a principal aims to motivate among his workers.

2.1 Seller's Decision Problem

The agent, or *seller*, has a window of time over which outcomes are evaluated (e.g., one month). During this window, the seller completes s^d sales and incurs c^d in private costs of sales activity for each dimension $d \in D$. Let $\boldsymbol{s}$ and $\boldsymbol{c}$ denote the vectors of these values across all dimensions.

When assessing each dimension of sales activity, the seller's ex-post utility is determined by $u^d(\phi^d(s^d), c^d)$, where $\phi^d(s^d)$ denotes the payment received for completing s^d sales and c^d denotes the net private costs that the seller incurs in the sales process. The incentive scheme $\phi^d(s^d)$ is meant to promote the pursuit of sales, so $\phi^d(s_1^d) \geq \phi^d(s_2^d)$ if $s_1^d > s_2^d$. Utility is assumed to be increasing in payments ($\phi^d(s^d)$) and decreasing in costs (c^d).

Over the window of time considered, the seller faces a long sequence of individual decisions of how to manage the sales process. The seller's decision is the choice of a *selling strategy*, which constitutes a complete contingent management plan over the window in question. Denote an individual selling strategy as σ, and the (finite) set of possible selling strategies as Σ. The decision-relevant consequence of the choice of a selling strategy is the determination of the joint distribution of sales ($\boldsymbol{s}$) and private costs ($\boldsymbol{c}$). Denote this joint distribution by $f(\boldsymbol{s}, \boldsymbol{c}|\sigma)$, and marginal distributions by, e.g., $f(\boldsymbol{s}|\sigma)$.

The seller's decision problem is therefore to choose $\sigma \in \Sigma$ to maximize expected utility, given by $\mathbb{E}[U|\sigma] = \int U(\boldsymbol{s}, \boldsymbol{c})\, df(\boldsymbol{s}, \boldsymbol{c}|\sigma)$, where $U(\boldsymbol{s}, \boldsymbol{c}) = \sum_{d \in D} u^d(\phi^d(s^d), c^d)$.

2.2 Introducing Loss Framing

The principal considers a modification to the incentive contract that introduces loss framing (e.g., through a clawback scheme). We assume that the consequence of introducing loss framing is to induce loss aversion relative to a dimension-specific target level of sales (R^d). If loss aversion is induced, the seller's utility function is modified to be $U^\Lambda(\boldsymbol{s}, \boldsymbol{c}) = \sum_{d \in D} u^d(\phi^d(s^d), c^d, \boldsymbol{z}^d) - \Lambda(\phi^d(R^d) - \phi^d(s^d)) \cdot I(\phi(s^d) < \phi(R^d))$. I denotes the indicator function, taking a value of 1 if the statement in parentheses is true and 0 otherwise. In this utility specification, for cases when the target is exceeded—a gain—utility is identical to the prior formulation, but for cases when the target is not met—a loss—an additional cost proportional to the lost incentive is imposed. The magnitude of this additional cost is governed by the excess weight placed on losses, $\Lambda \in \mathbb{R}^+$. This specification arises from the movement of reference points from 0 to R^d under the assumption of piecewise-linear gain/loss utility.[24]

[24] Formally, assume dimension-specific utility is governed by $f_1^d(\phi^d(s^d), c^d) + \eta f_2^d(\phi^d(s^d) - \phi^d(R^d))$, with $f_2^d(\phi^d(s^d) - \phi^d(R^d)) = (\phi^d(s^d) - \phi^d(R^d)) + (\lambda - 1) \cdot (\phi^d(s^d) - \phi^d(R^d)) \cdot I(\phi^d(s^d) < \phi^d(R^d))$. This model includes arbitrary "standard" utility represented by f_1^d, augmented with the additional consideration of f_2^d allowing for a piece-wise linear version of prospect theory. In this formulation, the coefficient of loss aversion is λ and the decision-weight placed on prospect theory is η. Translating this model into the version in text, define $u^d(\phi^d(s^d), c^d)$ to incorporate both standard utility as well as utility derived from gains/loss evaluation

In order to assess the impact of loss framing, we will consider how its introduction influences choices between two selling strategies. Before presenting our result, we begin with a simple example that helps illustrate the potential for negative consequences.

2.2.1 Example 1: Reducing Sales to Avoid a Loss

In this example, consider a seller who sells two types of goods, referred to as good 1 and good 2. Both goods are subject to the same incentive contract, $\phi(s) = s$, and the same dimension-specific utility function $u(\phi(s^d), c^d) = \phi(s^d) - c^d$. Throughout this example, we will assume there are no private costs of sales: $c = 0$ for all dimensions and in all strategies.

In this example, the seller has two potential selling strategies. Strategy 1 yields 10 sales of each good. Strategy 2 yields 9 sales for good 1 and 12 sales for good 2.

Absent loss framing, the seller's utility is higher pursuing strategy 2:

$$U(\boldsymbol{s}_1, \boldsymbol{c}_1) = 10 + 10 = 20 < 21 = 9 + 12 = U(\boldsymbol{s}_2, \boldsymbol{c}_2)$$

Now consider the consequence of the principal applying loss framing relative to a sales target of 10 for each good. Further assume that $\Lambda = 2$. This modifies the utility evaluation to:

$$U^{\Lambda}(\boldsymbol{s}_1, \boldsymbol{c}_1) = 10 + 10 = 20 > 19 = 9 + 12 - \Lambda \cdot 1 = U^{\Lambda}(\boldsymbol{s}_2, \boldsymbol{c}_2)$$

In this example the expected-sales maximizing strategy is chosen in the absence of loss framing, but is not chosen when loss framing is introduced. Choosing strategy 2 over strategy 1 effectively entails trading the opportunity to sell 1 unit of good 1 for the opportunity to sell 2 units of good 2. For the purposes of maximizing expected sales, this is a desirable trade.

without excess weighting of losses: $f_1^d(\phi^d(s^d), c^d) + \eta(\phi^d(s^d) - \phi^d(R^d))$. Defining $\Lambda = \eta \cdot (\lambda - 1)$, the equivalence of the models then follows. Note that in this formulation, we have assumed that gains and losses are evaluated with respect to the difference in payments relative to reference level. Some applications of prospect theory would instead assume that gains and losses are evaluated with respect to the utility difference relative to the reference level. In practice, this assumption will matter little for our results: conceptually similar results can be generated under the alternate assumption. An advantage of the current approach is that, by assuming that the magnitude of loss is influenced by lost payments as opposed to lost utility, we need not impose much structure on the direct utility function u. In contrast, if utility evaluations served as an input to gain/loss evaluation, substantially more structure would need to be imposed.

However, when framed as a loss, forgoing that 1 unit of good 1 is viewed as more costly. For a sufficiently loss averse seller, trading a one unit loss for a two unit gain is undesirable.

This simple example captures the core intuition that arises in the theory to come. Stated most simply, applying loss framing incentivizes the seller to avoid losses. If strategies are available that avoid losses at the cost of expected sales, then applying loss framing increases the incentives to pursue those strategies.

2.3 Characterizing the Impact of Loss Framing

In order to formalize the claims made in interpreting the previous example, we require a means of characterizing when strategies assist in avoiding losses. Define the *loss exposure of strategy* σ as

$$L(\sigma) = \int \sum_{d \in D} (\phi^d(R^d) - \phi^d(s^d)) \cdot I(\phi^d(s^d) < \phi^d(R^d)) \, df(\boldsymbol{s}, \boldsymbol{c}|\sigma). \tag{5}$$

In words, the loss exposure of a strategy is the expected magnitude of dimension-specific losses that arises from the distribution of outcomes that the strategy induces, summed across dimensions.

Using this definition, we may completely characterize the types of strategies that are incentivized by loss framing. When characterizing the strategies that are incentivized, we will refer to loss aversion making strategy σ_1 more attractive relative to σ_2 if $\mathbb{E}[U^\Lambda|\sigma_1] - \mathbb{E}[U^\Lambda|\sigma_2] > \mathbb{E}[U|\sigma_1] - \mathbb{E}[U|\sigma_2]$. For ordinal evaluation, we say that strategy σ_1 is preferred to strategy σ_2 if $\mathbb{E}[U|\sigma_1] > \mathbb{E}[U|\sigma_2]$, with the relevant definition of individual utility (U or U^Λ) applied.

Proposition 1 (The impact of inducing loss framing)**.** *Consider two selling strategies, σ_1 and σ_2. Inducing loss framing makes σ_1 more attractive relative to σ_2 if and only if $L(\sigma_1) < L(\sigma_2)$. Furthermore, for sufficiently large values of Λ, σ_1 is preferred to σ_2 if $L(\sigma_1) < L(\sigma_2)$ and only if $L(\sigma_1) \leq L(\sigma_2)$.*

Proof. Note that expected utility with loss framing may be expressed as $\mathbb{E}[U^\Lambda|\sigma] = \mathbb{E}[U|\sigma] - \Lambda \cdot L(\sigma)$. From this formulation, it follows immediately that the difference in expected utility resulting from σ_1 and σ_2 is $\mathbb{E}[U^\Lambda|\sigma_1]-\mathbb{E}[U^\Lambda|\sigma_2] = (\mathbb{E}[U|\sigma_1]-\mathbb{E}[U|\sigma_2])-\Lambda(L(\sigma_1)-L(\sigma_2))$. The first term of this expression is simply the utility difference experienced absent loss aversion, and the second term is the difference in exposure to losses scaled by the excess weight on losses. This second term is positive if and only if $L(\sigma_1) < L(\sigma_2)$, establishing the first claim of the proposition.

We next establish the claim that, for sufficiently large Λ, σ_1 is preferred to σ_2 if $L(\sigma_1) < L(\sigma_2)$. Consider an arbitrary pair of strategies drawn from Σ, denoted σ_i and σ_j, and assign labels such that $L(\sigma_i) < L(\sigma_j)$. Note that $\mathbb{E}[U^\Lambda|\sigma_i] > \mathbb{E}[U^\Lambda|\sigma_j]$ if and only if $\Lambda > \frac{\mathbb{E}[U|\sigma_i]-\mathbb{E}[U|\sigma_j]}{L(\sigma_i)-L(\sigma_j)} \equiv T_{i,j}$. Thus, if we define "sufficiently large" to mean $\Lambda > \max_{i,j} T_{i,j}$, for arbitrarily drawn pairs we are assured that $\Lambda > \frac{\mathbb{E}[U|\sigma_i]-\mathbb{E}[U|\sigma_j]}{L(\sigma_i)-L(\sigma_j)}$ and thus that σ_i is preferred to σ_j. The claim then follows.

The final claim, that for sufficiently large Λ σ_1 is preferred to σ_2 only if $L(\sigma_1) \leq L(\sigma_2)$, follows immediately from the claim just proved. $\square$

Proposition 1 fully characterizes the impact of inducing loss aversion in performance incentives. When loss aversion is induced, incentives to reduce exposure to losses are increased. If a given strategy has a lower degree of loss exposure, a sufficiently loss-averse seller can be motivated to choose it if the principal imposes loss framing. In situations where the set of strategies involve tradeoffs between loss exposure and expected sales, this drives the potential for negative consequences like those illustrated in the leading example above.

2.4 Reassessing Unidimensional Results

The results of the prior section establish that, in a relatively minimalist and general model, the performance effects of loss framing can be negative. This finding may be viewed as surprising in light of existing literature. It is in opposition to commonly applied intuitions regarding loss framing, and it would seem to contradict some formal results. This begs the question: what feature of our modeling approach drives these differences in conclusions? Since prior works have focused on unidimensional environments, and given our focus on the exacerbation of incentives for inefficient multitasking, one might be tempted to infer that the inclusion of additional dimensions is purely responsible for the differences. In this section we will document, however, that conceptually similar results arise in unidimensional environments.

2.4.1 Example 2: Reduced Sales in a Unidimensional Stochastic Setting

In this example, a seller is considering his approach to selling a single type of good. As in Example 1, assume that the incentive contract is $\phi(s) = s$ and the utility function is $u(\phi(s), c) = \phi(s) - c$ (with superscript d surpressed in this unidimensional case). Again assume that $\Lambda = 2$.

The seller has two potential strategies. Strategy 1 yields 10 sales with certainty. Strategy 2 results in an amount of sales drawn from the integers 7-14, each with an equal probability of occurring. This yields an expectation of 10.5 sales. Both strategies require the seller to incur 5 units of effort costs.

Absent loss framing, the seller's utility is higher pursuing strategy 2:

$$\mathbb{E}[u(\phi(s_1), c_1)] = 10 - 5 = 5 < 5.5 = 10.5 - 5 = \mathbb{E}[u(\phi(s_2), c_2)]$$

If the principal imposes loss framing relative to a sales target of 10, this modifies the utility evaluation to:

$$\mathbb{E}[u^{\Lambda}(\phi(s_1), c_1)] = 10 - 5 - \Lambda \cdot L(\sigma_1) = 5 > 4 = 10.5 - 5 - \Lambda \cdot L(\sigma_2) = \mathbb{E}[u^{\Lambda}(\phi(s_2), c_2)]$$

Note that strategy 1 has no loss exposure, so $L(\sigma_1) = 0$. Strategy 2, however, has the potential for losses of size 3, 2, or 1 if the realization of sales is 7, 8, or 9, respectively. Weighting each of these losses by its probability of occurring ($\frac{1}{8}$) yields $L(\sigma_2) = \frac{6}{8}$, generating the inequality above. The worker would therefore choose strategy 1 with loss aversion induced.

In this example, loss framing again leads the seller to pursue a strategy that does not maximize expected sales. As with Example 1, the strategies available involve a tradeoff between loss exposure and expected sales. In Example 1, avoidance of loss exposure was pursued through "gaming" of sales pursuit across dimensions. In this example, avoidance of loss exposure was pursued through "gaming" of tolerated risk. This illustrates that similar concerns described above can extend into the unidimensional environments considered in the literature as long as risk/reward tradeoffs are present.

2.5 A Modified Statement of the Conventional Wisdom

Example 2 demonstrates that the conceptual issues considered in this paper extend even to unidimensional environments that have been the focus of this literature. The element that is essential for a potential of negative consequences of loss framing is not multidimensionality, but rather the presence of strategies that trade off greater loss exposure for greater expected productivity. In the unidimensional environment, this leads us to a proposition that we believe provides an accururate but qualified refinement of the prevaling wisdom regaring loss framing.

Proposition 2 (Motivation of dominant sales strategies)**.** *Consider two (unidimensional) strategies, σ_1 and σ_2. Assume that $f(s|\sigma_1)$ first-order stochastically dominates $f(s|\sigma_2)$. Inducing loss framing cannot make σ_2 more attractive relative to σ_1, and for some potential reference values inducing loss framing makes σ_1 more attractive relative to σ_2. For those reference values, σ_1 is preferred to σ_2 if the additional weight placed on losses Λ is sufficiently large.*

Proof. The assumption that $f(s|\sigma_1)$ first-order stochastically dominates $f(s|\sigma_2)$ implies that $L(\sigma_1) \leq L(\sigma_2)$, with the inequality strict for some potential reference values. The claims therefore follow immediately from Proposition 1. □

Proposition 2 implies that loss framing is predicted to have an unambiguous (weakly) positive effect on productivity when the available strategies are ordered by first-order stochastic dominance. This aligns exactly with the intuitions just described: when strategies are so ordered, there exists no tradeoff between loss exposure and expected sales.

Proposition 2 additionally helps to reconcile our results with existing theoretical claims. For example, Prediction 1 of Imas, Sadoff and Samek (2016) and Prediction 1 of de Quidt (2017) both establish unambiguously positive effects of loss framing. The models in these papers involve effort choices where greater effort would indeed induce a first-order stochastically dominant outcome distribution. And indeed, we believe that this is a reasonable modeling assumption in these papers' experimental contexts, which present straightforward "real effort tasks" that are intentionally designed to be clean measures of effort provision. To the extent that most of the lab-experimental literature uses tasks like these, the conventional wisdom may perhaps be generally expected to hold in these papers. A notable exception occurs in the work of Ahrens, Bitter and Bosch-Rosa (2023), who document negative effects of loss framing among players of a coordination game facing strategic uncertainty. This experimental environment involves the type of loss exposure/productivity tradeoffs that can generate negative effects in our unidimensional case, and indeed the authors present their findings as a validation of our theory.

Beyond demonstrating the ways in which this past literature is typically correct, however, this analysis also demonstrates ways in which past literature has been limited. By focusing on environments where the only way to avoid losses is to increase effort, the lab-experimental literature has directed attention to cases where the channels for negative consequences do not exist. In the field settings where loss-framed interventions have been attempted—for example, among factory workers, salespeople, or teachers—productivity is clearly a substantially more complex object with more scope for meaningful tradeoffs. This creates the potential for negative productivity consequences like those we have documented can rise to first-order importance. This perhaps helps explain why these field studies have not systematically seen the large positive effects that have been observed, on average, in the lab literature (as documented in the metanalysis of Ferraro and Tracy, 2022).

3 Discussion

In recent years, the findings of behavioral economics have been broadly disseminated to the general public. This surge of public engagement has contributed to a wave of interest in designing policy interventions that harness behavioral-economic forces. While the potential for using loss framing to sharpen performance incentives has received a great deal of attention, formal field evaluation of such programs is currently limited to a relatively small number of cases (see Ferraro and Tracy, 2022, for a critical review). Furthermore, formal theoretical examination of the effect of these incentives has been limited to their application to somewhat narrow and stylized circumstances.

In this paper, we have sought to critically assess the desirability of loss framing for organizations and policy makers. In an unusually large field experiment, in which $66 million in bonus payments was subject to random assignment of loss framing, we have found strong indications that loss aversion influenced sales activity in a manner contrary to our typical expectations. When distilling our findings, some caution is merited due to the marginal significance of some of our main results. However, the overall patterns in our data, and especially those revealed in our DKDD analysis, provide clear reasons to believe that loss framing was ultimately harmful to sales efforts. Furthermore, despite common claims to the contrary, we have shown that such negative results are theoretically expected.

Our findings clearly illustrate a position summarized in Card, DellaVigna and Malmendier (2011): that field experiments can often benefit from the development of guiding theory, even ex-post.[25] The initial experimental findings of Hossain and List (2012) motivated substantial interest in loss-framing interventions. However, in papers that followed, some researchers faced difficulties when attempting to generate similar positive treatment effects.[26] Through a formalized reexamination of the underlying conceptual premise of this literature, we are able to suggest factors that could contribute to the mixed efficacy of these interventions, and are able to make more precise predictions about the situations in which loss framing may be counterproductive.

We note that although automotive retailing in the United States is an idiosyncratic (yet economically important) setting, the characteristics that generate our results are in no way unique to that setting. One survey reported that 72% of firms including bonus pay in their sales compensation systems. Of those firms, 76% use performance quotas (Joseph and Kalwani, 1998). Discrete bonus systems are used across numerous private and public sectors (e.g., Tzioumis and Gee, 2013; Fryer, 2013; Asch, 1990), most of which face the potential for gaming and multitasking problems. Our theoretical model explains why loss framing poses the same hazards in these settings as it does in auto dealer incentives.

[25] This point is particularly emphasized in fields where generalizability is a common criticism of experimentation (Di Stefano and Gutierrez, 2019).

[26] And indeed, the original work of Hossain and List (2012) did not find unambiguous positive effects in all situations studied.

Although the presence of loss aversion is undesirable in our setting, we note that this does not undercut the demonstration that loss aversion is *present*. Until relatively recently, demonstrations of loss aversion were largely restricted to lab settings, allowing reasonable researchers to question the field-relevance of the phenomenon. More recently, perhaps due to the increase in data availability, we have seen a proliferation of demonstrations of loss aversion in fundamentally economic field environments. Loss aversion has now been shown to influence job search behavior (DellaVigna et al., 2017), labor supply (see Camerer et al., 1997, and the many papers it inspired), house prices (Genesove and Mayer, 2001; Andersen et al., 2022), tax compliance (Engström et al., 2015; Rees-Jones, 2018), and more (for a thorough review, see O'Donoghue and Sprenger, 2018). We add to this growing list of field applications, and provide a demonstration with an unusually large estimated financial impact.

Because the dealers that we study are self-contained businesses themselves, an important novelty of our results is our demonstration that behavioral models apply not only to individual workers, but also to firms. While it is often argued that the experience, stakes, and competitive forces present among market actors should eliminate the role of behavioral-economic considerations (List, 2002, 2003, 2004*a*,*b*), our results suggest that this is not so in one large and important market. This applicability may be partially explained by the largely private ownership structure of car dealerships that is particularly dominant in our focal manufacturer's dealer network (98%). Because many dealerships have small management teams, they may operate more like individuals than the abstract ideal of a firm. Furthermore, franchising laws heavily limit entry and consolidation in this market, limiting some of the competitive pressure that is often assumed to discipline behavioral tendencies. Despite these caveats, we note that the ownership characteristics and competitive frictions in our setting are common in much of the world (Bloom and Van Reenen, 2007; Bloom, Sadun and Van Reenen, 2012), and that some frictions to idealized competition exist in many other markets. We therefore view our paper as strongly suggesting that behavioral-economic considerations apply to the analysis of firms in quantitatively important ways,[27] and thus that

[27] For additional papers emphasizing the role of loss aversion in corporate settings, see Loughran and McDonald (2013); Ljungqvist and Wilhelm Jr (2005); Baker, Pan and Wurgler (2012); Dittmann, Maug and Spalt (2010)

policy interventions guided by behavioral economics might be productively deployed to such populations.

References

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. "Synthetic control methods for comparative case studies: Estimating the effect of Californias tobacco control program." *Journal of the American Statistical Association*, 105(490): 493–505.

Ahrens, Steffen, Lea Bitter, and Ciril Bosch-Rosa. 2023. "Coordination under loss contracts." *Games and Economic Behavior*, 137: 270–293.

Andersen, Steffen, Cristian Badarinza, Lu Liu, Julie Marx, and Tarun Ramadorai. 2022. "Reference dependence in the housing market." *American Economic Review*, 112(10): 3398–3440.

Asch, Beth J. 1990. "Do incentives matter? The case of navy recruiters." *ILR Review*, 43(3): 89–S.

Athey, Susan, and Guido W Imbens. 2017. "The econometrics of randomized experiments." In *Handbook of Economic Field Experiments*. Vol. 1, 73–140. Elsevier.

Baker, George P, and Thomas N Hubbard. 2003. "Make versus buy in trucking: Asset ownership, job design, and information." *American Economic Review*, 93(3): 551–572.

Baker, Malcolm, Xin Pan, and Jeffrey Wurgler. 2012. "The effect of reference point prices on mergers and acquisitions." *Journal of Financial Economics*, 106(1): 49–71.

Balmaceda, Felipe. 2018. "Optimal task assignments with loss-averse agents." *European Economic Review*, 105: 1 – 26.

Barnatchez, Keith, Leland D. Crane, and Ryan A. Decker. 2017. "An assessment of the National Establishment Time Series (NETS) database." *Opportunity Inclusive Growth Institute: System Working Paper*, 17–29.

Benartzi, Shlomo, John Beshears, Katherine L. Milkman, Cass R. Sunstein, Richard H. Thaler, Maya Shankar, Will Tucker-Ray, William J. Congdon, and Steven Galing. 2017. "Should governments invest more in nudging?" *Psychological Science*, 28(8): 1041–1055. PMID: 28581899.

Bennett, Victor Manuel. 2013. "Organization and bargaining: Sales process choice at auto dealerships." *Management Science*, 59(9): 2003–2018.

Benson, Alan. 2015. "Do agents game their agents behavior? Evidence from sales managers." *Journal of Labor Economics*, 33(4): 863–890.

Bloom, Nicholas, and John Van Reenen. 2007. "Measuring and explaining management practices across firms and countries." *The Quarterly Journal of Economics*, 122(4): 1351–1408.

Bloom, Nicholas, Raffaella Sadun, and John Van Reenen. 2012. "The organization of firms across countries." *The Quarterly Journal of Economics*, 127(4): 1663–1705.

Brooks, Richard R. W., Alexander Stremitzer, and Stephan Tontrup. 2012. "Framing contracts: Why loss framing increases effort." *Journal of Institutional and Theoretical Economics*, 168(1): 62–82.

Brownback, Andy, and Sally Sadoff. 2020. "Improving college instruction through incentives." *Journal of Political Economy*, 128(8): 2925–2972.

Bulte, Erwin, John A List, and Daan van Soest. 2020. "Toward an understanding of the welfare effects of nudges: Evidence from a field experiment in the workplace." *The Economic Journal*, 130(632): 2329–2353.

Bulte, Erwin, John A List, and Daan van Soest. 2021. "Incentive spillovers in the workplace: Evidence from two field experiments." *Journal of Economic Behavior & Organization*, 184: 137–149.

Busse, Meghan, Jorge Silva-Risso, and Florian Zettelmeyer. 2006. "$1,000 cash back: The pass-through of auto manufacturer promotions." *American Economic Review*, 96(4): 1253–1270.

Busse, Meghan R, and Jorge M Silva-Risso. 2010. "“One discriminatory rent” or “double jeopardy”: Multicomponent negotiation for new car purchases." *American Economic Review*, 100(2): 470–74.

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

Card, David, Stefano DellaVigna, and Ulrike Malmendier. 2011. "The role of theory in field experiments." *Journal of Economic Perspectives*, 25(3): 39–62.

Chen, Yvonne Jie, Pei Li, and Yi Lu. 2018. "Career concerns and multitasking local bureaucrats: Evidence of a target-based performance evaluation system in China." *Journal of Development Economics*, 133: 84–101.

Chung, Doug J, and Das Narayandas. 2017. "Incentives versus reciprocity: insights from a field experiment." *Journal of Marketing Research*, 54(4): 511–524.

Chung, Doug J, Das Narayandas, and Dongkyu Chang. 2019. "The effects of quota frequency: Sales performance and product focus." *Harvard Business School Working Paper No. 17-059*.

Chung, Doug J, Thomas Steenburgh, and K Sudhir. 2013. "Do bonuses enhance sales productivity? A dynamic structural analysis of bonus-based compensation plans." *Marketing Science*, 33(2): 165–187.

Courty, Pascal, and Gerald Marschke. 2004. "An empirical investigation of gaming responses to explicit performance incentives." *Journal of Labor Economics*, 22(1): 23–56.

DellaVigna, Stefano, and Devin Pope. 2018. "What motivates effort? Evidence and expert forecasts." *The Review of Economic Studies*, 85(2): 1029–1069.

DellaVigna, Stefano, Attila Lindner, Balzs Reizer, and Johannes F. Schmieder. 2017. "Reference-dependent job search: Evidence from Hungary." *The Quarterly Journal of Economics*, 132(4): 1969–2018.

de Quidt, Jonathan. 2017. "Your loss is my gain: A recruitment experiment with framed incentives." *Journal of the European Economic Association*, 16(2): 522–559.

de Quidt, Jonathan, Francesco Fallucchi, Felix Kölle, Daniele Nosenzo, and Simone Quercia. 2017. "Bonus versus penalty: How robust are the effects of contract framing?" *Journal of the Economic Science Association*, 3(2): 174–182.

Di Stefano, Giada, and Cédric Gutierrez. 2019. "Under a magnifying glass: On the use of experiments in strategy research." *Strategic Organization*, 17(4): 497–507.

Dittmann, Ingolf, Ernst Maug, and Oliver Spalt. 2010. "Sticks or carrots? Optimal CEO compensation when managers are loss averse." *The Journal of Finance*, 65(6): 2015–2050.

Dumont, Etienne, Bernard Fortin, Nicolas Jacquemet, and Bruce Shearer. 2008. "Physicians multitasking and incentives: Empirical evidence from a natural experiment." *Journal of health economics*, 27(6): 1436–1450.

Ederer, Florian, Richard Holden, and Margaret Meyer. 2018. "Gaming and strategic opacity in incentive provision." *The RAND Journal of Economics*, 49(4): 819–854.

Engström, Per, Katarina Nordblom, Henry Ohlsson, and Annika Persson. 2015. "Tax Compliance and Loss Aversion." *American Economic Journal: Economic Policy*, 7(4): 132–164.

Ferraro, Paul J, and J Dustin Tracy. 2022. "A reassessment of the potential for loss-framed incentive contracts to increase productivity: a meta-analysis and a real-effort experiment." *Experimental Economics*, 25(5): 1441–1466.

Fryer, Jr., Roland G., Steven D. Levitt, John List, and Sally Sadoff. 2022. "Enhancing the efficacy of teacher incentives through framing: a field experiment." *American Economic Journal: Economic Policy*, 14(4): 269–99.

Fryer, Roland G. 2013. "Teacher incentives and student achievement: Evidence from New York City public schools." *Journal of Labor Economics*, 31(2): 373–407.

Genesove, David, and Christopher Mayer. 2001. "Loss aversion and seller behavior: Evidence from the housing market." *The Quarterly Journal of Economics*, 116(4): 1233–1260.

Gill, David, and Rebecca Stone. 2010. "Fairness and desert in tournaments." *Games and Economic Behavior*, 69(2): 346–364.

Goldberger, A.S. 1968. *Topics in Regression Analysis.* Macmillan.

Hallsworth, Michael, John A. List, Robert D. Metcalfe, and Ivo Vlaev. 2017. "The behavioralist as tax collector: Using natural field experiments to enhance tax compliance." *Journal of Public Economics*, 148: 14–31.

Herweg, Fabian, Daniel Müller, and Philipp Weinschenk. 2010. "Binary payment schemes: moral hazard and loss aversion." *American Economic Review*, 100(5): 2451–77.

Holmstrom, Bengt, and Paul Milgrom. 1991. "Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design." *Journal of Law, Economics, Organization*, 7: 24–52.

Homonoff, Tatiana A. 2018. "Can small incentives have large effects? The impact of taxes versus bonuses on disposable bag use." *American Economic Journal: Economic Policy*, 10(4): 177–210.

Hong, Fuhai, Tanjim Hossain, and John A. List. 2015. "Framing manipulations in contests: A natural field experiment." *Journal of Economic Behavior Organization*, 118: 372 – 382.

Hossain, Tanjim, and John A List. 2012. "The behavioralist visits the factory: Increasing productivity using simple framing manipulations." *Management Science*, 58(12): 2151–2167.

Imas, Alex, Sally Sadoff, and Anya Samek. 2016. "Do people anticipate loss aversion?" *Management Science*, 63(5): 1271–1284.

Jansen, Mark, Lamar Pierce, Jason Snyder, and Hieu Nguyen. 2023. "Product sales incentive spillovers to the lending market: Evidence from subprime auto loan defaults." *Management Science*, Forthcoming.

Joseph, Kissan, and Manohar U. Kalwani. 1998. "The role of bonus pay in salesforce compensation plans." *Industrial Marketing Management*, 27(2): 147–159.

Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler. 1991. "Anomalies: The endowment effect, loss aversion, and status quo bias." *Journal of Economic Perspectives*, 5(1): 193–206.

Kim, Minkyung, K Sudhir, and Kosuke Uetake. 2021. "A structural model of a multitasking salesforce: Incentives, private information and job design." *Management Science*, Forthcoming.

Kleven, Henrik Jacobsen. 2016. "Bunching." *Annual Review of Economics*, 8(1): 435–464.

Kőszegi, Botond, and Matthew Rabin. 2006. "A model of reference-dependent preferences." *The Quarterly Journal of Economics*, 121(4): 1133–1165.

Kuhn, Peter, and Lizi Yu. 2021. "Kinks as goals: Accelerating commissions and the performance of sales teams." Working paper.

Lafontaine, Francine, and Fiona Scott Morton. 2010. "Markets: State franchise laws, dealer terminations, and the auto crisis." *Journal of Economic Perspectives*, 24(3): 233–50.

Larkin, Ian. 2014. "The cost of high-powered incentives: Employee gaming in enterprise software sales." *Journal of Labor Economics*, 32(2): 199–227.

List, John A. 2002. "Testing neoclassical competitive market theory in the field." *Proceedings of the National Academy of Sciences*, 99(24): 15827–15830.

List, John A. 2003. "Does market experience eliminate market anomalies?" *The Quarterly Journal of Economics*, 118(1): 41–71.

List, John A. 2004*a*. "Neoclassical theory versus prospect theory: Evidence from the marketplace." *Econometrica*, 72(2): 615–625.

List, John A. 2004*b*. "Testing neoclassical competitive theory in multilateral decentralized markets." *Journal of Political Economy*, 112(5): 1131–1156.

Ljungqvist, Alexander, and William J Wilhelm Jr. 2005. "Does prospect theory explain IPO market behavior?" *The Journal of Finance*, 60(4): 1759–1790.

Loughran, Tim, and Bill McDonald. 2013. "IPO first-day returns, offer price revisions, volatility, and form S-1 language." *Journal of Financial Economics*, 109(2): 307–326.

Lu, Bo, Robert Greevy, Xinyi Xu, and Cole Beck. 2011. "Optimal nonbipartite matching and its statistical applications." *The American Statistician*, 65(1): 21–30.

Manski, Charles F. 1986. "Analog estimation of econometric models." In *Handbook of Econometrics*. Vol. 4, , ed. R. F. Engle and D. McFadden, Chapter 43, 2559–2582. Elsevier.

Maritz, Inc. 2018. "Proprietary Sales Incentive Data From Confidential Automotive Manufacturer *CarCo*." Washington University in St. Louis.

Misra, Sanjog, and Harikesh S Nair. 2011. "A structural model of sales-force compensation dynamics: Estimation and field implementation." *Quantitative Marketing and Economics*, 9(3): 211–257.

Murry, Charles, and Henry S Schneider. 2016. "The economics of retail markets for new and used cars." *Handbook on the Economics of Retailing and Distribution*, 343.

O'Donoghue, Ted, and Charles Sprenger. 2018. "Reference-dependent preferences." In *Handbook of Behavioral Economics - Foundations and Applications 1*. Vol. 1, , ed. B. Douglas Bernheim, Stefano DellaVigna and David Laibson, 1 – 77. North-Holland.

Oyer, Paul. 1998. "Fiscal year ends and nonlinear incentive contracts: The effect on business seasonality." *The Quarterly Journal of Economics*, 113(1): 149–185.

Pierce, Lamar. 2012. "Organizational structure and the limits of knowledge sharing: Incentive conflict and agency in car leasing." *Management Science*, 58(6): 1106–1121.

Rees-Jones, Alex. 2018. "Quantifying loss-averse tax manipulation." *The Review of Economic Studies*, 85(2): 1251–1278.

Ritz, Robert A. 2008. "Strategic incentives for market share." *International Journal of Industrial Organization*, 26(2): 586–597.

Robbins, Michael W, Jessica Saunders, and Beau Kilmer. 2017. "A framework for synthetic control methods with high-dimensional, micro-level data: evaluating a neighborhood-specific crime intervention." *Journal of the American Statistical Association*, 112(517): 109–126.

Roberts, Adrienne. 2018. "Car dealerships face conundrum: Get big or get out." *Wall Street Journal*.

Steenburgh, Thomas J. 2008. "Effort or timing: The effect of lump-sum bonuses." *Quantitative Marketing and Economics*, 6(3): 235.

Tzioumis, Konstantinos, and Matthew Gee. 2013. "Nonlinear incentives and mortgage officers' decisions." *Journal of Financial Economics*, 107(2): 436–453.

Walls & Associates, Inc. 2015. "National Establishment Time Series (NETS) Dataset."

A Appendix

A.1 Density of Dealers in DMAs

Figure A.1 shows the distribution of the number of participants in the DMAs assigned to each group. The outlier is the same control group DMA referenced in the discussion of Figure 4, containing the unusually large dealers.

Figure A.1: Distribution of DMA Dealer Count By Condition

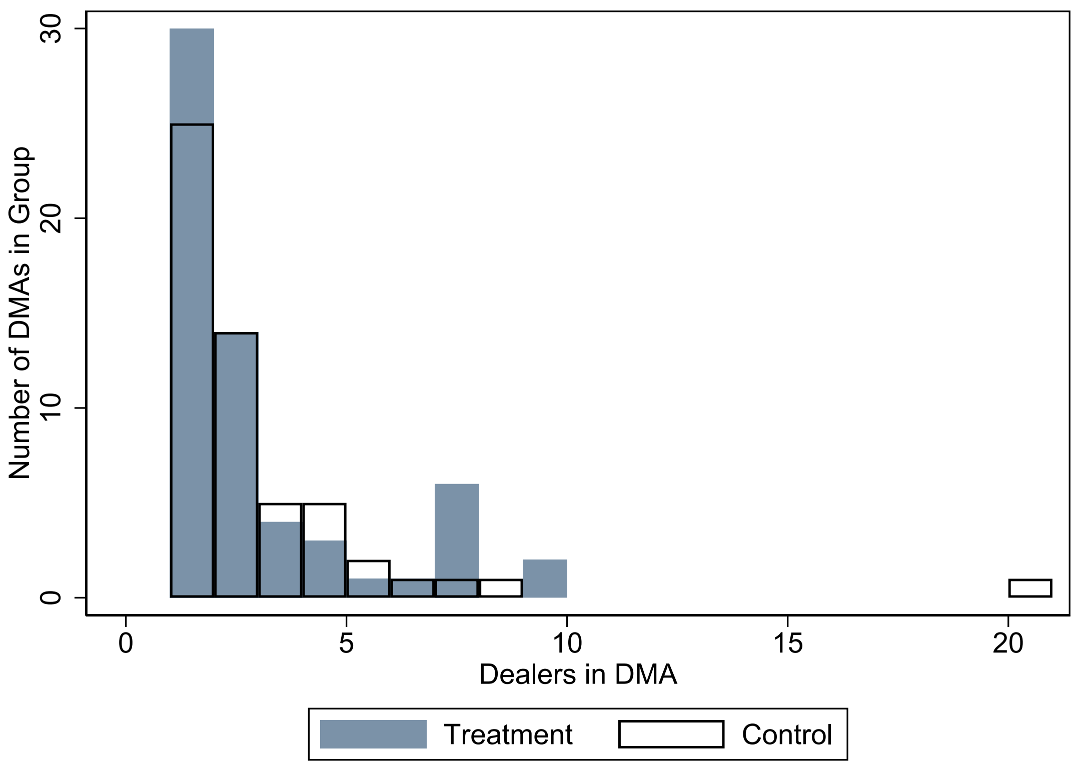
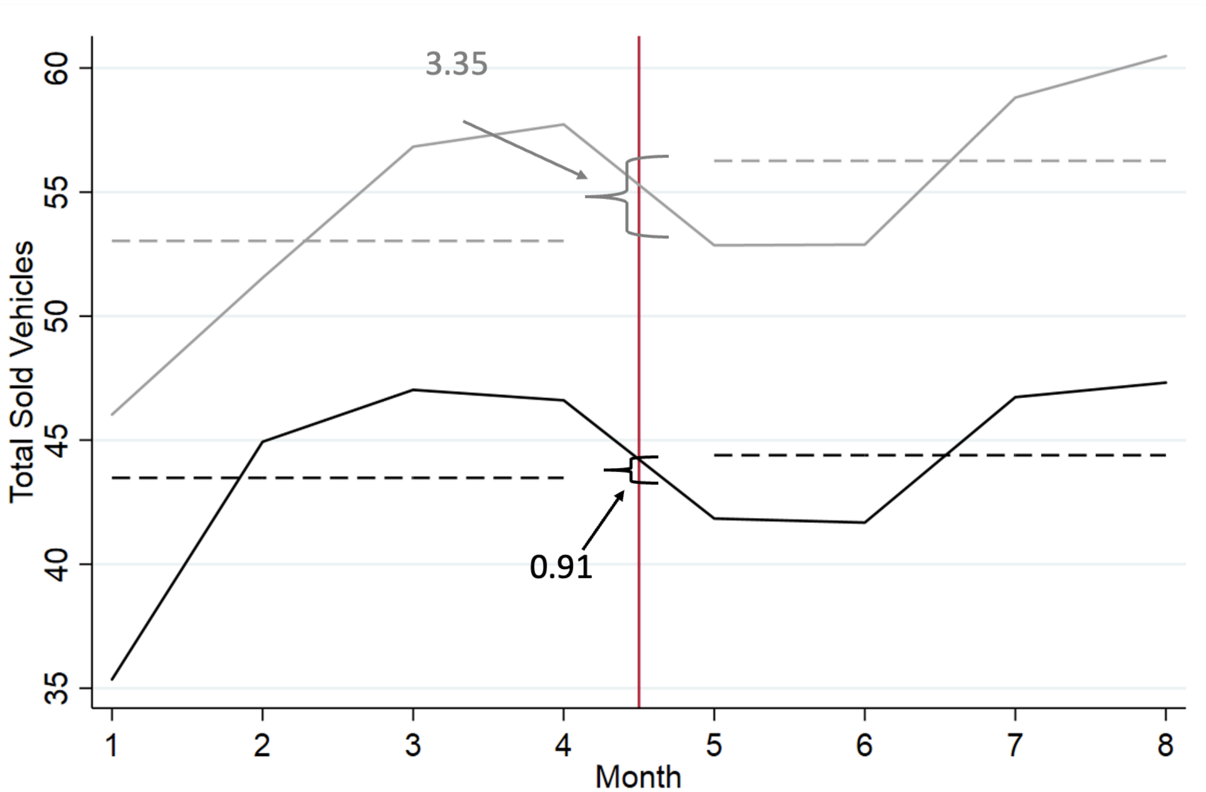


Figure A.2: Monthly Trends in Sold Vehicles for Participating Dealers



Notes: This figure shows average total dealer sales by month for dealers assigned to treatment (in black) and control (in grey). The first four months are the pre-treatment period and months 5-8 are those where the treatment group is assigned treatment. Horizontal dashed lines represent the average monthly sales across each four month period. Note that the period specific averages do not equal the sum of model-specific averages in Table 1 because some dealers only carry one model-group.

A.2 Parallel Trends Assumption

To help assess the parallel-trend assumptions that are central to difference in differences designs, Figure A.2 shows time-paths of sales over the first eight months of 2017. This window includes the four-month pre-treatment window and the four months during which the treatment group was prepaid. Several patterns are evident from this figure. First, the observable pretrends in both groups are encouragingly similar. Second, the estimated treatment effect is evident in the larger gap between groups in months 5-8. We believe this figure provides reassurance that our attempt to match pre-trends with our randomization procedure was successful, and some assurance that the treatment effects we study are clearly apparent in minimally processed data.

A.3 Assessment of Selection

In this paper, we have focused on the negative effects of loss framing *conditional on participation in our experiment*. While prior research has assumed that such effects should be positive, several papers have emphasized the potential for negative effects arising by selecting who participates (Imas, Sadoff and Samek, 2016; de Quidt, 2017). To the extent that treatment assignment was random conditional on participation, these concerns should not confound our estimated treatment effects, although they do influence our interpretation of who is averaged in average treatment effects.

For comparability to prior research and to completely examine existing accounts of the negative effects of loss framing, we now examine the predictors of selection into our experiment.

Despite the manufacturer's initial belief that this program would be highly desirable to dealers (due to providing early cash flow), comparatively few dealers opted into our experiment: 294 dealers opted in, 336 actively opted out, and 597 opted out through non-response. This low rate of opt-in could potentially be interpreted as prima facie evidence that dealers did in fact anticipate loss aversion, and thus avoided a situation that might induce it. While we do believe that this is a partial explanation, we note that several alternative explanations of the low participation rate are present. First, dealers may have failed to participate purely because they did not know of this opportunity. However, given the manner in which the opportunity was advertised, we believe this was unlikely to be a major factor.[28] Second, dealers may have feared that they would be unable to make the necessary clawback payments if targets were not met. In practice, several fail-safes ensured that dealers would face relatively few severe negative consequences if this situation arose, but lack of knowledge of those fail-safes or remaining concerns could possibly drive behavior.[29] Third, dealers may have been averse to accepting the administrative costs that a change in

[28]Multiple emails were sent to the incentive program's contact at each dealer. CarCo regional dealer representatives indicated that dealers almost certainly read at least one of the multiple invitation emails. Since the incentive program is crucial for their sales profits, program communications are high priority. In addition, CarCo's regional dealer representatives followed up with non-respondents to encourage participation. Based on these considerations, we believe effectively all non-respondents knowingly defaulted to non-participation.

[29]Based on the researcher's concerns about the negative consequences of failure to repay, CarCo agreed to automatically unenroll any dealer failing repayment. Among participating dealers, this condition was never triggered.

accounting practices would require.[30] In sum, while the low rate of participation is consistent with the concern that loss-averse agents will select out of loss framed contracts, we cannot firmly establish that this is this is the mechanism driving nonparticipation.

To examine how selection into the experiment influences the composition of our sample, we began by testing for differences between the three groups of dealers (opt-in, opt-out, and non-respondent) across the variables available in CarCo's internal data.

To begin, we emphasize that we find no statistically distinguishable differences in treatment group assignment across the three participation groups ($\chi^2 = 3.00, p = 0.223$). Treatment was assigned to 48.3% of participants, 54.0% of opt-outs, and 48.6% of non-respondents.

Among CarCo's internal data, we see evidence of selection into the program across several observable dimensions. Although participants and opt-out dealers had statistically indistinguishable monthly sales volume (48.5 vs. 43.2, $p = 0.162$), non-respondents had substantially fewer sales than both (31.1, $p < 0.001$).[31] Second, and consistent with these volume differences, there are significant cross-group differences in whether or not the dealers carry both model groups ($\chi^2 = 10.93, p = 0.004$). Non-respondents (86%) are less likely to do so than both participants (90%) and opt-out dealers (93%). Third, we find substantial differences across regions in participation rates, and particularly in the rate of opt-outs through non-responses.

These differences reject the hypothesis that dealers' willingness to be exposed to loss-framing through participation in our experiment is completely random conditional on observables. While the random assignment of treatment preserves the validity of estimated treatment effects within this group, we sought to collect more dealer-level data to help us better understand the features of the dealers who selected into our study.

We generate our dealer-level data by combining CarCo's sales and incentive records with supplemental data from the National Establishment Time Series (NETS). NETS aggregates data on most establishments in the United States (Barnatchez, Crane and Decker, 2017), including car dealers, providing data on ownership structure, employment, and financial

[30] Regional dealer representatives informally indicated that many of the dealers claimed that their lack of participation was due to avoidance of what they viewed as an accounting "headache," combined with little need for the early cash flow.

[31] These averages do not equal the sum of model-group average sales in Table 1 because some dealers carry only one model group.

Table A.1: Group Means by Participation Status for Dealers with NETS Data

	Selection Group			Group Difference Test
	Participants	Opt-Out	Non-Respondents	
Year Founded	1976.49	1976.67	1974.25	$F = 0.98$
	(27.16)	(27.76)	(27.39)	[0.38]
Employees	55.13	50.86	51.98	$F = 0.73$
	(46.70)	(45.30)	(40.27)	[0.48]
2015 DUNS Score	2.47	2.40	2.41	$F = 0.85$
	(0.60)	(0.52)	(0.56)	[0.43]
Has DUNS Score	0.70	0.62	0.66	$\chi^2 = 3.50$
	(0.46)	(0.49)	(0.47)	[0.17]
2015 Min Paydex	73.16	74.22	73.17	$F = 1.31$
	(8.73)	(6.34)	(9.59)	[0.27]
Has Paydex Score	0.81	0.75	0.80	$\chi^2 = 3.37$
	(0.39)	(0.43)	(0.40)	[0.19]
Publicly-Held	0.00	0.02	0.02	$\chi^2 = 3.78$
	(0.06)	(0.14)	(0.15)	[0.15]
Part of Group	0.18	0.23	0.26	$\chi^2 = 5.79$
	(0.39)	(0.42)	(0.44)	[0.06]

Notes: Summary statistics are presented for three groups: those who chose participation, those who opted-out, and those who did not respond and therefore were non-participants by default. Means with standard deviations in parentheses for each group. F-statistics from ANOVA are presented for continuous variables. Chi-squared tests are for dichotomous variables. P-values are in brackets. Each line only includes those dealers with a populated NETS field.

strength data from Dun & Bradstreet.[32] We successfully matched 877 non-participants and 276 participants via phone, address, name, and ownership data. We then examined differences in sales, ownership, age, employment, and financial health.[33]

Table A.1 presents group means and statistical tests of differences across the three groups for the 1,153 dealers with NETS data. We see few observable differences between the three groups. In particular, no statistical differences are detected in the age or size of the firm, in the Dun & Bradstreet measures of financial health, in the Paydex measures of reliable bill payment, or in publicly-held status. One potential difference is seen in the "Part of Group" variable. Participants are more likely to be stand-alone dealers (18%) than opt-out (23%) or non-respondents (26%). These differences could arise if dealers that were part of larger organizations were less likely to have autonomy to change accounting systems or rules.

In summary, we see relatively modest evidence of observable selection into participation in our experiment. We acknowledge that we cannot rule out perhaps interesting unobservable differences in dealers by participation status, but we note these would be very unlikely to bias our treatment estimates given our random assignment procedure.

A.4 Robustness to Outliers

As discussed in Section 1.4, notable differences exist in the average pre-intervention sales between the treatment and control groups. As we argued in our discussion of Figure 4, this difference can be attributed to the random assignment of an outlier DMA. A single DMA contained five of the six largest participating dealers, each with average monthly sales several multiples of the average of other dealers. Due to the natural concern that these influential outliers meaningfully drive results, we exclude this DMA as well as its matched partner, and present summary statistics in Table A.2 and difference in differences models in Table A.3. With its exclusion, average sales across both model groups in the pre-period are statistically indistinguishable ($p = 0.451$), with means of 43.5 and 45.5 for the treatment and control group, respectively. We note that these numbers are not simply the sum of the preperiod

[32] CarCo did not share internal data on dealership financial or ownership structure with us for legal reasons.

[33] The NETS data fields have varying levels of completeness. Employment data, for example, is complete for approximately 90% of the matched dealers. Dun & Bradstreet data, however, is complete for only about two-thirds.

Table A.2: Summary Statistics Excluding Outlier DMA and its Match

	LARGE-BONUS MODEL GROUP					
	Treatment Group			Control Group		
	Pre	Post	Total	Pre	Post	Total
Vehicles	29.59	31.56	30.58	34.42	36.49	35.46
Sold	(21.58)	(23.85)	(22.75)	(45.76)	(46.69)	(46.22)
Target	27.07	32.25	29.66	28.80	35.37	32.09
Sales	(19.48)	(22.17)	(21.02)	(30.73)	(38.66)	(35.07)
Hit 110%	0.56	0.44	0.50	0.63	0.49	0.56
Target	(0.50)	(0.50)	(0.50)	(0.48)	(0.50)	(0.50)
Hit 100%	0.61	0.49	0.55	0.68	0.54	0.61
Target	(0.49)	(0.50)	(0.50)	(0.47)	(0.50)	(0.49)
	SMALL-BONUS MODEL GROUP					
	Treatment Group			Control Group		
	Pre	Post	Total	Pre	Post	Total
Vehicles	15.14	14.16	14.65	14.29	14.71	14.50
Sold	(14.07)	(13.47)	(13.78)	(13.63)	(14.79)	(14.22)
Target	12.39	14.69	13.54	12.27	14.71	13.49
Sales	(9.91)	(11.92)	(11.02)	(10.79)	(12.91)	(11.96)
Hit 110%	0.64	0.44	0.54	0.62	0.46	0.54
Target	(0.48)	(0.50)	(0.50)	(0.49)	(0.50)	(0.50)
Hit 100%	0.70	0.50	0.60	0.69	0.51	0.60
Target	(0.46)	(0.50)	(0.49)	(0.46)	(0.50)	(0.49)

Notes: Summary statistics of monthly sales performance by model group and treatment excluding the largest DMA in the control group. Means and standard deviations presented in all cells. The first two rows present monthly sales and monthly target thresholds. The rows below summarize the probability of hitting earning the larger fixed bonus for exceeding the 110% threshold, the smaller fixed bonus for exceeding the standard target threshold.

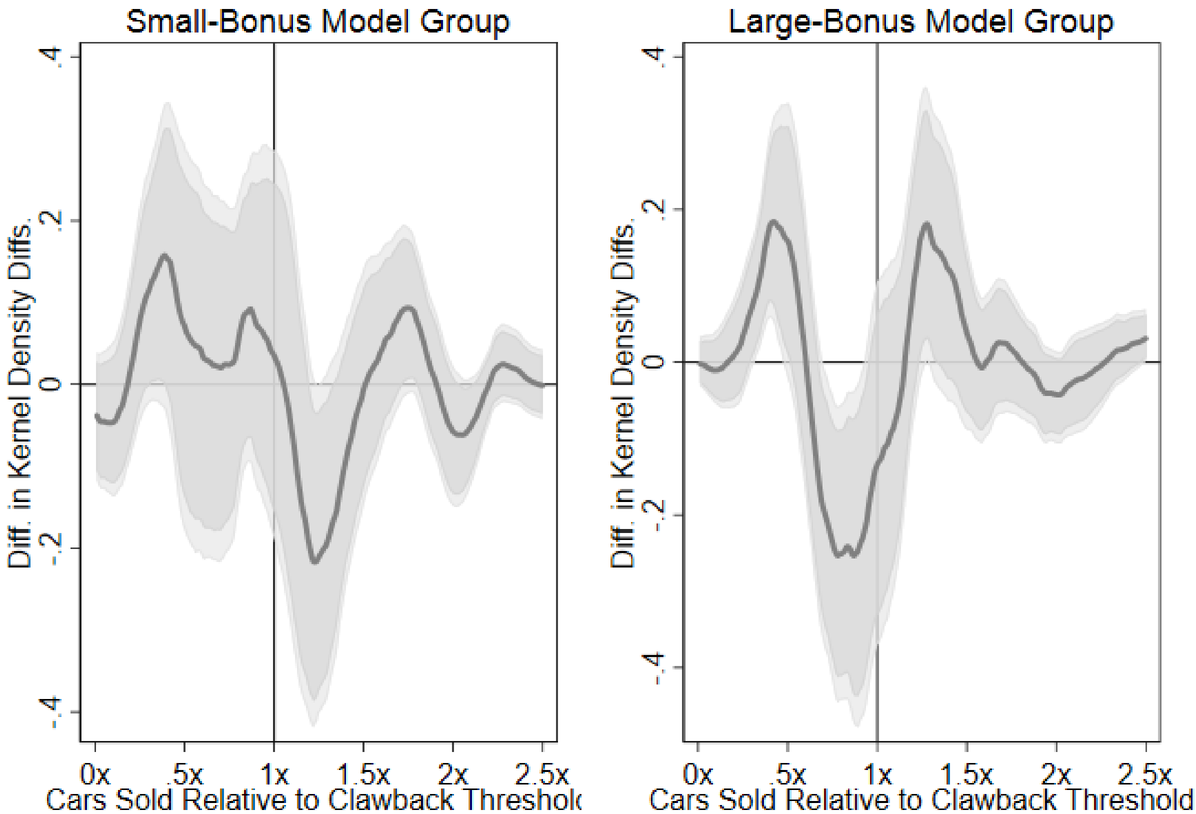
Table A.3: Difference-in-Difference Estimates Excluding the Outlier DMA and its Match

	(1) OLS	(2) PPML	(3) OLS	(4) PPML	(5) OLS	(6) PPML
$\text{Treated}_d \times \text{Treatment period}_m$	-1.57	-0.032	-0.16	0.004	-1.44	-0.098
	(1.14)	(0.023)	(0.88)	(0.023)	(0.73)	(0.045)
Model Group	Pooled	Pooled	L-B	L-B	S-B	S-B
Dealer Fixed Effects	X	X	X	X	X	X
Month Fixed Effects	X	X	X	X	X	X
Observations	2280	2280	2189	2189	2144	2144

Notes: This table presents regressions predicting the dealer/month-specific number of vehicles sold. Odd numbered columns present OLS estimates of equation 1. Even numbered columns present Poisson Pseudo-Maximum Likelihood (PPML) estimates of equation 2. We first present analysis pooling the two model groups together (cols 1 and 2), followed by analysis of large-bonus model group (cols 3 and 4) and small-bonus model group (cols 5 and 6). Standard errors are clustered at the DMA level.

in Table A.2 because not all dealerships carry both model groups. Regression results are presented in Table A.3. Average monthly sales decrease by 1.57 vehicles per month or 3% with nearly all losses coming from the small-bonus model group. With the outlier DMA excluded, the magnitude and significance of our diff-in-diff results are somewhat reduced, but the same general patterns remain. DKDD results, presented in Figure A.3, remain largely unchanged.

Figure A.3: The Impact of Loss Framing on the Distribution of Sales While Excluding the Outlier DMA and its Match



Notes: This figure shows difference-in-kernel-density-differences estimates of the impact of loss framing on the distribution of sales achieved. The x-axes show the amount of sales relative to clawback threshold. The two subfigures present estimates derived from each model group. Confidence regions, based on 10,000 bootstrap iterations resampled by DMA, are shaded. The 90% confidence region is shaded darkly and the 95% confidence region is shaded lightly. Kernel: Epanechnikov; Bandwidth: .1x; Sample Sizes: 2,108 (left panel) and 2,166 (right panel).

A.5 Robustness to the Exclusion of Four Flipped DMAs

As explained in the main text, two matched pairs of DMAs were reassigned by CarCo in order to avoid neighboring DMAs with cross-DMA competition having different treatment assignment. In Table A.4 we repeat our difference-in-difference models with similar but

Table A.4: Difference-in-Difference Estimates Excluding Four Flipped DMAs

	(1) OLS	(2) PPML	(3) OLS	(4) PPML	(5) OLS	(6) PPML
$\text{Treated}_d \times \text{Treatment period}_m$	-1.47	-0.030	-0.08	0.007	-1.41	-0.096
	(1.14)	(0.023)	(0.89)	(0.024)	(0.74)	(0.046)
Model Group	Pooled	Pooled	L-B	L-B	S-B	S-B
Dealer Fixed Effects	X	X	X	X	X	X
Month Fixed Effects	X	X	X	X	X	X
Observations	2256	2256	2165	2165	2120	2120

Notes: This table presents regressions predicting the dealer/month-specific number of vehicles sold. Odd numbered columns present OLS estimates of equation 1. Even numbered columns present Poisson Pseudo-Maximum Likelihood (PPML) estimates of equation 2. We first present analysis pooling the two model groups together (cols 1 and 2), followed by analysis of large-bonus model group (cols 3 and 4) and small-bonus model group (cols 5 and 6). Standard errors are clustered at the DMA level.

smaller effect sizes. DKDD results, presented in Figure A.4, remain largely unchanged.

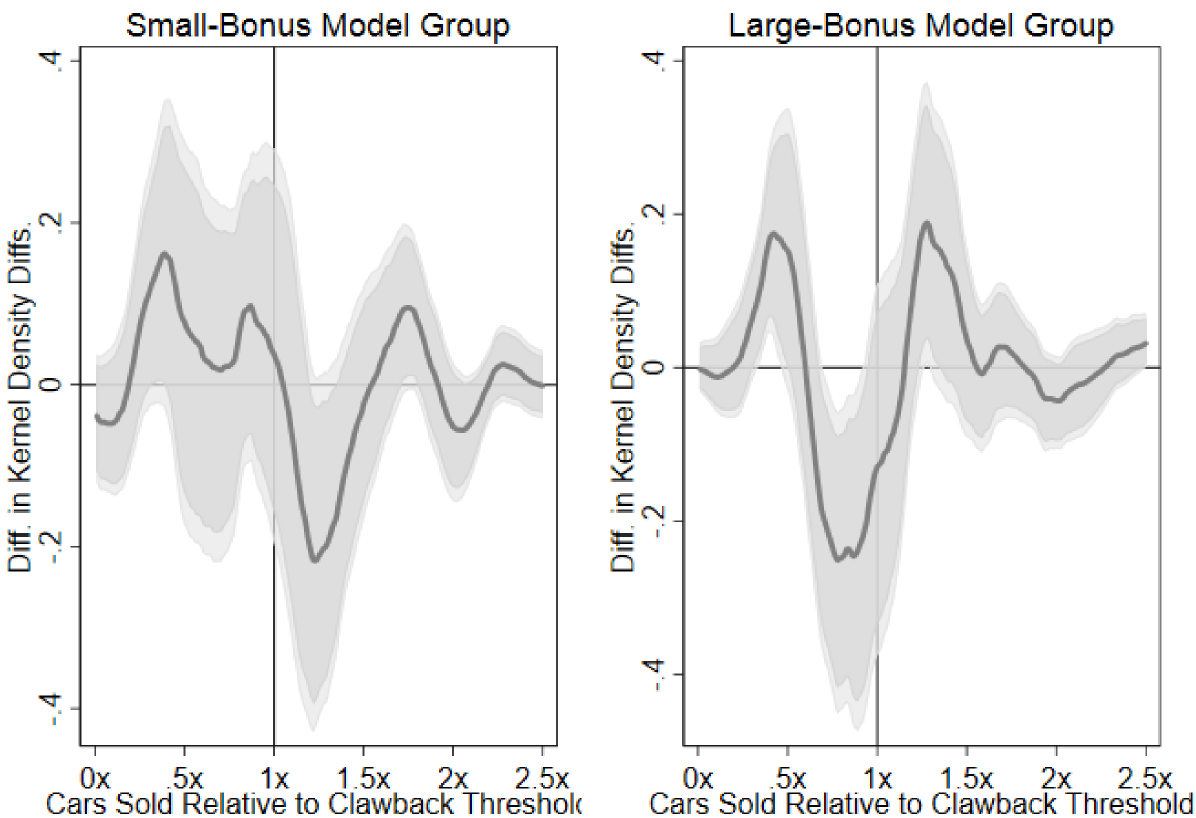
A.6 Treatment Effect in Second Treatment Window

As described in Section 1.2, an "equal treatment" requirement led us to design our intervention so that all experimental participants faced four months in treatment condition and four months in the control condition. We have focused our attention on the estimated impact of treatment in the first four-month treatment window, prior to conditions being flipped. Conceptually, we believe that analysis of treatment effects in the first window have the cleanest interpretation. When examining the second window, the prior treatment of the control group is expected to affect results, leading to a confounding of a difference-in-differences design. For completeness, we attempt to estimate treatment effects in this window using synthetic control methods (Abadie, Diamond and Hainmueller, 2010; Robbins, Saunders and Kilmer, 2017).

The aim of our synthetic control analysis is to compare the sales among dealers receiving treatment in the second window to that of comparable dealers who did not participate in the experiment.[34] We constructed the synthetic control group by matching on dealer

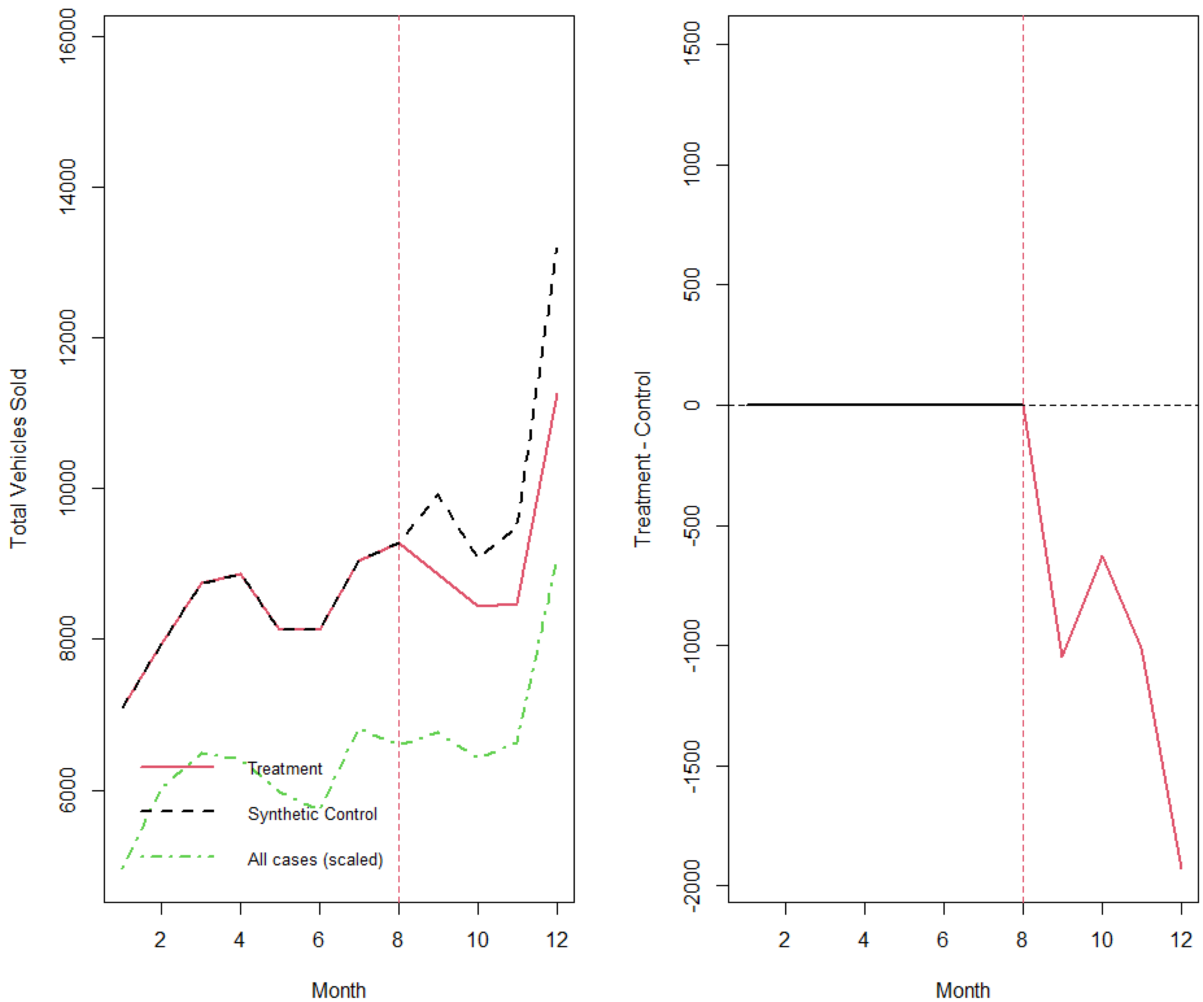
[34] Because random assignment occurred before participation decisions, we specifically restrict attention to non-participants who were assigned to control.

Figure A.4: The Impact of Loss Framing on the Distribution of Sales While Excluding the Four Flipped DMAs



Notes: This figure shows difference-in-kernel-density-differences estimates of the impact of loss framing on the distribution of sales achieved. The x-axes show the amount of sales relative to clawback threshold. The two subfigures present estimates derived from each model group. Confidence regions, based on 10,000 bootstrap iterations resampled by DMA, are shaded. The 90% confidence region is shaded darkly and the 95% confidence region is shaded lightly. Kernel: Epanechnikov; Bandwidth: .1x; Sample Sizes: 2,085 (left panel) and 2,142 (right panel).

Figure A.5: Synthetic Control Estimates of Treatment Effect in Second Treatment Window



Notes: This figure shows results from a synthetic control model, estimating the impact of the control group receiving the loss-framing treatment in months 9–12. The vertical axes represent the total monthly sales across the treatment and synthetic control groups. The synthetic control group is built from non-participants in the treatment group and matches on dealer size (in units sold), dual-dealer status, and pre-treatment sales trends. The model estimates a treatment of -11%.

size (in units sold), dual-dealer status, and the pre-trends in sales that were used in our algorithm for random assignment. The estimated model, represented in Figure A.5, shows a treatment effect of loss framing of -11% (p = 0.015)[35]—substantially larger than our

[35] All estimates are derived using the R package microsynth, with p-values generated through the permutation method with 10,000 permutations applied.

estimate of -5% in our primary analysis. Though we believe our difference-in-differences estimates are conceptually preferable to these synthetic control estimates (as they do not require comparisons to a group that has endogenously opted out of treatment), we note that both methods provide evidence of a negative effect of loss framing.

A.7 Interview and Survey Evidence

In this paper, we have focused on the observed reaction to the treatment using objective sales data. As a complement to this evidence, we conducted informal interviews and formal surveying of participants after the experiment. Fourteen dealers were interviewed, and sixty-nine dealers (23%) responded to the online survey. Overall, the interviews and surveys suggested reasonably universal perceived importance of the incentive program, but more heterogeneous reactions to the loss framing treatment. Some dealers indicated perceptions that were clearly in line with loss-averse evaluation, while others provided assessments of the program more in line with neoclassical considerations. The interviews also indicated both awareness of, and efforts directed towards, the need to optimize efforts across the two model groups, consistent with the exacerbation of multitasking issues that we have discussed.

A.7.1 Interview Evidence

The interviews were conducted over the phone in May, 2018, with either the general manager or principal (owner) of 14 participating dealers. Dealer interviews were arranged by CarCo with a mix of targeted regions and experimental conditions. Participation was not random nor representative of the dealers as a whole. The interviews, which lasted 20 to 30 minutes, were relatively unstructured, but focused on five key questions: 1. "How do the [program] targets influence formal policies or managerial attention in a given month for both you or other dealers?" 2. "How do you think about the two separate brands and respective targets, both before and after achieving the targets?" 3. "When you received the prepayments, did this affect any of your policies or attention, and how was the cash used?" 4. "How do you imagine other dealers might deal with these prepayments, and do you see them as generating potential benefits or problems for these other dealers?" 5. "What do you see as the immediate strengths and weaknesses of the current setup, where 110% has such a large immediate payoff?"

The interviews revealed relatively consistent views on the role of the incentive program in dealer strategy and policy. Respondents universally emphasized that the incentive program represents the majority of their sales profits. One dealer noted that "you gotta get there. If you don't hit the 110 number, your operating profit is next to nothing." Another stated that "it's critical to reach those plateaus for achieving profitability." They also consistently noted how the program dominates their sales strategy. "It's an integral part of how we shape and process just about everything." Another dealer explained that "(the incentive plan) is running our business. It's the most important thing. Financially, it's crucial—55-65% of operating profit." Several dealers also noted that hitting the objectives early changed their behavior, saying "it allows us to start focusing on margin," while also noting that it makes them "relaxed." Dealers also explained that the incentive program is tightly tied to the incentives of the sales managers and salespeople.

Many indicated that they carefully track progression toward the targets throughout the month, adjusting resources and pricing based on the likelihood of hitting the target. One dealer reported that "if we don't hit a fast start, then we back off. If we're close, we use incentives to get there." This resource and attention allocation also applied across brands. A dealer noted that "if we're cruising on (large-bonus model group), we'll put attention to (small-bonus model group). If we get to the 20th of the month, we'll focus on the one closest to the 110% target." One dealer noted that "we think about (the brands) entirely separately, until we're close. Then I'll move people." Another explained that "mid-month, when it looks likely to hit one more than the other, we allocate resources accordingly." Another dealer noted the difference between the brands. "We make a lot more money off (large-bonus model group) so we focus there. We're not going to take a (large-bonus model group) buyer and put them in (small-bonus model group)."

Although dealers were reluctant to discuss attempts to move customers across months and brands, several indicated that they had some ability to do so, and many more explained that they "know there are dealers who play that game," or opaquely explained "if we're having a slow month, we think about next month." Some dealers also indicated the importance of having smooth earnings across months. “You can’t have peaks and valleys in your sales year. You need to be consistent.”

The interviewed dealers indicated a mix of responses to the prepayment condition. One mentioned that “the reason I signed up was for an interest-free loan,” with some other dealers expressing similar sentiments. Another mentioned that “this is simply cash up front, and we put it in a separate account,” explaining that “my sisters are very accounting oriented.” Others, however noted the motivational effect it had on their business. One dealer noted that “nobody wanted to give it back.” Another explained that “it’s like giving my son $100—it’s going to be spent. Your back’s in the corner so you’ve got to produce.” Another also mentioned that “nobody wanted to give it back,” referring to it as “a little extra spice in the stew,” and explaining that he “didn’t want to have that conversation with the owner.” One successful dealer explained that the prepayment did not impact their behavior during the experiment because “I knew I would hit my goals. I would be hesitant to do it this year, because I don’t want it yanked back from me.” Another stated that it “always felt like there was a lot more pressure. It changed the intensity.” Two dealers also noted that the prepayment did help with cash flow.

A.7.2 Post-Experiment Survey

The post-experiment survey was conducted in July, 2018, and was distributed to the primary contacts at all 294 participating dealers. The survey asked about the use of the advanced funds as well as the respondent’s perception about how the advanced funds changed behavior throughout the month. Follow-up emails were used to attempt to increase participation, resulting in sixty-nine dealers (23%) completing the survey. Responses from the survey are presented in Tables A.5 and A.6.

Table A.5: Questions and Responses From the Post-Experiment Survey

1. Which of the following employees have financial incentives tied to SFE bonuses? (SELECT ALL THAT APPLY)

Total Responses	General Manager	Sales Manager	Sales Consultant	Service Technicians	Administrative Staff	Other
69	47	59	23	5	19	12
100.0%	68.1%	85.5%	33.3%	7.2%	27.5%	17.4%

2. Did any of these employees receive advanced funds when participating in the pilot program (i.e., received bonuses or incentives at the beginning of the month)? (SELECT ALL THAT APPLY)

Total Responses	General Manager	Sales Manager	Sales Consultant	Service Technicians	Administrative Staff	Other
69	6	5	1	1	0	10
100.0%	8.7%	7.2%	1.4%	1.4%	0.0%	14.5%

3. How did you immediately use the advanced funds when you received them at the first of each month? (SELECT ALL THAT APPLY)

Total Responses	Marketing/ Advertising	Customer Incentives	Advanced Employee Comp.	Pay Down Inventory	Recorded as Income	Set Aside Until End of Month	Other
68	7	8	2	1	7	49	6
100.0%	10.3%	11.8%	2.9%	1.5%	10.3%	72.1%	8.8%

4. How did the advanced funds influence the external pressure you felt to meet SFE targets?

Total Responses	Strongly decreased 1	2	3	No change 4	5	6	Strongly increased 7
69	3	1	5	37	2	7	14
100.0%	4.3%	1.4%	7.2%	53.6%	2.9%	10.1%	20.3%

5. How did the advanced funds influence the internal motivation you felt to meet SFE targets?

Total Responses	Strongly decreased 1	2	3	No change 4	5	6	Strongly increased 7
69	1	1	5	36	4	4	18
100.0%	1.4%	1.4%	7.2%	52.2%	5.8%	5.8%	26.1%

6. Compared to your usual management approach, how did the advanced funds change the way you approached the sales process to reach your monthly target?

	Less than before 1	2	3	No change 4	5	6	Much more than before 7
a. Closely tracked progress toward monthly SFE targets	0 0.0%	1 1.4%	4 5.8%	47 68.1%	4 5.8%	5 7.2%	8 11.6%
b. Emphasized meeting SFE targets with my salesforce	0 0.0%	1 1.4%	3 4.3%	46 66.7%	3 4.3%	4 5.8%	12 17.4%
c. Encouraged aggressive pricing to meet SFE targets	0 0.0%	1 1.4%	4 5.8%	45 65.2%	3 4.3%	4 5.8%	12 17.4%

Table A.6: Questions and Responses From the Post-Experiment Survey (Continued)

7. How did the Advanced Funds Pilot Program change your focus on either [small-bonus model group] or [large-bonus model group] before reaching either of the 110% targets?

Total Responses	No change	Increased focus on small-bonus group	Increased focus on large-bonus group	Increased focus on closest group to targets	Other
69	48	3	2	11	5
100.0%	69.6%	4.3%	2.9%	15.9%	7.2%

8. How did the Advanced Funds Pilot Program change your focus on either [small-bonus model group] or [large-bonus model group] after reaching both of the 110% targets?

Total Responses	No change	Increased focus on small-bonus group	Increased focus on large-bonus group	Other
69	54	5	4	6
100.0%	78.3%	7.2%	5.8%	8.7%

9. Do you believe the advanced funds increased or decreased your group's overall sales?

Total Responses	Strongly decreased sales 1	2	3	No change 4	5	6	Strongly increased sales 7
69	0	0	5	44	4	5	11
100.0%	0.0%	0.0%	7.2%	63.8%	5.8%	7.2%	15.9%

10. Given your experience with the Advanced Funds Pilot, what would be your preferred payment timing for SFE moving forward:

Total Responses	Strongly prefer end of month 1	2	3	Indifferent 4	5	6	Strongly prefer advanced funds 7
68	24	1	3	10	4	4	22
100.0%	34.8%	1.4%	4.3%	14.5%	5.8%	5.8%	31.9%

11. How would you describe your cash flow constraints in a given month?

Total Responses	No constraints 1	2	3	4	5	6	Very high constraints 1
68	21	10	9	11	3	2	10
100.0%	30.4%	14.5%	13.0%	15.9%	4.3%	2.9%	14.5%

Only 23 dealers reported distributing advanced funds to employees, with this distribution typically going to a general manager or sales manager. The majority of respondents (49) reported setting the funds aside until the end of the month. Twenty-three dealers reported that they agreed that the advanced funds increased pressure to meet program goals, while 26 reported increased motivation. Nineteen dealers also reported that they increased their emphasis on meeting program targets under the prepayment condition, while 17 indicated that the prepayment condition made them more closely track progress toward the monthly targets.

Sixteen dealers reported that the prepayment condition changed their approach to the two brands, with 11 reporting that it increased their focus on whichever brand was closest to the target. Twenty dealers believed that the prepayment system increased their sales, while five believed it slightly decreased them. Finally, 30 dealers would prefer advanced funds moving forward, while 28 would prefer end-of-the-month and 10 were indifferent.

Collectively, the interviews and survey indicate that for a select set of dealers, the prepayment condition changed behavior and mind-frame in many dealers while having little effect in others. This is consistent with there being a variety of management practices across car dealers resulting from both franchise law protections and other sources of heterogeneity. We draw several important conclusions from these self-reported data. First, the prepayment condition was salient enough to successfully treat some dealers with loss-framing. Second, this loss-framing was not universal, which deflates any average treatment effects estimated from the study. Third, the responses of some treated dealers align closely with the gaming behaviors we discuss in our theoretical treatment of loss framing.

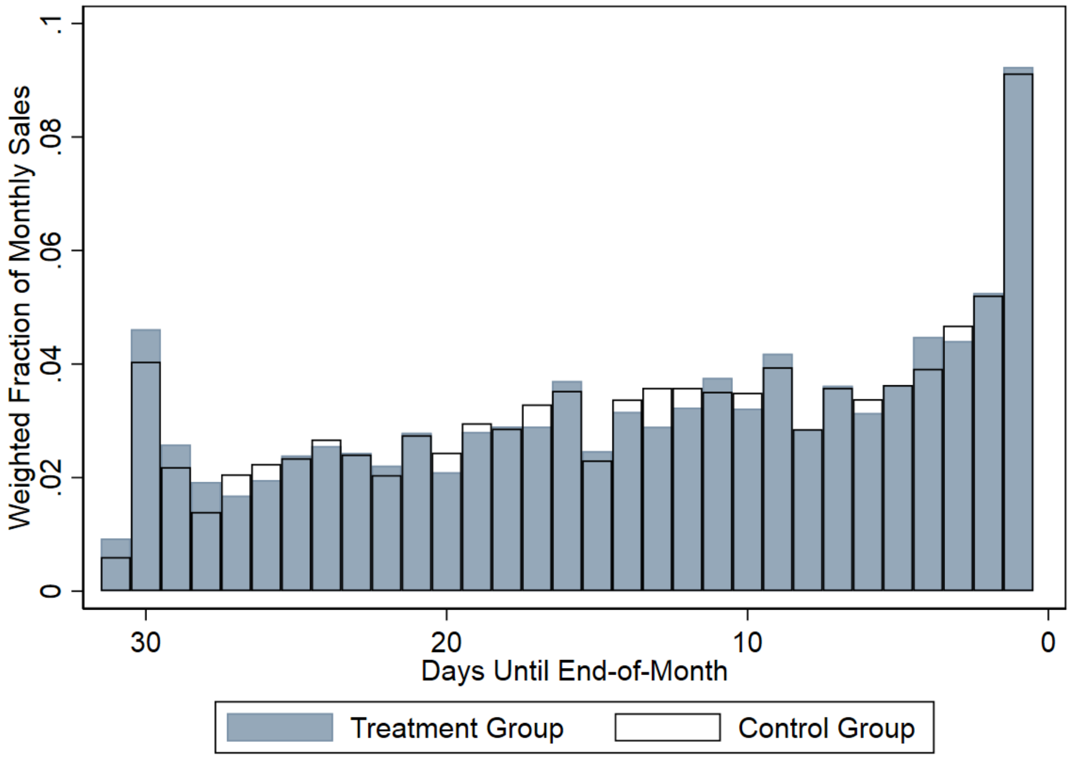
A.8 Timing of Sales

Another commonly considered "gaming" behavior among car dealers is attempting to shift sales not across model groups, but across time periods of evaluation. Ultimately, we are unpowered to provide interesting tests of this channel for potential gaming in our experiment, but we present some attempts to examine this here.

Figure A.6 shows the distribution of sales by day-of-the-month in each group during the pre-period, demonstrating some daily responsivity to the intertemporal incentives induced by these contracts.[36] As previously noted in, e.g., Larkin (2014), high-powered incentives measured over discrete time windows can result in high effort to close sales at the end of the window of measurement. Consistent with this consideration, we see that sales are particularly concentrated at the end of the month. Also consistent with Larkin (2014), we see increased sales at the beginning of the month, likely from when dealers delayed closing deals at the end of a month where the target was unreachable. As with our analysis of within-month responsivity to direct incentives, there is no statistically detectable difference in means between treatment and control groups (T-test: $p = 0.692$). There is a small distributional difference that is statistically distinguishable because of the large number $(57,005)$ of daily

[36]Each day-of-the-month is inverse-weighted by the number of days during the four-month period when dealers were open. Nearly all dealers were closed January 1-3 and on Sundays.

Figure A.6: Pre-treatment Daily Sales Timing by Treatment Assignment

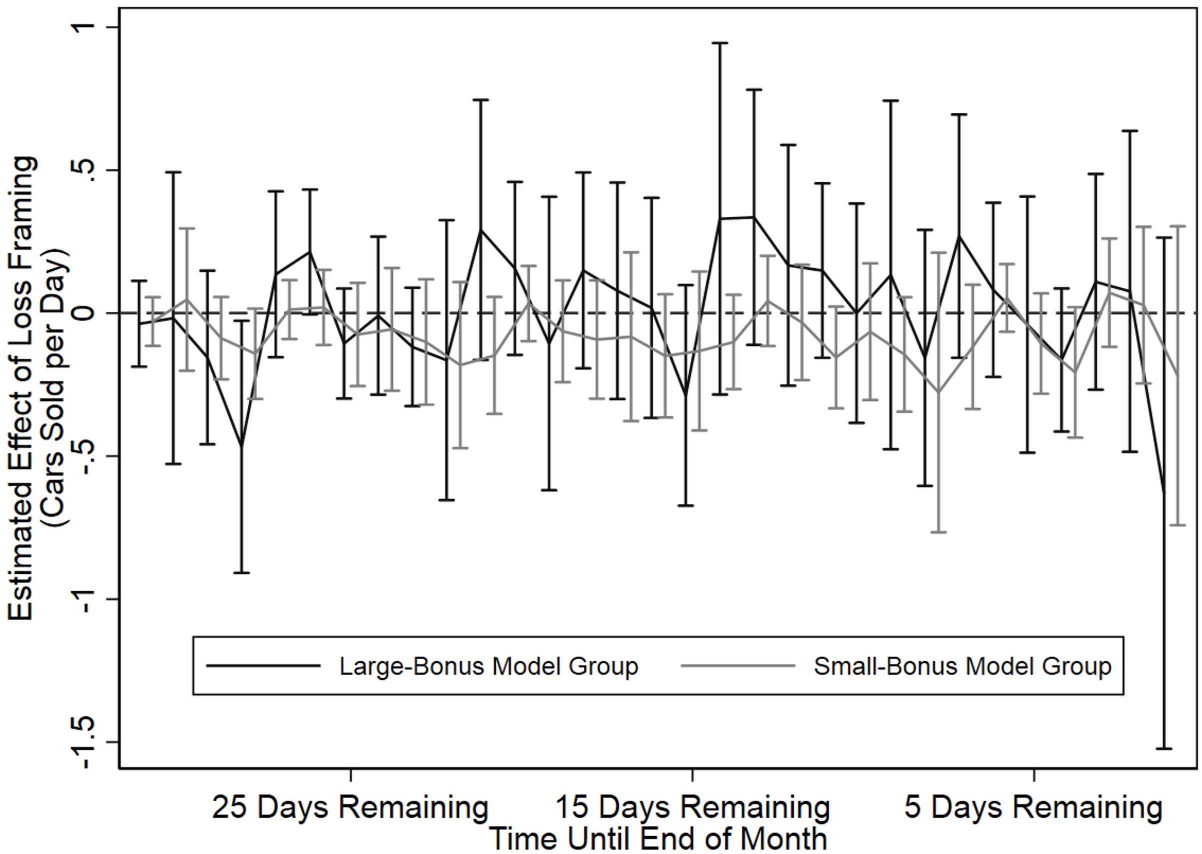


Notes: This figure shows the percentage of monthly sales occurring on each day of the month in the four months prior to the experiment. We correct for days when dealers are closed (Sundays and the New Year's holiday) by inverse-weighting sales by the number of open days during the four month period.

observations (Kolmogorov-Smirnov: $p = 0.033$).

To test for the possibility that loss framing influenced the incentives to sell cars on particular days (e.g., at the beginning or end of the month), we re-estimated regressions (3) and (5) from Table 2, restricting the data to each possible number of days until the end of the month. Results are presented in Figure A.7. As seen in the figure, we do not find evidence that treatment effects operate at a specific time of the month; however, given that the average dealer sells fewer than one car of either brand per day, on average, the size of the confidence intervals in this figure illustrate that we are underpowered to rule out substantial intertemporal effects.

Figure A.7: Day-Specific Estimated Treatment Effects on Cars Sold



Notes: This figure shows the estimated treatment effect from regressions (3) and (5) in Table 2, with the sample restricted to include only the day of the month indicated on the X-axis.