2015

Checklists and Worker Behaviour: A Field Experiment

C. Kirabo Jackson, *Northwestern University*
Henry Schneider, *Cornell University*

CHECKLISTS AND WORKER BEHAVIOR: A FIELD EXPERIMENT[1]

C. Kirabo Jackson
Northwestern University

Henry S. Schneider
Cornell University

November 22, 2014

We analyze data from a field experiment in which an auto-repair firm provided checklists to mechanics and monitored their use. Revenue was 20 percent higher during the experiment, and the effect is equivalent to that of a 1.6 percentage point (10 percent) commission increase. Checklists appear to boost productivity by serving both as a memory aid and a monitoring technology. Despite the large benefits to the firm, mechanics did not use checklists without the firm directly monitoring their use. We show that a moral hazard can explain why mechanics do not otherwise adopt checklists.

[1] We thank Dan Benjamin, George Iny, Eli Melnick, Steve Nafziger, Ted O'Donoghue, Joseph Schneider, Michael Waldman, and various seminar participants for helpful comments, and Peter Hlawitschka for excellent research assistance. Henry Schneider thanks the S.C. Johnson Graduate School of Management for research support.

Checklists involve a list of tasks to be completed and checked off, and are often employed as a means to improve worker outcomes. Checklists are used in aviation, health care, restaurant food preparation, and many other settings with the aims of reducing human error, facilitating the monitoring of workers, and providing explicit instructions to workers (Boorman 2001, Gawande 2009).[2] Checklist~~s~~ use, however, is far from ubiquitous, and there is often resistance from workers to using them (Pickering et al. 2013). In this study, we conduct the first experimental evaluation of checklists, providing evidence on their effectiveness and the mechanisms through which they may operate. We also show how moral hazard can hinder their adoption and may help to explain why their use is not more widespread.

The largest body of evidence on checklists and worker performance is in medicine.[3] In one influential study, Haynes et al. (2009) examined how the adoption of a surgery safety checklist, coupled with worker training and team building, affected patient outcomes in eight hospitals. Compared to pre-adoption levels, the rates of surgical complications and mortality fell by about 40 percent. However, Urbach et al. (2014) noted that these results, as well as those of other checklist studies, may not accurately identify the effect of checklists because the data are observational, involve hospitals that chose to adopt checklists, and include complementary procedures (e.g., worker training) that can have independent effects. Supporting these concerns, Urbach et al. (2014) compared outcomes before and after a government policy that mandated checklist use in Canadian hospitals and found no meaningful effect.

We analyze data from a field experiment conducted at a chain of auto-repair shops.

[2] For example, commercial pilots have scores of checklists, each one brief, covering a wide range of flight scenarios from taxiing on the runway to emergency situations like engine failure. Representative is the US Airways checklist for engine failure on a Cessna plane that has just six steps, starting with "FLY THE AIRPLANE" (Gawande 2009).

[3] Pronovost et al. (2006) is another landmark study in which hospitals participated in intensive-care-~~—~~unit interventions that involved checklists instructing doctors on actions such as hand washing to address the central-line-infection problem, accompanied by supporting activities. The intervention is associated with a reduction in infection rates by 66 percent. See Urbach et al. (2014) for additional references on checklist use in medicine, and Ko, Turner, and Finnigan (2011) for a critical review of this literature.

Mechanics at three shops were instructed to use a car-inspection checklist for approximately one month, while mechanics at another eight shops were used for comparison purposes. The checklist lists the components of a thorough car inspection, and made explicit to the mechanics the steps the firm wished them to conduct. Mechanics were asked to fill out the checklist ~~for each checklist~~ during the inspection and submit the completed checklists to a supervisor to both facilitate monitoring that the checklists were being used and to facilitate the monitoring of worker actions through the completed checklists. We examine visit-level outcomes for the shops from 2008 to 2013. Our main outcome measure is revenue generated by the mechanic. We also examine mechanic effort, the allocation of effort across tasks, and mechanic pay. To benchmark the checklist effect, we additionally estimate the effect of changes in performance pay, a common alternative approach to improving worker productivity.[4] We do so by exploiting quasi-experimental variation in commission rates during our sample period for the same mechanics.

There are three primary mechanisms through which checklists may operate. One is as a memory aid: An itemized list may help workers avoid errors due to forgetfulness, including from overestimating the reliability of one's memory (Ericson 2011) and ability to avoid mistakes (Camerer and Lovallo 1999, Bernheim et al. 2003), and limited attention (skipping steps in the presence of distractions). The second is to clarify best practices if workers are misinformed about the correct procedures (Bloom et al. 2013, Hanna, Mullainathan, and Schwartzstein 2014). The third is to facilitate the monitoring of worker actions, which in the current setting would involve workers documenting their inspections in writing on the checklist for supervisors to review.[5] We

[4] Over one-third of jobs in the U.S. use performance pay (Lemieux, MacLeod, and Parent 2009), and popular acceptance of its benefits motivates current proposals such as tying teacher pay to student test scores (Jackson, Rockoff, and Staiger 2014) and hospital compensation to patient outcomes (Mullen, Frank, and Rosenthal 2010).

[5] In the hospital study by Provonost et al. (2006), monitoring was enabled by empowering nurses to stop doctors when the designated procedures were not completed.

examine which of these mechanisms may be at play.[6]

Naturally, checklists can only work if they are actually used (Gawande 2014), and Bosk, Dixon-Woods, Goeschel, and Pronovost (2009) argue that simply providing checklists to workers may be ineffective "unless they are coupled with attitude changes and efforts to remove barriers to actually using them." While workplace-culture considerations are surely important, the literature on checklists has generally overlooked an inherent moral hazard problem: Just because checklist use *could* improve welfare does not mean that workers would find it privately optimal to incur the effort costs to ~~use them~~do so. To better understand the role of moral hazard, our intervention involved providing checklists to mechanics both with and without the monitoring of their use, and calculating the private benefits of checklist use to the firm versus the workers.[7]

We conduct both event study and difference-in-differences analyses, comparing the change in outcomes before, during, and after the treatment to the change in outcomes for untreated workers. During the treatment, checklist use increased from zero to 30 percent of cars, and treated mechanics generated 22 percent more revenue per car ($p<0.01$). After the treatment, checklist use returned to zero and revenue returned to pre-treatment levels. We present tests to show robustness to small-sample inference and to support a causal interpretation.

We then estimate the effect of performance pay, comparing the change in outcomes for mechanics who received a commission increase to the change for mechanics who did not receive a contemporaneous commission increase. Because the commission increases did not occur

[6] Authors in the medical literature emphasize the ability of checklists to facilitate communication in the workplace. While this is surely important, we note that better communication may facilitate worker monitoring, clarify instructions, and help as a memory aid. Because communication itself does not improve worker outcomes, we do not consider communication as a separate causal mechanism, but rather as a key facilitator of the other mechanisms.
[7] In related work, Schneider (2010) and Johnson, Schneider, and Waldman (2014) found considerable moral hazard in leasing markets. Jackson and Schneider (2011) found that social pressure could mitigate this moral hazard.

randomly, we present evidence that the results can be interpreted causally. We find that a one-percentage-point commission increase led to a 12.9 percent increase in revenue (p<0.01). A one-percentage-point increase represents a six percent increase relative to the mean commission rate, so that the checklist effect is equivalent to that of a 10 percent commission increase.

Our evidence on mechanisms points to checklists in the current setting acting as a memory aid rather than an instructional tool. First, we find little checklist effects without actual checklist use, either ~~between~~ on non-checklist visits during the treatment period, or after the treatment ended~~s~~, suggesting that the instruction and learning mechanism is not operative in our setting. Second, at different times in the experiment, mechanics used two different checklists with different item orderings. We find large increases in repairs for items at the top of the checklist compared to items at the bottom. This pattern is consistent with a memory-aid function, where the checklist prompts mechanics to check items they would have otherwise skipped, but then mechanics slacken or face time pressure to move to the next car as they proceed through the checklists.[8]

Next we explore the role of monitoring. Because it is impossible to monitor worker actions using a checklist without collecting the checklists, the collection of checklists by managers involved both (1) the monitoring of checklist use and (2) the monitoring of worker actions as evidenced on the checklists. We find that the combination of these two is important. Specifically, mechanics stopped using checklists immediately when collection of checklists ended even though mechanics had the option to continue using them.[9] Thus, the mere provision of checklists was insufficient to improve outcomes. Because the monitoring of checklist use can

[8] This also makes clear that the outcomes required the checklists themselves, and were not simply due to mechanics responding to a perception that their actions were being scrutinized by the owner or researchers (us).
[9] While mechanics were not warned of any specific punishments for not using checklists, it was clear to mechanics that ignoring the order would cause the owner and supervisors to be displeased.

only improve outcomes if checklists themselves induce more productive actions, checklists and the monitoring of their use appear to be strong complements.

To specifically explore the role of monitoring worker actions that is facilitated by checklists (as distinct from simply monitoring that the checklists were used at all), we examine worker effort allocation. In a simple principal-agent model, workers would increase effort if the firm were better able to monitor worker actions. Consistent with this enhanced monitoring, we observe an increase in the numbers of repairs conducted and hours worked during the treatment.[10] In summary, the treatment effect appears to ~~act~~ operate through checklists acting both as a memory aid and a monitoring technology, such that mechanics would not have exerted the extra effort of employing a memory aid without this monitoring, and the monitoring would not have improved worker outcomes without the memory aid.

Given the large revenue gains from checklist use, it is natural to wonder why mechanics only used checklists when their use was being monitored. Our calculations point to moral hazard. Mechanics receive only a modest share of the revenue they generate (the commission) but incur the full effort costs of using checklists. As a result, checklist use may benefit the firm but not the mechanics under the current pay structure. We find that the firm obtained an extra $317 in gross profit (revenue net parts costs and payroll) per mechanic-week during the treatment. However, mechanics earned only $15.84 per hour for the 3.48 extra hours they worked per treatment-week, which is *below* the business-as-usual rate of $16.23. Given that effort costs are convex and mechanics chose not to work the extra hours at the prevailing wage outside of the treatment, the

[10]In contrast, commission increases led mechanics to shift to higher-price repairs without increasing the number of repairs or hours. Another motivation for studying alternate approaches to improving worker productivity is that performance pay can induce undesirable worker behavior when the rewarded performance does not capture all dimensions of firm output (Holmstrom and Milgrom 1991, Baker 1992). This problem may be particularly acute in multi-task problems such as auto repair, which involves the dual tasks of doing inspections and conducting repairs, Other empirical studies of the trade-offs of performance pay include Jacob and Levitt (2001), Freeman and Kleiner (2005), Rosenthal and Frank (2006), Mullen, Frank, and Rosenthal (2010), and Larkin (forthcoming).

added pay is likely insufficient to incentivize mechanics to adopt checklists on their own. A more basic indication of moral hazard, however, is simply that mechanics chose to discontinue checklist use, and revenue returned to pre-treatment levels, when the monitoring ended, despite the large gains to the firm.

Note that the firm has adopted checklists in all of its shops as of one year after the experiment (Fall 2013). Conversations with the owner and supervisors revealed that checklists were not previously used because mechanics perceived checklists as requiring extra effort and interfering with their autonomy; and the owner and supervisors suspected benefits but were sufficiently unsure to not mandate their widespread use. Thus, checklist adoption was hindered by worker reluctance to exert the necessary effort under moral hazard, and by firm uncertainty about its effectiveness.

This study makes several contributions. First, it broadens our understanding of checklists as a workplace technology, including the mechanisms through which they may act, and when they may be most effective. Second, it adds to the literature on management practices (e.g., Lucas 1978, Bertrand and Schoar 2003, Bloom et al. 2013) by identifying the effects of individual procedures – checklists, performance pay, and monitoring – that are often components of bundled management interventions. Third, it demonstrates how moral hazard can cause suboptimal technology adoption by workers, and how monitoring may enable their use.

The paper proceeds as follows. In Section I, we describe the data and some institutional details. In Section II, we describe the checklist and commission-increase experiment and quasi-experiment, and report the graphical event-study results. Section III contains the main results. Sections IV and V contain robustness checks and an investigation of mechanisms. Section VI examines the moral hazard problem. Section VII concludes.

I. DESCRIPTION OF THE DATA AND AUTO-REPAIR BUSINESS

We worked with 11 shops in one metropolitan area from a United States auto-repair chain. The shops are centrally owned. Under a confidential data-sharing agreement, the firm provided detailed data on all customers, cars, repairs, and employees for the period June 23, 2008 ~~and~~ to June 22, 2013. The source of revenue for the shops is repair charges.[11] Table 1 shows the repair categories in which repairs were conducted in at least one percent of visits, and the mean charge for each. Roughly half of visits involved oil changes, followed by brake repairs (16 percent) and alignment/suspension repairs (12 percent). Among visits with oil changes, 37 percent involve additional repairs. A primary aim of the firm is to expand routine-maintenance visits (e.g., oil changes) to include repairs that are discovered while the car is in the shop.

A visit is defined as an invoice, which is essentially one contiguous repair visit by a particular car, and may entail a simple oil-change visit or involve the customer leaving the car overnight or returning several days later for scheduled work. The mean amount charged per visit, or mean revenue per visit, is $191~~0.61~~. This amount is labor charges plus parts charges minus coupons and other discounts, where 52 percent of visits involve a discount. Thus, our revenue measure reflects what the customer actually pays. (Results are similar when discounts are excluded.) Repairs range from inexpensive, such as oil changes and windshield wipers, which have mean charges of $24 and $25, respectively, to labor and parts-intensive work such as brake and suspension repairs, which are $300 or more.[12]

[11] The firm uses a rule-of-thumb to set repair charges such that the internal cost of parts represent approximately 20 percent of the total repair charge. The remaining 80 percent encompasses the labor charges based on the shop's posted hourly labor rate and the standard book time for that repair, and a mark-up that is allocated to parts charges. For example, if the internal cost to the shop for a part is $10, the standard labor time for that repair is half an hour, and the posted hourly labor rate at the shop is $70, then the customer would be charged about $90 total, with $35 reported to the customer as labor charges and $55 reported as parts charges.

[12] Appendix Table A1 reports the most common invoice items in the sample. Figure A1 shows revenue per visit and number of visits per shop by month for 2009-2011 (prior to the treatments). A seasonal trend appears, which will be important to control for.

There is a single owner for all shops, and a shop manager for each shop who oversees the mechanics and interacts with customers. Two to three mechanics are on duty in each shop at all times and conduct the inspections and repairs. During the sample period, 84 mechanics and 25 managers, who also conduct some repairs, worked at the shops.[13] Table 2 presents summary statistics on all mechanics during the sample period. Including repairs by managers, ~~Sixty-one~~ 61 percent of visits are handled by mechanics who are paid primarily on commission ("commission mechanics"). Commission mechanics receive the maximum of a commission and an hourly rate, such that their entire pay is commission except in rare very ~~particularly~~ slow weeks where the hourly rate ensures a minimum pay. These mechanics have a commission rate between 14 to 20 percent, with a mean of 17.6 percent, and an hourly rate between \$9 to \$12.[14,15] Most of the remaining visits are handled by mechanics that are paid on an hourly rate (\$9 to \$12 per hour) or a flat rate (\$20 to \$24 per hour of labor billed to customers). On average, mechanics work 45.5 hours per week over 4.89 days, earning \$705 per week. As a measure of experience, the average number of days that mechanics have worked at the firm (on duty) since 1998 is 703. Commission mechanics have even more experience, having worked for 839 days on average.[16]

While most of the data are complete, the payroll data set, which reports time worked and weekly pay, has some limitations. First, it contains only mechanics employed by the firm at the end of October 2012 (through the end of the treatment periods) but not mechanics that left the firm before October 2012.[17] Second, it is incomplete prior to January 1, 2012 (two months prior

[13] The~~se 109 mechanics includes~~ 25 managers in our sample ~~who~~ only occasionally conducted repairs~~,~~ and ~~who~~ are not included in Table 2 to avoid distorting several statistics in the table (e.g., number of visits per day).

[14] Mechanics are paid every Wednesday for work over the prior Monday to Saturday. Shops are closed on Sundays.

[15] Two mechanics whose compensation is primarily hourly also receive commissions, of 2 to 3 percent. Managers, who occasionally conduct repairs, have a commission rate for those repairs of 10 percent.

[16] Therefore, turnover is less central for this firm than for some others such as in Lazear (2000). Regardless, we examine worker-level outcomes, and so changes in worker composition do not directly play a role in our estimation.

[17] During the treatment periods (~~Between~~ March ~~and~~ to October 2012), two mechanics (out of 59 in 2012) left the firm and were replaced by new mechanics. Both were at non-treatment shops.

Formatted: Font color: Auto

Formatted: Font color: Auto

Formatted: Font color: Auto

to the treatments).[18] Thus, the hours and pay components of our analysis are based on the mechanics and customer visits starting on January 1, 2012 (treatments began in March 2012). The results that use non-payroll data (revenue, repairs, and other outcomes) are very similar when limited to the same time period.[19]

II. THE EXPERIMENTS

The primary objective of this study is to estimate the causal effects of a checklist intervention. For comparison purposes, we also wish to estimate the causal effects of an increase in incentive pay for the same population of workers and ~~set of~~ outcomes. The following two experiments provide independent sources of variation for estimating these effects: (1) a controlled experiment in which checklist were provided and checklist use was monitored, and (2) quasi-experiments in which commissions were increased.

A. The checklist experiment

To identify the effects of the checklist intervention we worked with the owner to implement a series of checklist interventions at three of the 11 shops. The owner asked the mechanics to use and fill ~~-~~out checklists for cars that came into the shop. The mechanics were told to use checklists on as many cars as they comfortably could, but were not instructed to use them on every car, nor were they told to use checklist on any particular type of car. An example of a checklist visit is as follows: a customer brings in her car for an oil change, and while the customer is waiting, the mechanic inspects the car and completes the checklist. The mechanic

[18] The payroll data also only documents commission and hourly pay but not salary pay, which is relevant for managers, who are primarily paid on salary but also receive commissions for any repairs they do. This is less relevant for our analysis because few repairs overall are done by non-managers.

[19] Note that we worked with the same firm on a different experiment that occurred primarily in 2013. For completeness, in all of specifications we include a dummy variable for the second experiment so as to difference out any of these other treatment effects. Results are unchanged when these treatment periods are excluded.

then shows the completed checklist to the manager and the manager suggests the repairs to the customer. The customer consents to none, some, or all of the recommended repairs. There were 717 visits during the treatment periods, of which 215 or 30.0 percent had a checklist.[20] Given the possible selection of which cars received checklists during the treatment period, we consider all visits during the treatment period as treatment visits, and compare these outcomes to those in the non-treatment period. However, we will show that the effects are driven by checklist use.

The rationale for the checklists is to induce the mechanics to conduct more thorough inspections in order to identify more repair work. Many other auto-repair firms use checklists, and the supervisor told us that all mechanics were well aware of checklists and none at the shops in our sample used them. Expecting that simply providing checklists would not generate meaningful checklist use, the owner required mechanics to submit the completed checklists to a supervisor every week, which were subsequently returned to the firm headquarters.[21] This is the monitoring aspect of the intervention. The subjects were aware that they were part of an experiment run by the firm with assistance from academics to learn about checklist efficacy and so the intervention could be considered a framed field experiment.[22]

There were four checklist treatments at three shops (one shop had two treatments). Each treatment occurred for a series of days so that the treatment turned on and off, allowing us to observe the effect of both occurrences.[23] As Table 3 shows, these three shops were similar to the

[20] Partial compliance is common in other checklist studies. For example, in Haynes et al. (2009), medical staff compliance with items on the surgical checklist increased from 34.2 percent to 56.7 percent during the intervention.
[21] Copies of the completed checklists, with links to the associated invoices, were provided to us.
[22] A framed field experiment is defined as having suitable subjects in their natural environment with real incentives that are aware of the experiment (see Table 1 in List and Rasul 2011 for a taxonomy).
[23] Treatments occurred at shop 1 from July 10 to August 11, 2012, at shop 2 from March 3 to March 26, 2012 and November 5 to November 15, 2012, and at shop 13 from July 10 to August 6, 2012.

eight non-treatment shops.[24] The general windows of treatment dates correspond roughly to when we were in contact with the firm. The exact treatment dates within the general windows correspond to when the ~~operations~~ supervisor visited the shops for unrelated reasons, and was able to drop off and pick up the checklists at the same time.[25]

During the treatment, mechanics were given checklists that listed different car parts and maintenance items that the mechanics should inspect. The items include checking the oil level, brake components, and so on. Two distinct checklists were used, which had very similar sets of items but different layouts and orderings. Checklists 1 and 2 were used in 34 and 66 percent of visits, respectively. Figure~~s~~ 1~~4~~ ~~and 5~~ shows a sample of checklist 1 (Appendix Figure A2 shows a sample of checklist 2)~~s~~.[26] The high level of detail written on many of the completed checklists indicates that the mechanics actually used the checklists and did not carelessly fill them out with erroneous information. We also show in Section V that the checklist layout affected which repairs were conducted, again indicating that mechanics actually used them. While filling out the checklist itself is not time consuming (taking a minute or two), actually conducting a more-thorough inspection requires a non-trivial time cost for the mechanic.

After the treatment periods, the mechanics were told that the checklists would no longer be collected, but the treatment shops still had a significant number of blank copies of the checklists on hand and ~~they~~ mechanics were free to use them. The supervisor then checked with the mechanics in the weeks subsequent to the treatment periods to see whether mechanics used

[24] The three shops were chosen due to their proximity to the central office where the supervisors are located. Because distance to headquarters is a time-invariant shop-level characteristic, this does not affect any of our results, which rely exclusively on within-shop variation over time.

[25] The ~~operations~~ supervisor visits all shops every week or two for routine business operations.

[26] The checklists are similar to a typical state safety inspection form. However, the shops are in a state without mandated safety inspections and so many of the cars may not have received a thorough inspection in the recent past. Nineteen U.S. states currently require annual or biennial safety inspections for passenger cars, typically including a checklist for the inspection process. See http://en.wikipedia.org/wiki/Vehicle_inspection_in_the_United_States for more information (last accessed on April 29, 2013).

any of the remaining checklists. This ~~will~~ revealed whether mechanics ~~would~~ continued ~~to~~ using~~e~~ checklists on their own accord or required monitoring.

In our primary analysis in Section III, we provide results from a range of regression models and outcomes. However, because the number of treatments is small, we first establish visually with an event-study analysis that the observed treatment effect is not ~~driven by~~ due to random chance. We regress the log revenue per visit on mechanic and shop indicator variables to remove mechanic and shop-level means, week indicators to remove any shocks that are common across shops, and a series of indicator variables for each four-week period ("month") relative to the treatment periods (i.e., indicators for event time). These event time indicators map out the dynamic treatment effect of the intervention.[27] Standard errors are clustered at the shop-~~week~~ year level.

Figure 2~~1~~ plots the estimated coefficients for each month relative to the treatment month for the 24 months before and 12 months after the intervention, ~~(~~along with one standard-deviation error bars on either side~~)~~. The treatment month is month zero and the estimated event time coefficients are centered on zero. Thus, this event-~~-~~time plot depicts the *change* in mean invoice amount (net of shop, mechanic, and time effects) for mechanics in treatment shops relative to the *change* over the same period for mechanics in control shops. To test the null hypothesis that the outcomes during the treatment month are different from those during the non-treatment months, we use the values for the 34 non-treatment monthly periods to compute random variability in monthly mean residuals. We then use this estimated variability to form a 95-percent confidence interval for the monthly non-treatment period outcomes. This interval

[27] We chose a monthly period because this is approximately the mean treatment-period duration. We exclude the second treatment period for shop 2 from the event-study analysis because the graphs use the periods before and after the treatment periods, causing the months around the first and second treatment periods for shop 2 to overlap. The second treatment period at shop 2 ~~also~~ was only ~~lasts for~~ eleven days, compared to an average of 28 days for the other three treatment periods, which interferes with statistical inference.

represents the range within which 95 percent of the monthly mean residuals would fall by random chance if the treatment had no effect. The 95-~~-~~percent confidence interval is also included in Figure 2~~1~~.~~ A direct visual test of the hypothesis that the treatment monthly mean differs from the control monthly means is whether the treatment monthly mean lies outside the 95 percent confidence interval.~~

Before showing the effect on revenue (our measure of output), we show the effect on the fraction of visits in which mechanics used a checklist (our measure of treatment intensity) in the top panel of Figure 2~~1~~. Consistent with increased checklist use due to the increased monitoring (i.e., checklist collection), checklist use jumped during the treatment period among treatment mechanics. Checklist use then returned to zero after the treatment period, showing that mechanics took no independent initiative to use checklists. If the checklist treatment had a causal effect on revenue, there should be a spike in revenue coinciding with the treatment~~ period~~.

The bottom panel of Figure 2~~1~~ shows the event-~~-~~study for log revenue per visit. There is a clear increase in revenue during the treatment (~~large ~~black dot) and a return to pre-treatment levels afterward. The variability in the data suggests that this spike did not happen by chance. The standard deviation of the monthly means is 0.089 such that 95 percent of period means would lie within 0.178~~5~~ and -0.178~~5~~ by random chance. The actual deviation from the non-treatment mean is 0.260, 2.92~~65~~ times the standard deviation of the non-treatment monthly means, and would happen by random chance with less than one percent probability. Furthermore, the mean residual for the treatment period is so far outside the underlying variability that it is the largest spike in revenue during the entire five-year sample period, and is more than twice as far from the grand mean as the next highest value. The timing of this "outlier" spike at exactly the treatment period is compelling evidence of a causal interpretation of the treatment on revenue.

Furthermore, because differential trending between treatment and control workers appears to be absent, the causal effect of this experiment should be identifiable from a difference-in-difference analysis. We describe this empirical strategy in Section III. We also show in the top panel of Figure A3~~2~~ in the Appendix the kernel density plots of revenue shifting for the treatment.

B. The commission-increase quasi-experiment

To identify the effects of paying a higher commission, we exploit within-worker variation in commission rates. For mechanics that received a commission increase during the sample period ("CI mechanics"), we compare their outcomes before and after the commission increase. This within-worker approach avoids using between-mechanic variation in commission rates, which could be contaminated by between-mechanic unobserved productivity differences. To account for any unrelated events that may coincide with the timing of the commission increases, we compare the change in outcomes for CI mechanics to the change in outcomes for mechanics without coincident commission rate changes ("non-CI mechanics"). This approach is valid as long as the CI mechanics did not have changes at the time of~~as~~ the commission increases that were systematically different than those of the non-CI mechanics. We provide evidence of this now.

There were 16 instances of commission increases ~~among the 108 mechanics~~ in our sample, which are shown in Table 4. The commission increases (with the exception of one) occurred in ~~single~~ one-percentage-~~-~~point steps. Because the mean commission rate among commission mechanics is 17.6 percent, these one percentage point increases represent a compensation increase of approximately six percent. Interpreting any change in outcomes corresponding to the commission increases as causal requires that (1) the outcomes of CI mechanics were not already on an upward trajectory before the commission increases, and (2) differences in outcomes between CI and non-CI mechanics coincides with the timing of the commission increases. For evidence on these two conditions, we present a graphical event-study analysis.

Our event-study analysis for commission increases uses the same method as that for the checklist experiment. Specifically, we regress the log of revenue per visit on mechanic and shop indicator variables to remove mechanic and shop-level means, week indicators to remove any shocks that are common across shops, and a series of indicator variables for each four-week period relative to the ~~four-week period~~ date of the commission increases (i.e., indicators for event time). These event-time indicators map out the dynamic ~~treatment~~ effect of the commission increase.[28] As before, standard errors are clustered at the shop-~~week~~ year level. Figure 3~~2~~ plots estimated monthly effects for 24 months ~~relative prior~~ before through 24 months after a commission increase, ~~(~~along with one standard deviation error bars on either side~~)~~. The ~~month of the~~ commission increase occurs at the beginning of ~~is~~ month zero and the event-~~–~~time coefficients are centered on zero for the pre-increase months. Thus, this event-~~–~~time plot compares the *change* in mean invoice amount (net of shop, mechanic, and time effects) before and after a commission increase to the *change* over the same period for mechanics that did not have coincident commission increases~~ at the same time~~.

As with the treatment analysis, we use the values of the pre-increase monthly periods to compute the confidence interval within which 95 percent of the monthly mean residuals would lie by random chance under no commission effect. The 95-percent confidence interval also appears in Figure 3~~2~~.~~ A direct test of the hypothesis that the post-increase monthly means differ from the pre-increase monthly means is whether the post-increase means fall outside the 95-percent confidence interval more than five percent of the time.~~

[28] We chose a monthly period to match the monthly period of the checklist-treatment figure.

The top panel shows the event-time graph for the commission level. As expected, there is no appreciable change in the commission rate prior to the increases. However, at time zero, ~~the month of the commission increase,~~ the commission increases sharply by about one percentage point. If there is a causal effect of the commission rate on revenue, there should be a similar increase in revenue in month zero that is sustained over time. The lower panel presents the event-time graph for revenue. The standard deviation of the pre-commission-increase monthly mean is 0.074 such that 95 percent of monthly means should be between -0.148 and 0.148 of the pre-commission-increase mean if the commission effect were zero. Consistent with no pre-existing trending, none of the monthly means leading up to the commission increase are outside of this range. However, after the commission increase, there is a visible increase in the invoice amount for CI mechanics relative to non-CI mechanics. In the 24 months after the commission increase, 14 months (58 percent of months) were above the 95-~~-~~percent confidence interval ~~upper bound~~ for the pre-increase monthly means, and all of the post-increase monthly means are above the pre-increase grand mean, indicating a clear statistically significant increase in revenue that occurs at the exact time of a commission increase. The bottom panel of Figure A3~~2~~ in the Appendix shows the kernel density plots of this shifting in revenue for the commission-increase quasi-experiment.

In summary, Figure 3~~2~~ shows that (1) CI mechanics had larger revenue increases after a commission increase relative to non-CI mechanics, (2) this increase is not due to pre-existing trend differences between CI and non-CI mechanics, (3) the timings of the revenue increases and commission increases coincide exactly, and (4) the increases are too large and persistent to be explained by random chance. These results indicate that the commission quasi-experiments are amenable to a difference-in-difference estimation strategy. We present this below.

III. MAIN RESULTS

A. Empirical strategy

We combine the checklist experimental variation with quasi-experimental variation in commission rates to measure the effect of checklists and performance pay on worker productivity. The key variation that we exploit is within mechanics over time, using a difference-in-difference regression framework. We estimate the change in outcomes at the times of the checklist intervention and commission-rate changes for the affected mechanics and compare this to the change in outcomes at the same times for the control mechanics. This comparison is captured in the following equation, which we estimate by OLS,

[1] $$Y_{istv} = \beta_1 Commission_{it} + \beta_2 Checklist_{st} + \delta X_{istv} + \alpha_i + \gamma_s + \tau_t + \epsilon_{istv}.$$

Y_{istv} is revenue amount (or log revenue amount) for mechanic i in shop s at time t and visit v. $Commission_{it}$ is the commission rate for mechanic i at time t. $Checklist_{st}$ is equal to one if the checklist treatment is occurring at shop s at time t, and zero if not. α_i are mechanic fixed effects, which control for between-mechanic differences in revenue per visit. γ_s are shop fixed effects, which control for between-shop differences in revenue per visit. τ_t are year-by-week fixed effects, which control for any seasonality or other firm-wide time-specific effects. X_{istv} includes car and customer characteristics. Exploiting the high-frequency data at the visit level and including X_{istv} helps to improve statistical precision given the modest number of treatment mechanics. Finally, ϵ_{istv} is an idiosyncratic error term.

The key identifying assumption that allows us to uncover the causal effect of the checklist intervention is that no other systematic changes that would affect revenue coincided with the treatment periods. This condition is likely satisfied because the timing of the treatments was determined by when a supervisor visited the shops for unrelated reasons and was able to drop off checklists. Also, we were given assurances by the owner and supervisors that everything was "business as usual" during the experiment. The key identifying assumption that allows us to uncover the causal effect of commission increases is that nothing else that would increase worker revenue coincided with the timing of the commission increases. This condition is likely satisfied because mechanics received commission increases as a reward for good work and to increase retention, and hence the increases are based on past rather than projected future performance. We report further evidence in support of the identifying assumptions in Section IV.

Because the checklist experiment occurs at the shop level, we cluster standard errors at the shop-year level. Because these standard-error calculations rely on large-sample asymptotic methods, we show robustness to other small-sample robust-inference methods in Section IV.

B. Effects of the checklist treatment and performance pay on revenue

The main regression results are in Table 5. Columns 1 and 2 show the coefficients on the treatment and commission variables. These baseline models control for time effects with a year-week indicator, day-of-week indicators, shop effects with a shop indicator, and mechanic effects with a mechanic indicator. The unit of observation is the individual visit and the unit of measurement is dollars. As such, the estimated coefficients on treatment and commission represent the difference in the mean revenue per visit within a mechanic during the treatment versus before and after the treatment, and before versus after a commission increase.

The results in columns 1 and 2 indicate that the treatment and the commission increases led to more revenue per visit. In column 1, coefficient on treatment is 41.60 ($p<0.01$) indicating that the mean revenue per visit is about $42 higher during the treatment period. The coefficient on the commission rate is 28.82 ($p<0.01$) indicating that the mean revenue per visit is about $29 higher after a one percentage point commission increase. Column 2 additionally includes a comprehensive set of controls about the car itself, including mileage (as dummies in 25,000 mile intervals), age (as dummies in two-year intervals), and make, customer gender, and whether the customer is a business or has an account with the shop~~ or not~~. Results are similar, confirming that the timing of the treatments and commission increases are unrelated to these factors.

Because it is helpful to present results in relative terms, and because the log-revenue model fits the data better than the level-of-revenue model, columns 3 and 4 present the same models but using log~~ of~~ revenue. As expected, the treatment increases revenue per visit by 21.6 percent ($p<0.01$) while a one-percentage-point increase in the commission rate increases revenue per visit by 12.9 percent ($p<0.01$). The ratio of these two estimates provides the increase in the commission rate that would have an equivalent effect as the treatment. The results in c~~C~~olumn (4)~~7~~ show that the treatment effect is equivalent to that of a 1.7 percentage point commission increase ($p<0.01$). The mean commission rate among mechanics that are primarily paid on commission is 17.6 percent, and so the treatment effect is equivalent to that of a 10 percent increase in the commission rate. The 95~~-~~percent confidence interval bounds for this ratio are 1.1 and 2.3 percentage points, which is equivalent to an increase in the commission rate of 7 to 14 percent.[29]

We next compute the standard deviation of mechanic skill and use this to benchmark the treatment effect in terms of skill. We follow Jackson and Schneider (2011) and proxy for skill with the residual revenue associated with the worker during non-treatment periods. Specifically, we estimate equation [1] without mechanic fixed effects, using data from the non-experimental periods. We then take the mean mechanic-level residual as our output-based measure of skill.

[29] Column (1) of Table 6~~7~~ shows that the treatment effect for operating profit per visit, measured as revenue minus parts and labor costs per visit, is similar to that for revenue per visit.

This measure identifies how much revenue a mechanic tends to generate after controlling for shop and time effects, compensation structure, and the types of cars and customers they handle. The standard deviation of the mechanic ~~fixed effects~~ mean residual is 0.59, indicating that the treatment is equivalent to an increase in skill of 0.37 standard deviations, or approximately the difference between a median-skill worker and one with skill at the 63rd percentile of the distribution.[30]

IV. ROBUSTNESS CHECKS

Before presenting evidence on the underlying mechanisms, we summarize several robustness checks ~~we conduct~~ that help to establish that our estimates represent real causal effects, including ~~to demonstrate~~ that our results are robust to small-sample inference given our small sample size.[31]

As mentioned above, we cluster the standard errors throughout at the shop-year level to account for the possibility that outcomes within shops are correlated over time. We also follow Cameron, Gelbach, and Miller (2011) and show that our tests of significance are robust to multi-way clustering at both the shop-year level and week level. However, as pointed out in Bertrand, Duflo, and Mullainathan (2004~~1~~), clustering the standard errors can cause an over-rejection of a true null hypothesis. To address this possibility, we follow their suggestion and estimate the effects of placebo treatments. The placebo treatments were created for the actual treated shops, and, using the actual temporal spacing of the true treatments (the treatment in shop 2 occurred four months before those in shops 10 and 13), we created placebo treatments for each of the 1318

[30] Note that because these worker effects are a generated variable, they may overstate the true variation in skill and hence the treatment effect may represent an even larger range of the skill distribution.
[31] Note that having a small number of treated subjects is not uncommon in labor-market field experiments. For example Shearer (2004) has nine treated subjects and Bloom et al. (2013) has 11 treated firms (and 14 plants).

possible increasing number of days prior to the actual treatment dates, with no overlap with the true treatment shop-dates.[32] Similarly, we estimate the effect of 1600 placebo commission increases per mechanic, where each placebo involves a randomly timed commission increase for each of the 16 CI mechanics.[33] The distributions over the estimated placebo treatment and commission effects are in Figure 4~~3~~. For both variables, the actual estimated effect is larger than 97.5 percent of the placebo replications. As a final check, we follow Cameron, Gelbach, and Miller (2008) who argue that a more conservative test is to use a wild bootstrap clustered t-statistic. For both variables of interest, the estimated t-statistic is larger than 97.5 percent of the wild bootstrap clustered t-statistics.

In the Appendix, we report several additional checks, including to show that the commission results are not explained by business stealing from other mechanics at the same shop or shop-level changes that might coincide with the timing of commission increases. We also show that the treatment and commission effects are not due to mechanics shifting effort across time, and report results of permutation tests for the individual checklist treatments (across shops) and commission increases to show that the results are not due to outlier treatments or mechanics.

V. MECHANISMS

The results thus far indicate that the estimated effects can be interpreted causally. For the remainder of the paper, we investigate the underlying mechanisms and their implications.

A. Are the revenue increases moving forward future business or hurting repeat business?

Given that the majority of customers represent repeat business, one possibility is that the increased work from the interventions are viewed unfavorably by customers and reduce repeat

[32] We exclude the second treatment period for shop 2, which is only 11 days long, because including it would cause an overlap of the placebo treatments for shop 2 with the actual earlier treatment period treatment for shop 2.
[33] ~~Note that w~~We control for the actual commission increases to account for any overlap with real increases.

business. We test this by estimating the likelihood that a car returns for another visit within six months; ~~(~~because there is more than nine months after the most recent treatment period in the sample, right censoring is not a significant concern~~)~~. This result is in column (2) of Table 6~~7~~. The linear probability model indicates no effect on the probability of repeat business.[34]

B. Are mechanics actually finding more repairs?

Because customers do not consent to all recommended repairs, it is possible that the positive treatment effect on revenue is due to customers agreeing to more repairs rather than mechanics finding more repair work. To test this, we hand ~~-~~entered the recommended repairs for all visits of each treatment, the month before each treatment, and the corresponding month in the year before each treatment.[35] If the revenue increases are only due to customers consenting to more recommended repairs, then there will be no increase in recommended repairs. We test this in column (3~~7~~) of Table 6~~7~~. The treatment effect for the log of revenue per visit for recommended repairs is 0.407 ($p<0.05$), which is larger than that for actual repairs and suggests that mechanics conduct even more thorough inspections than suggested by the actual revenue increases.

C. Are mechanics working on more cars or working harder on each car?

Another possibility is that mechanics are generating more revenue per car but working on fewer cars. We test this by estimating the model at the mechanic-week level (from Section IV) but using number of visits per week by the mechanic as the dependent variable. The results are in column (4)~~3~~ of Table 6~~7~~. The mean number of visits per week is 16.54 and the estimated effects

[34] We also estimate a duration model (a Cox proportional hazard model) that uses the full sample accounting for the right censoring that occurs at the end of the sample period. Results are very similar. We additionally estimate models in which the dependent variable is the revenue of the subsequent visit, and the treatment and commission increases have no meaningful effect here as well.

[35] The recommended repairs and charges were in an inaccessible computer format and required hand entering, which precluded~~s~~ having recommended repairs for the entire sample period.

are -0.634 (standard error of 0.836) for the checklist treatment and 0.489 (standard error of 0.358) for the commission increases. Thus, mechanics are not servicing fewer cars.

Given the increase in revenue per car, but not in number of cars per mechanic-week, mechanics were either doing more repairs per car (of the same types of repairs) or instead shifting to higher-price repairs without doing more repairs per car. We investigate this by examining the number of repairs conducted per visit, in column (5)~~4~~, and the number of minutes each mechanic worked per week, in column~~s~~ (6)~~5~~. These two outcomes reveal notable differences between the effects of commission increases and the checklist treatment. Commission increases did not affect the number of repairs per visit or the amount of time at work. In contrast, checklists increased the number of repairs per visit by approximately 12 percent ($p<0.01$) and the number of minutes per week by 7 percent or 213~~2~~ minutes ($p=0.05$). In summary, checklists generated greater effort as measured by the number of repairs and time worked, while commission increases had little effect on the number of repairs or time worked.

The above results suggest that checklists improved outcomes in part because they allow the firm to better monitor specific worker actions. We can see this because, whether or not checklists *also* function as a memory aid or learning tool, we find that mechanics did not exert the effort to use checklists without the firm collecting the checklists. If the monitoring of specific mechanic actions via checklists were not relevant, then mechanics could have submitted the checklists without exerting extra effort (by randomly marking up the checklist or not filling out much of the checklists at all). But given the increase in mechanic effort associated with checklist use, it is evident that the monitoring of mechanic actions via checklists played a role.

D. Are mechanics doing more expensive repairs?

A remaining possibility is that the revenue increases reflect mechanics shifting to more expensive repairs. We test whether the treatment and commission changes increase revenue by inducing mechanics to substitute from low to high-price repairs. We estimate the effects of the two interventions on the probability of each repair type, and then conduct a rank-order test of the correlation between the estimated effect and the mean price of each repair type. If mechanics are substituting to more expensive repairs, the correlation will be positive.

Table 7~~8~~ shows repair types ordered by mean repair price, and the estimated treatment and commission effects on the probability of each repair type. The correlation between the rank of the repair price and the rank of the commission effect is 0.85. In contrast, the correlation for the treatment is -0.17. Thus, checklists led to more repairs overall in a way that was modestly skewed toward lower-price repairs. In contrast, commission increases led to higher-price repairs, which is consistent with the stronger incentives to generate revenue.

For additional evidence on this point, we return to the recommended-repairs analysis of Section V.B. If commission increases generate more revenue because they induce mechanics to seek out higher-price repairs, we should see more recommended repairs. Column (3~~7~~) of Table 6~~7~~, however, shows that commission increases did not affect recommended-repair revenue despite the increase in actual-repair revenue. Thus, commission increases appear to have led mechanics to convince customers to *consent* to higher-price repairs, generating a "pushy salesman" effect. Insofar as this shifting toward more expensive repairs reflects mechanics exploiting their informational advantage over customers (Hubbard 1998, Schneider 2012), the results highlight the possible deleterious side effects of high-powered incentives: workers may maximize the incentivized output (in this case, revenue) without improving behavior on the primary inputs (in this case, effort and thoroughness).

E. Did the checklists themselves affect mechanic behavior?

During the treatment, mechanics conducted more repairs per car and worked longer hours. The likely mechanism is that checklists prompt mechanics to check for repairs that they would otherwise overlook. For evidence that the content of the checklist itself induces mechanics to check for additional repairs, we compare the outcomes between visits in which two distinct checklists were used. The two checklists contain very similar sets of items but have different orderings (Figures 1~~4~~ and A2~~5~~ are samples). Because mechanics likely start at the top of the checklist and work down, the effects will be largest for items at the top of the checklist if there is any decay in mechanic attention over the course of the inspection. We test for this possibility.

We estimate the treatment effect for the items at the top of each checklist to see if checklist order affects repair frequency, and hence whether checklists themselves affect behavior. The items at the top of checklist 1 are windshield wipers ("Wiper Blades") and lights ("Stop Lamps," "Headlamps," "Tail Lamps," "Marker Lamps," and "License Lamps"). Providing excellent variation, wipers and lights are the bottom of checklist 2. Similarly, tires are at the top of checklist 2, but are in the middle and bottom of checklist 1.

If checklists themselves affect inspection behavior, more wiper and light repairs would occur under checklist 1 than checklist 2, while more tire repairs would occur under checklist 2 than checklist 1. As Table 8~~9~~ shows, this is exactly what we find. Furthermore, wipers, lights, and tires are among the least expensive repairs (all with mean prices under $50), again indicating that checklists induce mechanics to address problems with smaller monetary returns. We interpret these results as consistent with checklists acting as a memory aid. In contrast, commission increases led to *fewer* wiper, light, and tire repairs.[36]

For additional evidence on the direct role of checklist use, we compute the share of visits in each mechanic-week that a checklist was used, and test if higher checklist use is associated with more revenue. These results are in Appendix Table A2~~6~~, and show that this is indeed the case: a one-percentage-point increase in checklist use per mechanic-week is associated with a $9.81 (p=0.04) increase in mechanic-week revenue.

F. Is monitoring of checklist use required to promote checklist use?

To check if the treatment effect was solely due to the productivity-enhancing role of checklists, such that monitoring of checklist use was unnecessary, we tracked whether checklist use continued after the treatment periods, when ostensible monitoring was absent. At the end of the treatment periods, extra blank checklists were left at the treatment shops and mechanics were allowed to continue using them, but they never did. That is, even after checklists were shown to be effective at increasing revenue, mechanics chose not to use checklists and outcomes returned to pre-treatment levels, in the absence of ~~this~~ monitoring. The top panel of Figure 2~~1~~ is compelling evidence that this monitoring was central for the observed treatment effects.

G. Is the checklist effect a Hawthorne effect?

As with many experiments, one may wonder if the outcomes are driven by a Hawthorne effect in which the very act of observing workers and collecting data affects their behavior. There are several reasons why this is unlikely to explain our results. First, while mechanics were aware that they were being monitored, monitoring was a central part of the intervention. Insofar as monitoring improves outcomes, it is likely due to the stakes attached to the monitoring rather

[36] Note that alignment and suspension components are at the top of both checklist 1 ("Springs/Vehicle Height," "Inner Tie Rod Ends") and checklist 2 ("Alignment (2 & 4)," "Steering/Suspension"), not providing variation, though both show a modest increase.

than simply being part of an experiment. Second, unlike the well-known Hawthorne experiment, the mechanics were never observed by researchers nor were their activities interfered with (apart from the checklists). That is, while mechanics knew the firm was observing their outcomes, their actions were not being directly observed. Third, the intervention (the checklist) is routine practice in much of the industry, and was introduced to mechanics as something the owner would like to try out, rather than mechanics being told that they were in a control or treatment group. Fourth, we showed in Section V.F that the checklists themselves were driving behavior, which is evidence against a more general improvement in outcomes that might occur under a Hawthorne effect. In summary, we find no ~~evidence~~ indication of a Hawthorne effect.

VI. MORAL HAZARD AND POST-EXPERIMENT BEHAVIOR

Given the apparently large benefits of checklists, it is initially puzzling why mechanics did not conduct thorough inspections (perhaps with checklists) outside of the treatment. One possibility is that the workers were simply unaware of the benefits, as has been found elsewhere (e.g., Bloom et al. 2013, Hanna, Mullainathan, and Schwartzstein 2014). While this explanation may play a role here, moral hazard appears to be a more basic problem. Specifically, mechanics must exert more effort to conduct better inspections, yet receive only some of the added revenue from this effort (the commission). The thorough inspections ~~will~~ only benefit the mechanics if the extra repair revenue multiplied by the commission rate exceeds the extra effort cost.

While we cannot directly observe mechanic effort under checklists, we can indirectly observe some of this effort via the time spent on the job. Commission mechanics are in the shop for 46.7 hours per week and are paid \$758 per week on average (from Table 2), and hence earn \$16.23 per hour. Given convex effort costs, we can use \$16.23 as a lower bound estimate of the hourly rate that would make workers indifferent between working the longer hours to do thorough inspections versus not. That is, mechanics would require at least $16.23 per hour to work the longer hours on their own accord.

Columns (6~~5~~) and (7~~6~~) of Table 6~~7~~ show that mechanics worked an extra 213 minutes and were paid an extra $56.12 per week during the treatment. This represents a marginal hourly rate of $15.84 for the extra time during the treatment. Thus, the actual marginal pay is *below* the lower bound of what is required to induce mechanics to work longer hours. Because more-thorough inspections likely impose other costs as well, such as conducting more tedious repairs and exerting greater effort within any given hour, mechanics may require much more than $16.23 per hour to conduct more-thorough inspections on their own.[37] In summary, the extra pay from checklist inspections may not be worth the extra private effort required of the mechanics.

The firm, however, received $423 more per mechanic-week during the treatment (Appendix Table A3~~2~~, column 3~~2~~), while compensation costs increased by $56 (Table 6~~7~~, column 7~~6~~) and parts costs increased by $50 (not reported elsewhere), leaving $317 in operating profit per mechanic-week from the more-thorough inspections (perhaps modestly less due to taxes). Therefore, if inspection thoroughness were more observable (e.g., with checklists), the firm could pay mechanics for the extra effort and still keep significant added profits. That is, a Pareto improvement may be possible with the right monitoring technology and compensation structure.

Given the large checklist effects, it is also natural to wonder why the firm did not require checklist use all along. We asked the owner and supervisors this question directly, and learned that checklists were not previously adopted for two reasons. First, the mechanics had resisted this

[37] Also note that hourly-pay mechanics are paid 1.5 times their hourly rate for any hours worked over 40, suggesting that the mechanics require a significantly higher pay rate to incentivize the longer hours.

possibility is the past because checklists were perceived as requiring extra effort and interfering with their autonomy. Second, the owner and supervisors were sufficiently unsure about the magnitude of the benefits to not force the issue with mechanics. In summary, employee reluctance to adopt checklists under moral hazard, combined with management uncertainty about their effectiveness, appears to explain their non-adoption.

It is also natural to wonder if checklists were adopted after the experiment. If our conclusion about the benefits of checklists is correct, then a profit-maximizing owner should have adopted a policy of checklist use after the experiment. To assess this, we asked the supervisors about checklist use one year after the experiment. They indicated that the firm had adopted checklists in all of its shops by that time (Fall 2013). We interpret this as compelling evidence that the estimated checklists benefits were real.

VII. CONCLUSIONS

Through a field experiment at an auto-repair chain, we find that worker productivity increased significantly with checklist use. Using within-mechanic variation in commission rates, we find that the checklist effect on firm revenue is equivalent to that of a 10 percent commission increase. Mechanics in the checklist treatment increased revenue by doing more repairs per car and working longer hours. In contrast, mechanics receiving commission increases generated more revenue by substituting from low to high-price repairs, with no increase in number of repairs or work time. Thus, the two approaches to improving worker productivity generated superficially similar outcomes but functioned quite differently.

Our results also suggest that in our setting checklists act primarily as a memory aid rather than as an instructional or learning tool, that the monitoring of worker actions facilitated by the checklists plays an important role, and more basically that the submission of the completed checklists to the firm was required to induce mechanics to use this technology. That is, the checklist effect is driven by a combination of the content of the checklists and the monitoring of their~~its~~ use, and the mere provision of checklists was insufficient to change worker behavior. This was likely because workers lacked sufficient incentives to use checklist when their use was not monitored. We also calculate that a modest transfer of profits from the firm to mechanics may be sufficient to compensate mechanics for their extra effort under checklist use, making possible a sizable Pareto improvement. This type of moral hazard might help to explain the mixed findings on checklist efficacy in the medical literature, and supports the assertion in Bosk, Dixon-Woods, Goeschel, and Pronovost (2009), Gawande (2014), and others that checklist interventions require supporting activities in order to be effective.

A number of questions about the checklist technology remain, which are ripe for further investigation. First, we find important checklist effects during four-week treatment periods, but would complacency set in such that the effects weaken over time? Second, given that mechanics used checklists on only 30 percent of cars during the treatment, would better compliance generate larger gains? Third, our primary supporting activity was the monitoring of checklist use, but medical checklist interventions typically involve many more activities, such as worker training and an emphasis on communication: What is the optimal bundle of activities?

Checklists represent an inexpensive and potentially powerful technology to improve worker productivity. In this study, we have provided the first experimental evidence on the effects of this technology on worker behavior, providing support for the large benefits found in some of the previous medical literature, but also identifying some conditions that may be necessary for obtaining these benefits.

REFERENCES

~~Akerlof, George (1982) "Labor Contracts as Partial Gift Exchange," Quarterly Journal of Economics, 97(4), 543-569.~~

Baker, George (1992) "Incentive Contracts and Performance Measurement," Journal of Political Economy, 100(3), 598–614.

Bernheim, B. Douglas, Lorenzo Forni, Jagadeesh Gokhale, and Laurence J. Kotlikoff (2003) "The Mismatch Between Life Insurance Holdings and Financial Vulnerabilities: Evidence from the Health and Retirement Study," American Economic Review, 93(1), 354-365.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan (2004) "~~"~~How Much Should We Trust Differences-in-Differences Estimates?," ~~"~~ Quarterly Journal of Economics, 119(1), 249-275.

Bertrand, Marianne, and Antoinette Schoar (2003) "Managing with Style: The Effect of Managers on Firm Policies," Quarterly Journal of Economics, 118(4), 1169-1208.

Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts (2013) "Does Management Matter? Evidence From India," Quarterly Journal of Economics, 128(1), 1-51.

Boorman, Daniel (2001) "Today's Electronic Checklists Reduce Likelihood of Crew Errors and Help Prevent Mishaps," International Civil Aviation Organization Journal, 1.

Bosk, Charles L., Mary Dixon-Woods, Christine A. Goeschel, and Peter J. Pronovost (2009) "Reality Check for Checklists," The Lancet, 374(9688), 444-445.

Camerer, A. Colin, and Dan Lovallo (1999) "Overconfidence and Access Entry: An Experimental Approach," American Economic Review, 89(1), 306-318.

Camerer, A. Colin, Jonah B. Gelbach, and Douglas L. Miller (2008) "Bootstrap-Based Improvements for Inference with Clustered Errors," Review of Economics and Statistics, 90(3), 414-427.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller (2011) "Robust Inference With Multiway Clustering," Journal of Business & Economic Statistics, American Statistical Association, ~~American Statistical Association,~~ 29(2), 238-249.

Ericson, Keith M. (2011) "Forgetting We Forget: Overconfidence and Memory," Journal of the European Economic Association, 9(1), 43-60.

~~Foster, A., and M. Rosenzweig (1994) "A Test for Moral Hazard in the Labor Market: Contractual Arrangements, Effort, and Health," Review of Economics and Statistics,~~

~~76(2), 213-227.~~

Freeman, R.~~-~~B., and M.M. Kleiner~~, M. M.~~ (2005) "The Last American Shoe Manufacturers: Decreasing Productivity and Increasing Profits in the Shift from Piece Rates to Continuous Flow Production," Industrial Relations: A Journal of Economy and Society, 44(2), 307-330.

Gawande, Atul (2009) *The Checklist Manifesto: How to Get Things Right*, Metropolitan Books.

Gawande, Atul (2014) "When Checklists Work and When They Don't," Incidental Economist, http://theincidentaleconomist.com/wordpress/when-checklists-work-and-when-they-dont/ [last accessed 10/16/14].

Hanna, Rema, Sendhil Mullainathan, and Joshua Schwartzstein (2014) "Learning Through Noticing: Theory and Experimental Evidence in Farming," Quarterly Journal of Economics, 129(3), 1311-1353.

Haynes, Alex B., et al. (2009) "A Surgical Safety Checklist to Reduce Morbidity and Mortality in a Global Population," New England Journal of Medicine, 360(5), 491-499.

Holmstrom, Bengt, and Paul Milgrom (1991) "Multitask Principal-Agent Analysis: Incentive Contracts, Asset Ownership, and Job Design," Journal of Law, Economics, and Organization, 7, 24-52.

Hubbard, Thomas N. (1998) "An Empirical Examination of Moral Hazard in the Vehicle Inspection Market," Rand Journal of Economics, 29(2), 406-426.

Jackson, C. Kirabo, Jonah Rockoff, and Douglas Staiger (2014) "Teachers'~~'~~ Effects and Teacher-Related Policies," Annual Review of Economics, 6.

Jackson, C. Kirabo, and Henry S. Schneider (2011) "Do Social Connections Reduce Moral Hazard? Evidence From the New York City Taxi Industry," American Economic Journal: Applied Economics, (3)3, 244-267.

Johnson, Justin P., Henry S. Schneider, and Michael Waldman (2014) "The Role and Growth of New-Car Leasing: Theory and Evidence," Journal of Law and Economics, 57(3).

Larkin, Ian (forthcoming) "The Cost of High-Powered Incentives: Employee Gaming in Enterprise Software Sales," Journal of Labor Economics.

Lazear, E.~~-~~P. (2000) "Performance Pay and Productivity," American Economic Review, 90(5), 1346-1361.

Lemieux, Thomas, W. Bentley MacLeod, and Daniel Parent (2009) "Performance Pay and Wage Inequality," Quarterly Journal of Economics, 124(1), 1-49.

~~Lindenauer, Peter K. et al. (2007) "Public Reporting and Pay for Performance in Hospital~~

~~Quality Improvement," New England Journal of Medicine, 356(5), 486-496.~~

List, John and Imran Rasul (2011) "Field Experiments in Labor Economics," *Handbook of Labor Economics*, volume 4a, Elsevier B.V., 103-228.

Lucas, Robert (1978) "On the Size Distribution of Business Firms," Bell Journal of Economics, 9(2), 508–523.

Mullen, Kathleen J., Richard G. Frank, and Meredith B. Rosenthal (2010) "Can You Get What You Pay F~~f~~or? Pay-for-Performance and the Quality of Healthcare Providers," Rand Journal of Economics, 41(1), 64-91.

Pickering, S.~~-~~P., E.R. Robertson, ~~E. R.,~~ D. Griffin~~, D.~~, M. Hadi~~, M.~~, L.J. Morgan~~, L. J.~~, K.C. Catchpole~~, K. C.~~, S. New~~, S.~~, G. Collins~~, G.,~~ and P. McCulloch~~, P.~~ (2013) "Compliance and Use of the World Health Organization Checklist in UK Operating Theatres," British Journal of Surgery, 100,~~:~~ 1664–1670.~~ doi: 10.1002/bjs.9305~~

Pronovost, Peter, Dale Needham, Sean Berenholtz, et al. (2006) "An Intervention to Decrease Catheter-Related Bloodstream Infections in the ICU," New England Journal of Medicine, 355(26), 2725-2732.

Rosenthal, Meredith B., and Robert. G. Frank (2006) "What Is the Empirical Basis for Paying for Quality in Health Care?," Medical Care Research and Review, 63(2),135–57.

Schneider, Henry S. (2010) "Moral Hazard in Leasing Contracts: Evidence from the New York City Taxi Industry," Journal of Law and Economics, 53(4), 783-805.

Schneider, Henry S. (2012) "Agency Problems and Reputation in Expert Services: Evidence from Auto Repair," Journal of Industrial Economics, 60(3), 406-433.

Shearer, B. (2004) "Piece Rates, Fixed Wages and Incentives: Evidence from a Field Experiment," Review of Economic Studies, 71, 514–534.

Urbach, David R., Anand Govindarajan, Refik Saskin, Andrew S. Wilton, and Nancy N. Baxter (2014) "Introduction of Surgical Safety Checklists in Ontario, Canada," New England Journal of Medicine, 370(11), 1029-1038.

Table 1: Repairs types and characteristics

	Visits with repair (percent)	Mean revenue ($)	Standard deviation revenue ($)
A/C repair, inspection	2.5	232	271
Alignment, suspension	12.2	285	325
Battery repair, service, inspection	3.8	95	58
Belts, pulley, tensioner	3.2	154	148
Brake repair, fluid, flush, inspection	16.3	311	248
Coolant fluid, flush	3.9	74	65
Engine cleaner, flush	3.0	20	21
Exhaust repair, inspection	9.3	216	219
Filters (air, cabin, fuel, PCV)	4.8	38	33
Fuel cleaner, service	1.4	124	28
Lights	2.9	34	49
Oil change	54.2	25	15
Radiator, hoses, fan, thermostat, water pump	2.4	257	203
Spark plugs, wires, coil, rotor, distributor	2.6	237	159
Tire rotation, repair, balance	9.2	46	112
Transmission fluid, service	1.6	122	66
Windshield wipers	2.2	24	23

Notes: Repair types that occur in at least one percent of visits are included. N=155,049 observations.

Table 2: Summary statistics on mechanics

	All mechanics	Commission mechanics	Non-commission mechanics
Commission rate (percent)	10.8	17.7	0.1
	[8.7]	[1.5]	[0.4]
Days per week	4.89	4.93	4.76
	[1.17]	[1.13]	[1.27]
Visits per week	19.1	18.6	20.5
	[10.4]	[8.5]	[14.9]
Revenue per week	3,700	4,027	2,669
	[2,410]	[2,265]	[2,560]
Payroll hours per week	45.5	46.7	42.6
	[14.4]	[13.3]	[16.3]
Pay per week ($)	705	758	585
	[415]	[401]	[422]
Workdays at firm since 1998	703	839	492
	[755]	[770]	[689]
N mechanics	84	51	33
N visits	146,852	108,892	37,960

Notes: "All mechanics" includes mechanics with pay structures of commission, hourly, and flat rate. Twenty-five managers are excluded from the table because they only occasionally conducted repairs, and including them would distort statistics in the "All mechanics" column. Visits for which mechanic information is missing are excluded from the table (approximately five percent of visits). "Workdays at firm since 1998" is the number of days between 1998 and 2013 that the mechanic worked on at least one car at the firm. Standard errors are reported in brackets.

Table 3: Characteristics of treatment and control shops

	Treatment shops	Control shops
N visits per shop-day	11.8	12.0
N mechanics per shop-day	3.17	2.68
Workdays at firm since 1998	729	682
Mileage per car	94812	107597
Proportion of v~~V~~isits with oil change	0.568	0.531
Revenue per visit	191	190

Notes: Treatment and control shops are those in which checklist were and were not used, respectively. “Workdays at firm since 1998” is the number of days between 1998 and 2013 that the mechanic worked on at least one car at the firm.

Table 4: Description of commission-rate changes

Mechanic ID	Commission change	Date of change
32	17% to 18%	11/29/2009
203	16% to 17%	8/14/2011
206	2% to 3%	8/26/2011
302	16% to 17%	8/7/2011
302	17% to 18%	10/21/2012
303	16% to 17%	2/14/2010
402	16% to 17%	11/29/2009
412	17% to 18%	1/18/2009
601	17% to 18%	2/13/2013
602	18% to 19%	7/18/2010
704	16% to 17%	9/4/2011
920	17.5% to 18%	10/8/2008
920	18% to 19%	5/11/2009
920	19% to 20%	3/11/2012
1201	17% to 18%	8/30/2009
1205	0% to 3%	8/19/2012

Notes: The two mechanics with commission rates between 0 and 3 percent are paid primarily on an hourly rate, which is supplemented with the indicated commission rate. The remaining mechanics are paid on commission, with a guaranteed minimum base pay in case of very low commissions in that pay period.

Table 5: Estimated models of revenue

	(1) Revenue per visit	(2) Revenue per visit	(3) Log revenue per visit	(4) Log revenue per visit
Treatment	41.60***	38.02***	0.216***	0.190***
	[13.44]	[11.52]	[0.046]	[0.039]
Commission	28.82***	25.19***	0.129***	0.112***
	[6.95]	[6.15]	[0.032]	[0.027]
Treatment/commission	1.44**	1.51**	1.68***	1.70***
	[0.58]	[0.58]	[0.58]	[0.58]
Car and customer controls		X		X
Observations	154,722	154,722	151,810	151,810
R-squared	0.11	0.16	0.20	0.26

Notes: "Treatment" is an indicator for visit during the checklist treatment. "Commission" is the commission rate of the mechanic in units of percent. "Treatment/commission" is the ratio of the estimates. The models are estimated by OLS. Heteroskedasticity-robust standard errors clustered at the shop-year level are reported in brackets. All models include mechanic, shop, and year-week fixed effects, day-of-week dummies, and an intercept term. "Car and customer controls" includes car age, mileage, and make, and indicators for female, business customer, and having an account with the firm. ** and *** indicate statistical significance at the five and one percent levels, respectively.

Formatted: Space After: 8 pt, Don't add space between paragraphs of the same style, Line spacing: Multiple 1.08 li

Formatted Table

Formatted: Don't add space between paragraphs of the same style

Formatted: Indent: First line: 0 pt, Space After: 8 pt, Don't add space between paragraphs of the same style, Line spacing: Multiple 1.08 li

Formatted: Don't add space between paragraphs of the same style

Formatted: Don't add space between paragraphs of the same style

Formatted: Indent: First line: 0 pt, Space After: 8 pt, Don't add space between paragraphs of the same style, Line spacing: Multiple 1.08 li

Formatted: Don't add space between paragraphs of the same style

Formatted: Don't add space between paragraphs of the same style

Formatted: Indent: First line: 0 pt, Space After: 8 pt, Don't add space between paragraphs of the same style, Line spacing: Multiple 1.08 li

Formatted: Don't add space between paragraphs of the same style

Formatted: Don't add space between paragraphs of the same style

Formatted: Indent: First line: 0 pt, Space After: 8 pt, Don't add space between paragraphs of the same style, Line spacing: Multiple 1.08 li

Formatted: Don't add space between paragraphs of the same style

Formatted: Don't add space between paragraphs of the same style

Formatted: Don't add space between paragraphs of the same style

Formatted: Indent: First line: 0 pt, Space After: 8 pt, Don't add space between paragraphs of the same style, Line spacing: Multiple 1.08 li

Formatted: Indent: First line: 0 pt, Space After: 8 pt, Don't add space between paragraphs of the same style, Line spacing: Multiple 1.08 li

Formatted: Don't add space between paragraphs of the same style

Formatted: Don't add space between paragraphs of the same style

~~Table 6: Interaction of the checklist treatment and rate of checklist use~~

	~~(1)~~	~~(2)~~
	~~Revenue per mechanic-week~~	~~Log revenue per mechanic-week~~
~~Treatment~~	~~115.73~~	~~0.199**~~
	~~[259.85]~~	~~[0.083]~~
~~Treatment x fraction checklist use by mechanic-week~~	~~980.67**~~	~~0.490*~~
	~~[464.21]~~	~~[0.284]~~
~~Commission~~	~~414.21***~~	~~0.170***~~
	~~[135.02]~~	~~[0.049]~~
~~Observations~~	~~9,358~~	~~9,315~~
~~R-squared~~	~~0.64~~	~~0.753~~

~~Notes: The specifications in columns (1) and (2) are the same as in columns (2) and (4) of Table 5, respectively, but aggregated to the mechanic-week level and with the addition of the explanatory variables, "Treatment x fraction checklist use by mechanic-week" and a stand-alone term for fraction checklist use by mechanic-week. Heteroskedasticity-robust standard errors clustered at the shop-year level are reported in brackets. *, **, and *** indicate statistical significance at the ten, five and one percent levels, respectively.~~

Table 6~~7~~: Estimated models for number of visits, repairs, minutes, and pay

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Log of profit per visit	Indicator for return within six months	Log of recommended revenue per visit	Number of visits per mechanic-week	Number of repairs per visit	Minutes per mechanic-week	Non-salary pay per mechanic-week ($)
Treatment	0.257***	-0.004	0.407**	-0.634	0.119***	212.56**	56.12*
	[0.063]	[0.011]	[0.148]	[0.836]	[0.031]	[101.43]	[30.03]
Commission	0.182***	-0.005	-0.056	0.489	-0.010	-50.33	-0.53
	[0.042]	[0.009]	[0.147]	[0.358]	[0.013]	[34.30]	[15.03]
Mean of dep. var.		0.333		16.54	1.440	2,868	556
Observations	148,539	138,282	2,835	9,358	149,853	2,489	2,489
R-squared	0.29	0.08	0.25	0.73	0.04	0.51	0.72

Formatted Table

Notes: "Treatment" is an indicator for a visit during the treatment. "Commission" is the commission rate of the mechanic in units of percent. "Log of recommended revenue per visit" is the estimate quoted to the customer, and some of these repairs are sometimes not consented to by the customer. The models are estimated by OLS. Heteroskedasticity-robust standard errors clustered at the shop-year level are reported in brackets. All models include controls for shop, mechanics, customer characteristics, car characteristics, and an intercept term. *, **, and *** indicate statistical significance at the ten, five, and one percent levels, respectively.

Table 7~~8~~: Rank-order correlation tests of repair-type revenue and intervention effects

	Mean revenue ($)	Treatment effect	Commission effect	Rank of mean revenue	Rank of treatment effect	Rank of commission effect
Brake repair, fluid, flush, inspection	311	0.0288**	0.0211***	1	1	1
Alignment, suspension	285	0.0112	0.0093	2	8	2
Radiator, hoses, fan, thermostat, water pump	257	0.0005	0.0033*	3	15	8
Spark plugs, wires, coil, rotor, distributor	237	-0.0037	0.0056*	4	16	3
A/C repair, inspection	232	0.0005	0.0054**	5	14	4
Exhaust repair, inspection	216	0.0120	0.0006	6	6	11
Belts, pulley, tensioner	154	0.0077	0.0022	7	10	9
Fuel cleaner, service	124	0.0131***	0.0039**	8	4	6
Transmission fluid, service	122	0.0045	0.0020	9	12	10
Battery repair, service, inspection	95	0.0016	0.0051*	10	13	5
Coolant fluid, flush	74	0.0120***	0.0038*	11	6	7
Tire rotation, repair, balance	46	0.0089	-0.0052	12	9	14
Filters (air, cabin, fuel, PCV)	38	0.0075	-0.0034	13	11	12
Lights	34	0.0177**	-0.0073***	14	3	15
Oil change	25	-0.0420**	-0.0481***	15	17	17
Windshield wipers	24	0.0130***	-0.0037**	16	5	13
Engine cleaner, flush	20	0.0198	-0.0132*	17	2	16
Rank correlation between mean revenue and treatment effect:						-0.17
Rank correlation between mean revenue and commission effect:						0.85

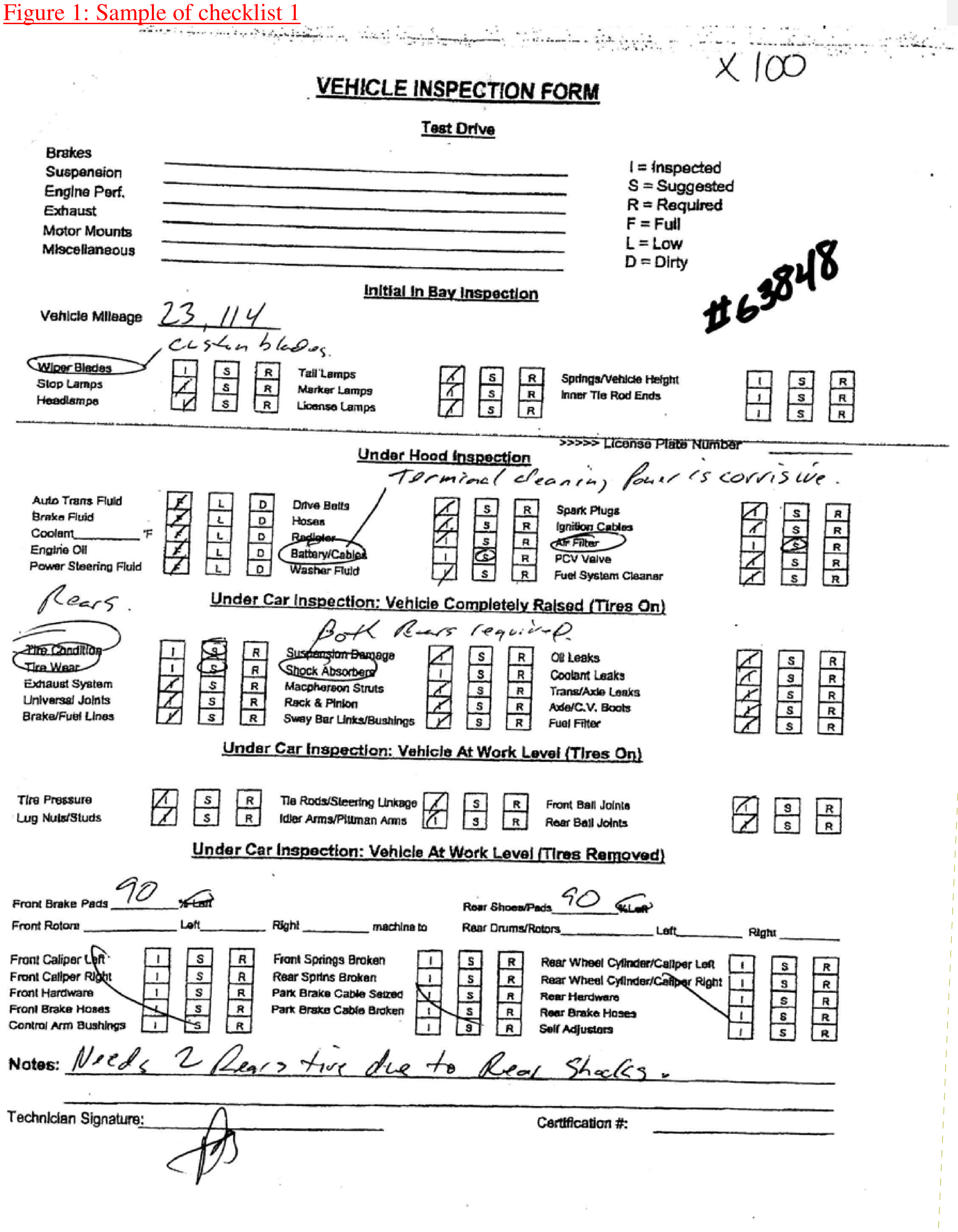
Notes: "Treatment effect" is the estimated effects of the treatment on the probability of repair for each repair type. "Commission effect" is the estimated effects of a one percentage point increase in the commission rate for that repair type. Repairs types that are present in at least one percent of visits are reported. Models are estimated for each repair type individually, and include shop, mechanic, time, customer, and car controls, and an intercept term. Results from a rank-order correlation test are reported at the bottom of the table. *, **, and *** indicate statistical significance at the ten, five, and one percent levels.

Table ~~8~~9: Effect of checklist order on probability of individual repair types

	(1)	(2)	(3)	(4)
	Lights	Windshield wipers	Tires	Alignment/ suspension
	Top of checklist 1, bottom of checklist 2	Top of checklist 1, bottom of checklist 2	Bottom of checklist 1, top of checklist 2	Top of checklist 1, top of checklist 2
Checklist 1 (treatment)	0.0527***	0.0293***	-0.0133**	0.0121
	[0.0057]	[0.0048]	[0.0060]	[0.0127]
Checklist 2 (treatment)	0.0021	0.0048	0.0150	0.0136
	[0.0107]	[0.0034]	[0.0101]	[0.0120]
Commission	-0.0077***	-0.0031*	-0.0051	0.0093
	[0.0022]	[0.0017]	[0.0031]	[0.0059]
Checklist 1 – checklist 2	0.0506***	0.0245***	-0.0283***	-0.0014
	[0.0127]	[0.0040]	[0.0112]	[0.0145]
Checklist 1 – commission	0.0604***	0.0325***	-0.0082	0.0028
	[0.0062]	[0.0073]	[0.0073]	[0.0148]
Checklist 2 – commission	0.0098	0.0079	0.0201*	0.0043
	[0.0111]	[0.0053]	[0.0099]	[0.0181]
Observations	154,722	154,722	154,722	154,722

Notes: "Checklist 1 (treatment)" and "Checklist 2 (treatment)" are indicators for visits during the treatment with checklist 1 and checklist 2, respectively. "Commission" is the commission rate of the mechanic in units of percent. The models are estimated by OLS and the unit of observation is the visit. The statistics in the bottom panel are the differences in the indicated estimates. Heteroskedasticity-robust standard errors clustered at the shop-year level are reported in brackets. All models include controls for shop, mechanics, time, customer, car, and an intercept term. *, **, and *** indicate statistical significance at the ten, five, and one percent levels, respectively.

Figure 1: Sample of checklist 1



X 100

VEHICLE INSPECTION FORM

Test Drive

Brakes ______
Suspension ______
Engine Perf. ______
Exhaust ______
Motor Mounts ______
Miscellaneous ______

I = Inspected
S = Suggested
R = Required
F = Full
L = Low
D = Dirty

#63848

Initial In Bay Inspection

Vehicle Mileage 23,114

custom blades.

Wiper Blades	I	S	R	Tail Lamps	I	S	R	Springs/Vehicle Height	I	S	R
Stop Lamps	I	S	R	Marker Lamps	I	S	R	Inner Tie Rod Ends	I	S	R
Headlamps	I	S	R	License Lamps	I	S	R		I	S	R

>>>>> License Plate Number ______

Under Hood Inspection

Terminal cleaning four is corrosive.

Auto Trans Fluid	F	L	D	Drive Belts	I	S	R	Spark Plugs	I	S	R
Brake Fluid	F	L	D	Hoses	I	S	R	Ignition Cables	I	S	R
Coolant ____ °F	F	L	D	Radiator	I	S	R	Air Filter	I	S	R
Engine Oil	F	L	D	Battery/Cables	I	S	R	PCV Valve	I	S	R
Power Steering Fluid	F	L	D	Washer Fluid	I	S	R	Fuel System Cleaner	I	S	R

Rears.

Under Car Inspection: Vehicle Completely Raised (Tires On)

Both Rears required.

Tire Condition	I	S	R	Suspension Damage	I	S	R	Oil Leaks	I	S	R
Tire Wear	I	S	R	Shock Absorbers	I	S	R	Coolant Leaks	I	S	R
Exhaust System	I	S	R	Macpherson Struts	I	S	R	Trans/Axle Leaks	I	S	R
Universal Joints	I	S	R	Rack & Pinion	I	S	R	Axle/C.V. Boots	I	S	R
Brake/Fuel Lines	I	S	R	Sway Bar Links/Bushings	I	S	R	Fuel Filter	I	S	R

Under Car Inspection: Vehicle At Work Level (Tires On)

Tire Pressure	I	S	R	Tie Rods/Steering Linkage	I	S	R	Front Ball Joints	I	S	R
Lug Nuts/Studs	I	S	R	Idler Arms/Pitman Arms	I	S	R	Rear Ball Joints	I	S	R

Under Car Inspection: Vehicle At Work Level (Tires Removed)

Front Brake Pads 90 %Left   Rear Shoes/Pads 90 %Left

Front Rotors ______ Left ______ Right ______ machine to   Rear Drums/Rotors ______ Left ______ Right ______

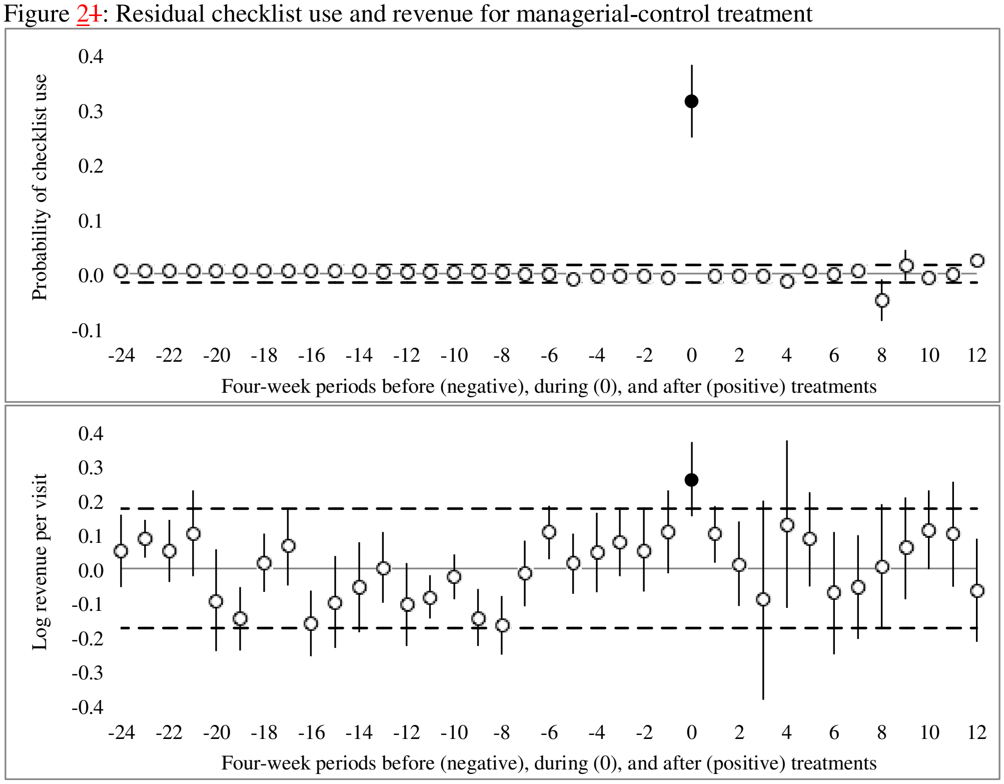
Front Caliper Left	I	S	R	Front Springs Broken	I	S	R	Rear Wheel Cylinder/Caliper Left	I	S	R
Front Caliper Right	I	S	R	Rear Sprins Broken	I	S	R	Rear Wheel Cylinder/Caliper Right	I	S	R
Front Hardware	I	S	R	Park Brake Cable Seized	I	S	R	Rear Hardware	I	S	R
Front Brake Hoses	I	S	R	Park Brake Cable Broken	I	S	R	Rear Brake Hoses	I	S	R
Control Arm Bushings	I	S	R		I	S	R	Self Adjusters	I	S	R

Notes: Needs 2 Rears tire due to Real Shocks.

Technician Signature: ______   Certification #: ______

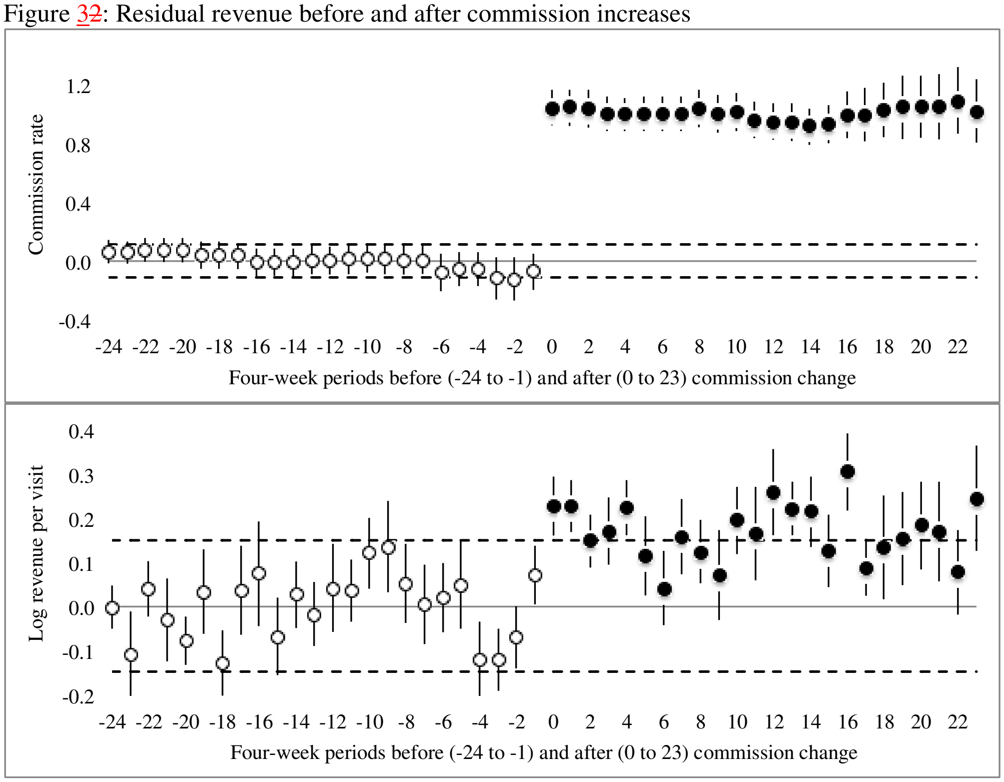
Formatted: Font: (Default) Times New Roman, 12 pt

Figure 2~~1~~: Residual checklist use and revenue for managerial-control treatment



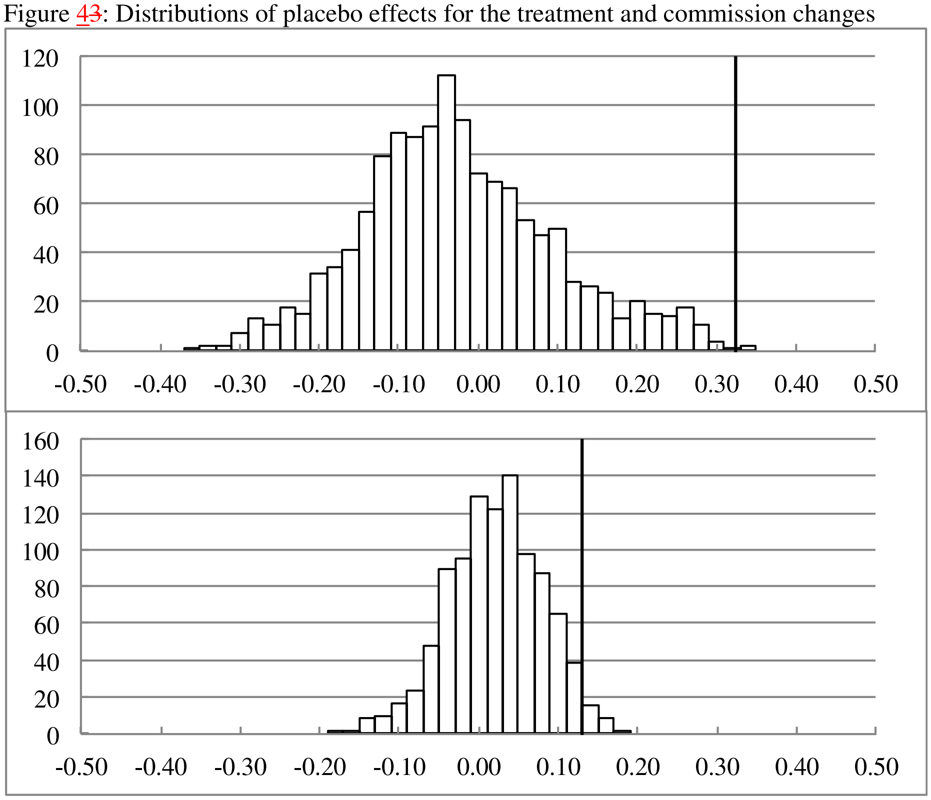
Notes: The circles represent the means of the regression residuals over the 24 four-week periods before the treatment period, the four-week periods of the treatment period (indicated by the solid circle), and 12 four-week periods after the treatment periods (when the sample period ends). The unit of observation of the regression is the visit. The dependent variable in the top panel is an indicator for whether a checklist was used in that visit, and in the bottom panel is log revenue for that visit. Standard error bars are included, and the ~~95 percent confidence~~95-percent confidence interval on the mean four-week residuals is indicated by the horizontal dashed lines.

Figure 3~~2~~: Residual revenue before and after commission increases



Notes: The circles represent the means of the regression residuals over 24 four-week periods before (~~light~~open circle) and after (~~dark~~solid circle) a commission increase. The dependent variable in the top panel is the mechanic commission rate, and in the bottom panel is log revenue per visit. Standard error bars are included, and the ~~95 percent confidence~~95-percent confidence interval on the mean four-week residuals is indicated by the horizontal dashed lines.

Figure 4~~3~~: Distributions of placebo effects for the treatment and commission changes



Notes: The top panel shows estimates from the 1318 placebo managerial-control treatments generated by offsetting the actual treatment periods by an increasing number of days for all days in the sample prior to the actual treatment periods. The bottom panel shows estimates from the 1~~,~~600 placebo treatments for the 16 commission increases. The vertical lines indicate the estimated effects for the true treatment period (top panel) and true commission rate changes (bottom panel).

~~Figure 4: Sample of checklist 1~~

X 100

VEHICLE INSPECTION FORM

Test Drive

Brakes ____
Suspension ____
Engine Perf. ____
Exhaust ____
Motor Mounts ____
Miscellaneous ____

I = Inspected
S = Suggested
R = Required
F = Full
L = Low
D = Dirty

#63848

Initial In Bay Inspection

Vehicle Mileage 23,114

custom blades.

Wiper Blades	I	S	R	Tail Lamps	I	S	R	Springs/Vehicle Height	I	S	R
Stop Lamps	I	S	R	Marker Lamps	I	S	R	Inner Tie Rod Ends	I	S	R
Headlamps	I	S	R	License Lamps	I	S	R		I	S	R

>>>>> License Plate Number ____

Under Hood Inspection

Terminal cleaning, power is corrosive.

Auto Trans Fluid	F	L	D	Drive Belts	I	S	R	Spark Plugs	I	S	R
Brake Fluid	F	L	D	Hoses	I	S	R	Ignition Cables	I	S	R
Coolant ____ °F	F	L	D	~~Radiator~~	I	S	R	~~Air Filter~~	I	S	R
Engine Oil	F	L	D	Battery/Cables	I	S	R	PCV Valve	I	S	R
Power Steering Fluid	F	L	D	Washer Fluid	I	S	R	Fuel System Cleaner	I	S	R

Rears.

Under Car Inspection: Vehicle Completely Raised (Tires On)

Both Rears required.

~~Tire Condition~~	I	S	R	~~Suspension Damage~~	I	S	R	Oil Leaks	I	S	R
~~Tire Wear~~	I	S	R	Shock Absorbers	I	S	R	Coolant Leaks	I	S	R
Exhaust System	I	S	R	Macpherson Struts	I	S	R	Trans/Axle Leaks	I	S	R
Universal Joints	I	S	R	Rack & Pinion	I	S	R	Axle/C.V. Boots	I	S	R
Brake/Fuel Lines	I	S	R	Sway Bar Links/Bushings	I	S	R	Fuel Filter	I	S	R

Under Car Inspection: Vehicle At Work Level (Tires On)

Tire Pressure	I	S	R	Tie Rods/Steering Linkage	I	S	R	Front Ball Joints	I	S	R
Lug Nuts/Studs	I	S	R	Idler Arms/Pitman Arms	I	S	R	Rear Ball Joints	I	S	R

Under Car Inspection: Vehicle At Work Level (Tires Removed)

Front Brake Pads 90 %Left Rear Shoes/Pads 90 %Left

Front Rotors ____ Left ____ Right ____ machine to Rear Drums/Rotors ____ Left ____ Right ____

Front Caliper Left	I	S	R	Front Springs Broken	I	S	R	Rear Wheel Cylinder/Caliper Left	I	S	R
Front Caliper Right	I	S	R	Rear Sprins Broken	I	S	R	Rear Wheel Cylinder/Caliper Right	I	S	R
Front Hardware	I	S	R	Park Brake Cable Seized	I	S	R	Rear Hardware	I	S	R
Front Brake Hoses	I	S	R	Park Brake Cable Broken	I	S	R	Rear Brake Hoses	I	S	R
Control Arm Bushings	I	S	R		I	S	R	Self Adjusters	I	S	R

Notes: Needs 2 Rears tire due to Rear Shocks.

Technician Signature: ____ Certification #: ____

Formatted: Font: (Default) Times New Roman, 12 pt

~~Figure 5: Sample of checklist 2~~

Formatted: Font: (Default) Times New Roman, 12 pt

99 SATURN SL Series 7/20/12

INV# 55882

Inspection Item Required	Condition			Explain/Action
TIRES				
1. Tires - Front	OK	NA	Needs Attention	
2. Tires - Rear	OK	NA	Needs Attention	
3. Alignment (2 & 4)	OK	NA	Needs Attention	
BRAKES AND SUSPENSION				
4. Steering/Suspension	OK	NA	Needs Attention	BOTH OUTER TIE RODS / GREASE BOOTS TORN
5. R&P/Pump/Hoses/Fluid	OK	NA	Needs Attention	
6. Front Brakes	80	% Left	Needs Attention	
7. Rear Brakes	15	% Left	Needs Attention	SHOES / HARDWARE
8. Rotors and Drums	OK	NA	Needs Attention	MACH. NB
9. Hydraulics & Brake Fluid	OK	NA	Needs Attention	
10. Axle Shafts/Boots/U Joints	OK	NA	Needs Attention	
11. Front Struts/Shocks	OK	NA	Needs Attention	
12. Rear Struts/Shocks	OK	NA	Needs Attention	
EXHAUST AND ENGINE				
13. Exhaust System	OK	NA	Needs Attention	
14. Engine/Transmission Leaks	OK	NA	Needs Attention	
15. Differential Fluid Leaks	OK	NA	Needs Attention	
MAINTENANCE AND MISC.				
16. Check Engine Light	OK	NA	Needs Attention	
17. Drive Belts & Accessories	OK	NA	Needs Attention	CRACKS
18. Transmission Fluid	OK	NA	Needs Attention	
19. Cooling System Fluid	OK	NA	Needs Attention	
20. Radiator/Hoses	OK	NA	Needs Attention	
21. Heater Hoses	OK	NA	Needs Attention	
22. Battery & Terminales	OK	NA	Needs Attention	
23. Battery Cables	OK	NA	Needs Attention	
24. Tune-Up Components	OK	NA	Needs Attention	
25. Filters (Cabin/Fuel/Air/PCV)	OK	NA	Needs Attention	
26. Engine Oil (Condition and Level)	OK	NA	Needs Attention	
27. Lights (Brake, tail, turn, etc.)	OK	NA	Needs Attention	
28. Headlights/Adjustment	OK	NA	Needs Attention	
29. Wiper Blades (Front and Rear)	OK	NA	Needs Attention	
30. Factory Scheduled Maintenance Mileage: ______			Due:	3K 7.5K 15K 30K 60K 90K

Miscellaneous Comments:

APPENDIX

A. The commission effect reflects other changes coinciding with the commission increase

The event-study analysis in Section II shows that the timing of the revenue increase coincided with the commission increase and was not driven by pre-existing trend differences. However, these results could still be biased if the shop had better overall outcomes coinciding with the commission increases or if the commission increases are a response to improved overall shop outcomes rather than the reverse. Because shops have at least two mechanics working per day, we can test these possibilities directly by using the other mechanics at the same shop at the same time as controls. We conduct this test by including shop-week fixed effects into the revenue per visit regression models (from ~~columns 2 and 4 of~~ Table 5). If shop-specific time shocks coinciding with commission increases explain the effect, then including shop-week fixed effects will eliminate the commission effect. The estimate of the commission effect in the level-of-revenue model is 22.05 ($p<0.01$), which is similar to the estimate of 25.19 ($p<0.01$) from column (2) of Table 5; while the estimate in the log-~~of~~-revenue model is 0.119 ($p<0.01$), which is similar to the estimate of 0.112 ($p<0.01$) from column (4) of Table 5.

B. The commission effect reflects business stealing from other mechanics at the same shop

Given that commission increases are not random, one might worry that when mechanics receive commission increases they may have greater control over the cars or repairs they work on ~~or repairs they conduct~~, ~~taking away~~ displacing work from other mechanics at the shop. If a business-stealing effect is driving the commission effect, then including shop-week fixed effects will *increase* the commission effect. This is because other mechanics at the same shop will have less revenue as some of their usual repairs are transferred to the CI mechanic ~~with the~~

~~commission increase~~. The results mentioned in the subsection above, which include shop-week fixed effects, show no change in the commission estimate.

C. The treatment and commission effects reflect shifting across time or mechanics

To ensure that temporal shifting of effort within mechanics does not explain the ~~treatment and commission~~ observed effects (i.e., doing more repairs per car but ~~working on~~ attending to fewer cars overall), we aggregate revenue across visits for a given worker to the week level and estimate the revenue model at the mechanic-week level. ~~The r~~Results ~~for the level-of-revenue and log-of-revenue models~~ are in columns (3) and (4) of Table A3~~2~~, and are qualitatively similar to the visit-level results. One may also wonder whether increases for one mechanic are offset by decreases for other mechanics or decreases in the total number of cars repaired at the shop. We test this directly by aggregating the revenue across all mechanics to the shop-week level. Because commission level is a mechanic-defined variable, commission level is not included. If shifting across workers or servicing fewer cars ~~overall~~ explains the results, then estimating the treatment effect on revenue aggregated to the shop would cause the effect to disappear. The results ~~for the level-of-revenue and log-of-revenue models~~ are in columns (5) and (6) of Table A3~~2~~. The estimated effects are positive and significant, indicating that ~~a~~ effort shifting ~~of effort~~ across cars or mechanics cannot explain the effect.[38]

~~D. The sample is too small for valid statistical inference~~

~~Because the number of treated mechanics and treated shops in our sample is small, it is important to show that our estimates are robust to small-sample inference tests. We do this several ways. First, we cluster the standard errors throughout at the shop-year level to account for the possibility that outcomes within shops are correlated over time. However, as pointed out in Bertrand, Duflo, and Mullainathan (2004), clustering the standard errors can lead to an over-~~

[38] While the shop-week effect is modestly smaller than the visit-level effect, the p-value of the difference is 0.28.

~~rejection of a true null hypothesis. To address this possibility we follow their suggestion and estimate the effects of 1,318 placebo treatments corresponding to shifting the timing of the actual treatment periods by each of an increasing number of days prior to the actual treatment dates, with no overlap with the true treatment dates.[39] Similarly, we estimate the effect of 1,700 placebo commission increases, where for each we chose 17 mechanics at random, assign one percentage point placebo commission increases at a random time for each mechanic, and estimate the model.~~

~~The distribution of the placebo treatments and placebo commission increases are in Figure 3. For both variables the actual estimated effect is larger than 97.5 percent of the placebo replications. As a further check on the robustness of our inferences to small samples we follow Camerer, Gelbach, and Miller (2008) who argue that a more conservative test is to use a wild bootstrap clustered t-statistic. For both variables of interest, the estimated t-statistic is larger than 97.5 percent of the wild bootstrap clustered t-statistics.~~

~~E.~~D. The managerial-control and commission effects are driven by a few mechanics or shops

When the number of treated units is small, one may worry that the results are due to outlier mechanics or shops. We can demonstrate that this is not the case with a permutation test that plots the distribution of the estimated effects when dropping any two mechanics or all mechanics at any two shops. These results are in Figures A4~~3~~ (excluding any two mechanics) and A5~~4~~ (excluding any two shops). All of the permutations yield positive estimated effects, and all are reasonably close to those in the table. Thus, we find that the estimated effects are not sensitive to individual mechanics or shops, and hence are robust to small-sample inference.

~~[39] We exclude the second treatment period for shop 2, which is only 11 days long, because including it would cause an overlap of the placebo treatments for shop 2 with the actual earlier treatment period treatment for shop 2.~~

Table A1: Most frequent line-item repairs

Invoice line item	N	Mean charge ($)	CDF
environmental disposal fee	83735	2	0.124
oil filter	83709	8	0.248
top off fluids	70581	4	0.352
shop supplies	57197	18	0.437
5w30 oil	53387	11	0.516
5w20 oil	19877	17	0.545
disc brake rotor	11146	145	0.562
four wheel tire rotation	10248	11	0.577
brake inspection	4898	4	0.584
computerized diagnostic test	4598	59	0.591
air filter	4542	19	0.597
5w30	4100	11	0.604
wiper blade	3879	14	0.609
brake system flush -84032	3064	87	0.614
serpentine belt	2987	78	0.618
5w30 syn blend oil	2895	11	0.623
clean & adjust rear drum brks	2753	29	0.627
exhaust inspection	2723	3	0.631
coolant flush & fill	2686	50	0.635
bg coolant flush kit	2543	30	0.638
ceramic disc brake pads	2480	114	0.642
bleed brakes	2125	25	0.645
exhaust gasket	2071	20	0.648
friction fighter	2043	10	0.651
fuel filter	1991	62	0.654
r134a freon (1/2 lb.)	1937	60	0.657
quickstop brake pad	1897	134	0.660
hardware-gasket	1884	24	0.663
brake shoes	1858	123	0.665
gold extended-life antifreeze	1844	27	0.668
two wheel alignment	1844	45	0.671
reman caliper assy.	1818	164	0.674
machine rotors	1803	41	0.676
mega-tron battery	1734	119	0.679
tie rod end	1686	146	0.681
4 wheel alignment	1664	69	0.684
transmission flush	1661	61	0.686
wagner quickstop	1652	134	0.689
a/c evacuation & recharge	1581	64	0.691
bulbs	1572	13	0.693

Table A2: Interaction of the checklist treatment and rate of checklist use

	(1) Revenue per mechanic-week	(2) Log revenue per mechanic-week
Treatment	115.73	0.199**
	[259.85]	[0.083]
Treatment × fraction checklist use by mechanic-week	980.67**	0.490*
	[464.21]	[0.284]
Commission	414.21***	0.170***
	[135.02]	[0.049]
Observations	9,358	9,315
R-squared	0.64	0.753

Notes: The specifications in columns (1) and (2) are the same as in columns (2) and (4) of Table 5, respectively, but aggregated to the mechanic-week level and with the addition of the explanatory variables, "Treatment × fraction checklist use by mechanic-week" and a stand-alone term for fraction checklist use by mechanic-week. Heteroskedasticity-robust standard errors clustered at the shop-year level are reported in brackets. *, **, and *** indicate statistical significance at the ten, five and one percent levels, respectively.

Formatted: Justified, Space After: 0 pt, Don't add space between paragraphs of the same style, Line spacing: single

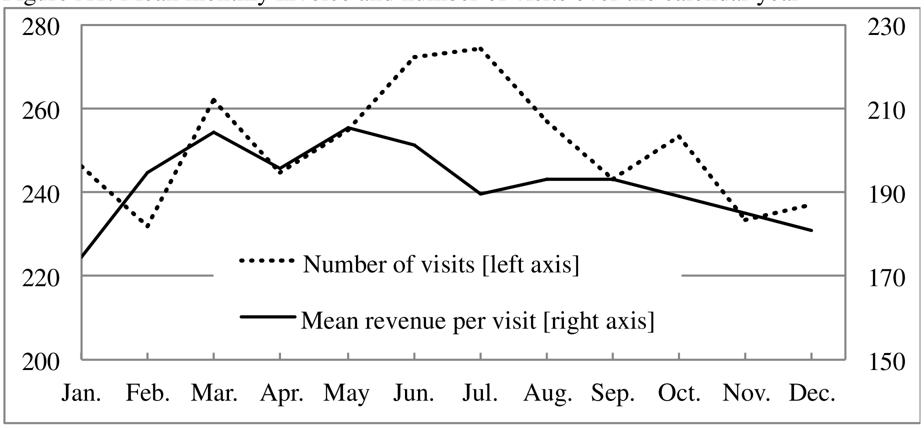
Formatted: Pattern: Clear

Table A~~2~~3: Models of revenue

	(1)	(2)	(3)	(4)	(5)	(6)
	Revenue per visit	Revenue per visit	Revenue per mechanic-week	Log revenue per mechanic-week	Revenue per shop-week	Log revenue per shop-week
Treatment			422.71**	0.350***	1,438.97**	0.125**
			[205.10]	[0.106]	[686.33]	[0.051]
Commission	22.05***	0.119***	413.01***	0.169***		
	[8.13]	[0.039]	[134.50]	[0.049]		
Mechanic FEs	X	X	X	X		
Shop-year-week FEs	X	X				
Shop FEs			X	X	X	X
Year x week FEs			X	X	X	X
Car/customer controls			X	X		
Observations	154,722	151,810	9,358	9,315	2,843	2,843
R-squared	0.19	0.29	0.64	0.75	0.64	0.63

Notes: "Treatment" is an indicator for visit during the checklist treatment. "Commission" is the commission rate of the mechanic in units of percent. The models are estimated by OLS. Treatment is omitted in columns (1) and (2) because the specifications include shop-year-week fixed effects. Commission is omitted from columns (5) and (6) because commission is a mechanic-level variable. The observations in columns (5) and (6) are weighted by the number of mechanics in that shop-week for comparability with the mechanic-week level results in columns (3) and (4). Heteroskedasticity-robust standard errors clustered at the shop-year level are reported in brackets. All models include an intercept term. ** and *** indicate statistical significance at the five and one percent levels, respectively.

Figure A1: Mean monthly invoice and number of visits over the calendar year



Notes: The data are for the full years 2009, 2010, and 2011. Subsequent data are excluded because treatments occurred at that time.

Figure A2: Sample of checklist 2

99 SATURN SL Series 7/20/12
VIN# 55882

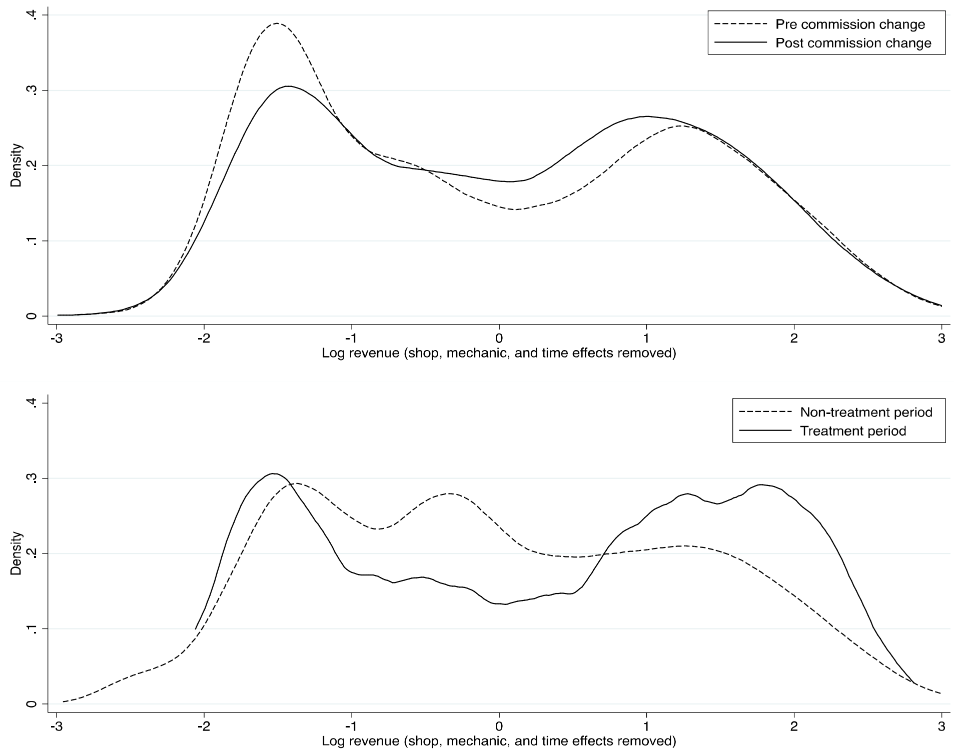
Inspection Item Required	Condition			Explain/Action
TIRES				
1. Tires - Front	OK	NA	Needs Attention	
2. Tires - Rear	OK	NA	Needs Attention	
3. Alignment (2 & 4)	OK	NA	Needs Attention	
BRAKES AND SUSPENSION				
4. Steering/Suspension	OK	NA	Needs Attention	BOTH OUTER TIE RODS / GREASE BOOTS TORN
5. R&P/Pump/Hoses/Fluid	OK	NA	Needs Attention	
6. Front Brakes	80	% Left	Needs Attention	
7. Rear Brakes	15	% Left	Needs Attention	SHOES / HARDWARE
8. Rotors and Drums	OK	NA	Needs Attention	MACH. NB
9. Hydraulics & Brake Fluid	OK	NA	Needs Attention	
10. Axle Shafts/Boots/U Joints	OK	NA	Needs Attention	
11. Front Struts/Shocks	OK	NA	Needs Attention	
12. Rear Struts/Shocks	OK	NA	Needs Attention	
EXHAUST AND ENGINE				
13. Exhaust System	OK	NA	Needs Attention	
14. Engine/Transmission Leaks	OK	NA	Needs Attention	
15. Differential Fluid Leaks	OK	NA	Needs Attention	
MAINTENANCE AND MISC.				
16. Check Engine Light	OK	NA	Needs Attention	
17. Drive Belts & Accessories	OK	NA	Needs Attention	CRACKS
18. Transmission Fluid	OK	NA	Needs Attention	
19. Cooling System Fluid	OK	NA	Needs Attention	
20. Radiator/Hoses	OK	NA	Needs Attention	
21. Heater Hoses	OK	NA	Needs Attention	
22. Battery & Terminales	OK	NA	Needs Attention	
23. Battery Cables	OK	NA	Needs Attention	
24. Tune-Up Components	OK	NA	Needs Attention	
25. Filters (Cabin/Fuel/Air/PCV)	OK	NA	Needs Attention	
26. Engine Oil (Condition and Level)	OK	NA	Needs Attention	
27. Lights (Brake, tail, turn, etc.)	OK	NA	Needs Attention	
28. Headlights/Adjustment	OK	NA	Needs Attention	
29. Wiper Blades (Front and Rear)	OK	NA	Needs Attention	

30. Factory Scheduled Maintenance Mileage: ________ Due: 3K 7.5K 15K 30K 60K 90K

Miscellaneous Comments:

Formatted: Font: (Default) Times New Roman, 12 pt

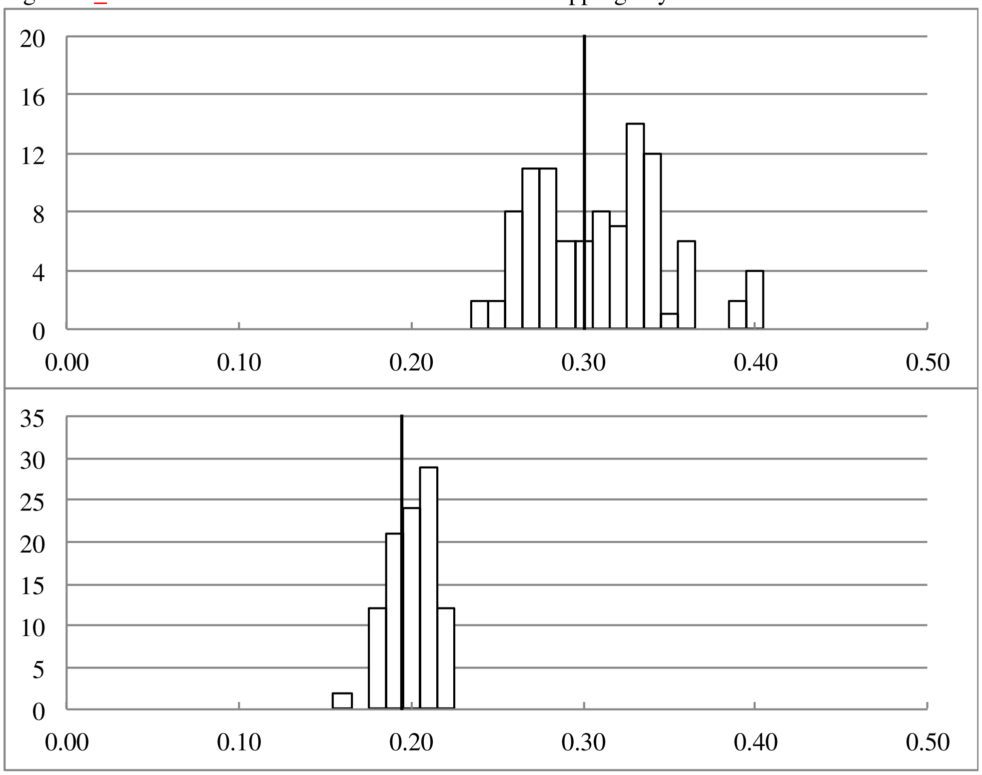
Figure A3~~2~~: Density plots of log revenue per visit



Notes: The residuals for commission mechanics from a regression of log revenue on fixed effects for shop, mechanic, and year-week are shown. The residuals are plotted with a kernel density smoother with a bandwidth of 0.2. In the top panel, the pre-commission change and post-commission change curves are for the five four-week periods before and after the commission change, respectively. In the bottom panel, the non-treatment period and treatment period curves are for the ten four-week periods before and after the treatment, and the approximately four-week period comprising the treatment periods, respectively.

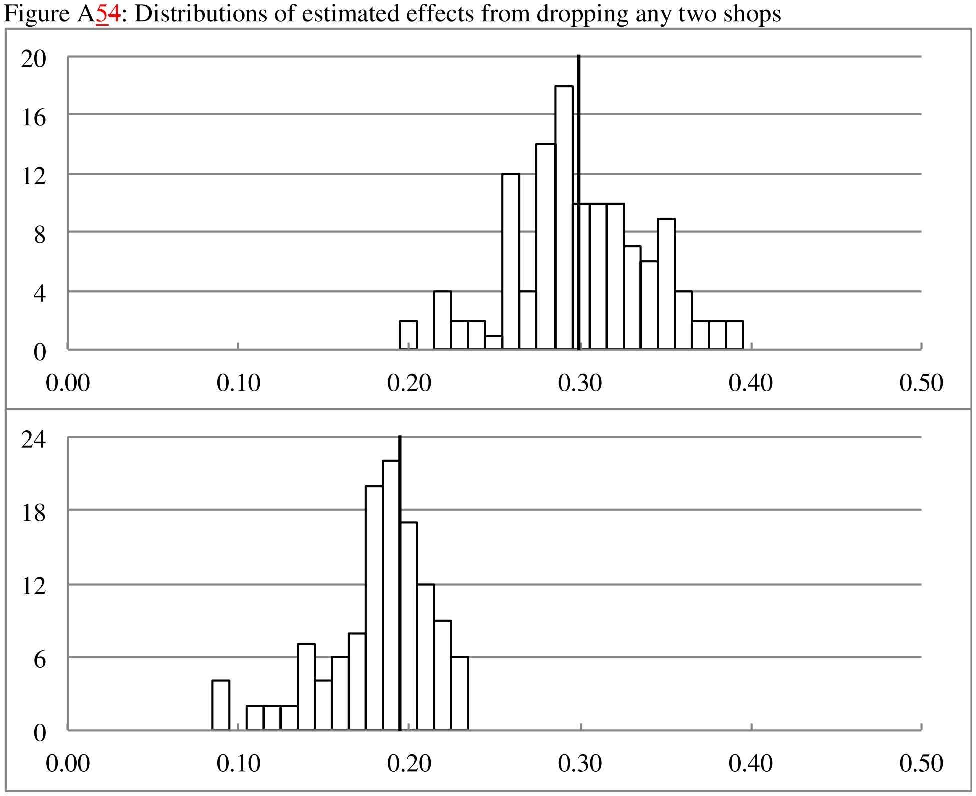
Formatted: Justified

Figure A4~~3~~: Distributions of estimated effects from dropping any two treated mechanics



Notes: The top panel shows estimates of the managerial-treatment effect. The bottom panel shows estimates of the commission effect. All combinations of any two mechanics receiving the treatment or commission increases are dropped and the mechanic-week model is estimated (from Table A3~~2~~). The vertical lines indicate the estimated effects with no dropped mechanics. The models are estimated at the mechanic-week level.

Figure A5~~4~~: Distributions of estimated effects from dropping any two shops



Notes: The top panel shows estimates of the treatment effect. The bottom panel shows estimates of the commission effect. The figures show the distribution of estimates from the mechanic-week model (from Table A3~~2~~) in which all combinations of any two shops are dropped. The vertical lines indicate the estimated effects with no dropped mechanics.