USING ARTIFICIAL INTELLIGENCE TO REDUCE FOOD WASTE

Yu Nu, Elena Belavina, Karan Girotra
SC Johnson College of Business, Cornell University/Cornell Tech, New York, NY 10044
{yn292, belavina, girotra}@cornell.edu

Abstract. In this study, we estimate the reduction in food waste that arises from the deployment of a system that digitally records instances of food items discarded in a commercial kitchen. We also shed light on the mechanisms that drive this impact. In a quasi-experimental setting, where the system was deployed in approximately 900 kitchens in a staggered manner, we estimate the impact using the synthetic difference-in-differences method. We find that three months after adoption, kitchens generate 29% lower food waste, on average, than they would have in the absence of the system— without any corresponding reductions in sales. Utilizing a long-short-term-memory fully-convolutional-network classifier, we document that these reductions are accompanied by a 23% decrease in demand chasing, a known bias in human inventory management. Upgrading to a system that uses computer vision to automate waste classification leads to a further 30% reduction in food waste generated by the kitchen a year after the upgrade. This further reduction is due to the accurate recording of infrequent but very high-impact instances of food wasted that employees avoid entering manually. We also observe substantial effect heterogeneity. Smaller kitchens and those with buffet service (vs. table service) experience almost double the reduction in food waste from the adoption of the system and also from the computer vision upgrade. Low and high-demand-variability sites have higher reductions from adoption than those with medium-demand-variability (42% vs 25%). The impacts of the upgrade are not detectably different with different demand variability.

1. Introduction

One-third of all food produced is wasted. In the hospitality and food service industries alone, wasted food costs over US$100 billion annually (Athmanathan 2021). The cost of wasted food can often be equivalent to a typical firm's net profits (Winnow Solutions 2024). As big as the financial consequences are, the impact of food waste on global warming is even more pernicious. About 8% of global greenhouse gas emissions come from food waste (Food and Agriculture Organization 2020). In line with the United Nations Sustainable Development Goal Target 12.3 (United Nations 2021), many nations and firms have committed to reducing their food waste by half by 2030. However, the path to achieving this goal remains unclear.

Key words and phrases. Food Waste; Artificial Intelligence; Sustainability; Computer Vision; Impact Analysis; Perishable Inventory Management.

To better manage food waste, accurately measuring and monitoring what is wasted is essential. However, retail-scale food production and service environments are notoriously data-scarce. A typical restaurant only records the total raw materials purchased on a weekly or monthly cycle, with no data on the precise quantities of ingredients that pass through each stage of food production and service. In the rare cases that food waste is recorded, it is typically just the total weight of organic waste collected by a waste collection company at a weekly or even monthly level. This makes it hard to track what specific items are wasted and identify the root causes of this waste. Further, in contrast with other modern production processes, there is no data on work-in-process inventory, amount of finished goods produced, on-hand inventory, leftover inventory, or lost sales. In buffets, there is also often no "sales" data. The wide variety of non-standardized raw ingredients, prepared items, components, and final products in food production have traditionally necessitated manual measurement and tracking, which makes measurement not economically viable. Recent advances in AI/image classification technologies have the potential to automate this measurement.

Winnow Solutions, a UK-based company, has developed a digital food waste tracking system. There are two versions of the system: the original "Winnow Classic" system and a more recently upgraded "Winnow Vision" system. The Winnow Classic system comprises a digital scale and a connected digital tablet that are retrofitted on existing trashcans in commercial kitchens. Whenever food is thrown away in a trashcan, the weight of the item wasted is automatically recorded. The user is then manually prompted to identify the item via an easy-to-use menu of options available on the tablet and select the reason for the waste. Winnow Vision goes further and eliminates the need for manual identification of the items wasted. The vision system includes a digital camera that is inconspicuously mounted over the trashcan. Using modern AI/image classification algorithms, the system automatically identifies what item is wasted to complement the weight measurement. Food waste is automatically categorized, with little need for manual input. Overall, using Winnow systems, kitchens can easily measure and stratify their food waste down to the level of each disposal transaction, classifying waste as different items at different stages of preparation and processing. The collected data is made available to kitchen managers in daily, weekly, and monthly site reports. Kitchen managers also have access to a portal to dig deeper into the data and conduct further analytics.

This study is based on a long partnership with Winnow Systems. In collaboration with Winnow and with the data collected by the deployed Winnow systems at hand, in this study, we set out to

(1) Provide the first systemic census of the food wasted in different types of food service establishments;

(2) Estimate the impact of the availability of detailed waste data on the amount of food wasted;
(3) Estimate the additional benefits, if any, from the AI/Image-classification-enabled automated data collection and
(4) Identify the mechanisms that drive the benefits of data availability and AI-enabled automation of data collection.

Our study exploits the staggered deployment of the original waste recording system, Winnow Classic, in hundreds of commercial kitchens and the subsequent staggered upgrade of some of these systems to the automated Winnow Vision system. We provide causal estimates of the change in food waste generated due to the adoption of the system and the subsequent upgrade. We also try to explore the potential mechanisms that lead to the change. We use the synthetic difference-in-difference method (SynthDiD) with staggered treatment times (Berman and Israeli 2022, Porreca 2022) to identify the effect of the adoption and that of the upgrade. This method allows us to overcome several identification challenges arising from the infeasibility of running large-scale, randomized controlled trials in the field where the subjects are many different firms. We confirm the robustness of our results using several other methods.

To explore the mechanisms of action, we utilize a machine learning classifier to detect behavioral patterns from waste data and compare the patterns before and after the adoption/upgrade. Specifically, we develop a time-series classifier based on the long-short-term-memory fully-convolutional networks (LSTM-FCN) algorithm. Based on the waste data collected, this classifier allows us to identify, with high accuracy, the incidence of common behavioral biases. We then use the DiD approach to examine whether this incidence changes on account of the adoption/upgrade of Winnow. Overall, our analysis sheds light on how information collection enabled by digital technologies and AI can influence the management of food waste, or, more generally, how AI and digital technologies can help better management of inventories and operations in fast-paced, dynamic, high-uncertainty, high-obsolescence production settings.

Our main findings are:

Census of Food Waste. Our census of food waste gives us a sense of the extent of food waste in establishments. We find that the average pre-adoption waste level of buffets is nearly six times higher than that of counter-service and double that of table-service settings. They waste 2042, 358, and 826 grams/cover, respectively. Hotel-based food service establishments waste more than double that of independent restaurants (2361 vs. 856 grams/cover), which, in turn, waste more than double that of quick-service, staff, and healthcare food service settings (422, 260, and 342 grams/cover, respectively). Most food waste originates from leftover/expired inventory (48%: 40%

prepared food and 8% raw), followed by food trimmings, plate waste, and cooking & handling errors (14%, 10%, and 7%, respectively).

Impact of Adoption and Upgrade. We find that, on average, kitchens that adopted the Winnow Classic system produce *23% less food waste* than they would have otherwise produced two months after adoption. This figure rises to *29% three months post-adoption.* Notably, this reduction happens without any corresponding decrease in sales. Further, on average, the sites that upgrade to the computer-vision-based Winnow Vision experience an additional food waste reduction of about *26% nine months post-upgrade, rising to about 30% twelve months post-upgrade* (also without any loss in sales). To put this in perspective, ambitious food service establishments have set targets to halve their food waste within a decade (World Resources Institute 2016). By adopting Winnow Classic, a commercial kitchen could achieve a 29% waste reduction in a mere three months! With Winnow Vision, the 50% reduction goal could likely be achieved within a year. In other words, a simple monitoring system can bring remarkable food waste reduction benefits and provide a clear path to meet even very ambitious climate goals. Note that these reductions are achieved with no revenue changes and a significant cost decrease. Thus, this brings about a win for the business and the environment, all achieved by using digital technology and AI.

Heterogeneous Effects. To sharpen our findings and make them generalizable to a wider variety of food service establishments, we explore how the above impacts vary by site characteristics. Three months post-adoption, smaller establishments see almost double the reduction compared to larger ones (42% for those serving below 800 covers per day vs. 22% for those above). Buffets also see a higher reduction than table sites (36% vs. 26%, respectively). Sites with very low or very high demand variability (coefficient of variation below .5 and above 1, respectively) also see a higher reduction than those with medium demand variability (42% for high- and low-variability sites vs. 25% for medium-variability sites). Similar to the effect of Winnow Classic adoption, one-year post-vision upgrade, smaller sites (buffets) experience double the reduction compared to larger sites (restaurants) on account of the upgrade: 21% vs. 57%.

Mechanisms–Adoption. We find that the primary driver of the waste reduction due to the adoption of Winnow Classic is the significant decrease in the leftover inventory of prepared and raw food items. These are about 32% down, on average, in three months. The reductions in other kinds of food waste (trimmings, cooking errors) are less pronounced. Using our machine-learning-based classifier, we next attempt to provide some evidence on the behavioral phenomena that drive waste reduction. We measure the change in how often the inventory decisions are biased away from optimal decisions in several specific ways previously identified in the behavioral inventory management literature. We observe that the waste reduction on adoption comes with a significant decrease in the likelihood of demand chasing. Three months post-adoption, we observe a 23% reduction in the incidence of demand chasing, with a corresponding increase in the likelihood of having waste outcomes consistent with optimal/bias-free decision-making. The incidences of the other biases (overreaction, static production plans) do not experience a significant change.

Mechanism–Upgrade. We find that within a year, the Winnow Vision upgrade significantly decreases leftover inventory, cooking errors, and trimmings (30%, 21%, and 43%). These additional benefits come from a better "quality" of information. Interestingly, with Winnow Classic, on average, only 3% of food waste "events" are uncategorized. However, these 3% of waste events account for about 26% of the total waste amount by weight, suggesting that when large amounts of food waste are generated, manual operators do not categorize these large waste transactions. Winnow Vision, however, leaves no option not to categorize such waste events. Our data shows a virtually complete categorization of all waste (by number of events or by weight) after the upgrade. In effect, the automated vision system prevents large waste events (be it massive overproduction/over-purchase or cooking errors) from being hidden by incomplete manual entries. The ability to see and control these large waste events perhaps drives the additional food waste reduction in leftover inventory and cooking errors (and, as a result, the corresponding reduction in trimmings) that we observe on account of the vision upgrade. We do not see a significant change in the behavioral biases after the upgrade, likely because the additional events captured by the vision system are larger straightforward errors (inventory and cooking) and have less to do with the managerial response to demand changes.

Finally, we see some changes in the consumer-level waste—an average 16% plate waste reduction three months after Winnow Classic adoption and an additional 18% reduction one year after the Winnow Vision upgrade. Although these reductions are not statistically significant and their genesis is unclear, these observations suggest that kitchen managers are able to influence the consumer-driven portion of food waste as well, possibly by reducing portions, changing plate sizes, etc.

Our study has important implications for practicing managers and for building our understanding of the use and effectiveness of technology in production and service operations.

For practitioners in the food service, hospitality, and other related industries, the implications are clear. The broad insights from our food-waste census on the prevalence, genesis, and variation of food waste provide a report of the status quo at food-service establishments and suggest clear improvement paths. Our estimates provide rigorous benchmarks on the potential benefits of adopting

Winnow or other food waste monitoring technologies.[1] Sites similar to the ones in our sample can reasonably expect to achieve environmental and financial outcomes like the ones estimated in our study. Our results on the effect heterogeneity can help potential adopters obtain sharper benchmarks based on the particular characteristics of their food operation (size, service type, and demand variability).

For practitioners in other industries, while our estimates may not be representative of the achievable gains, the structural phenomena identified in our study are likely to continue to hold: Using digital technology to monitor excess inventories in production and service systems reduces excess inventory on account of lower incidences of behavioral biases. Automated classification reduces the ability of employees to disguise large-impact errors, among other findings.

For civil society organizations, policymakers, and multilateral institutions that are working on climate goals and reducing food waste, our work provides an evidence-based, low-cost, profit-increasing path to achieving very significant reductions in carbon emissions.

For the technology, AI impact, sustainability, and operations management research communities, our work advances our understanding of how technology mediates better operational and environmental outcomes. To the best of our knowledge, we provide the first field evidence on how the availability of digital information on leftover inventories changes these outcomes, and the particular behavioral biases that drive the gains. We also provide the first real-world evidence of the benefits of using AI to increase operational efficiency and the potential pathways. Methodologically, our machine-learning-based time-series classifier is a new method for the online detection of biases in inventory operations, a sort of AI-driven warning system.

2. Literature Review

Our paper contributes to ongoing research streams that study (1) food waste and sustainable operations management, (2) perishable inventory management, (3) digital-technology-enabled interventions, and (4) AI and productivity.

2.1. Food Waste & Sustainable Operations Management.

Prior studies have examined food waste at different tiers of the food supply chain. At the retail level, Akkas et al. (2018) first identified the drivers of in-store product expiration using large-scale data. Akkas (2019) analyzes the impact of allocated shelf space on product expiration and formulates a shelf space selection problem. Via structural econometric modeling, Sanders (2020) evaluates the welfare of two potential remedies for food waste in grocery retail: dynamic pricing and organic waste landfill bans. More recently, Akkas

[1] Most of the existing food waste tracking solutions are very similar to Winnow. Like Winnow, they provide (almost identical) tracking/descriptive feedback.

and Honhon (2022) study how shipment policies (i.e., the rules to determine the quantity and age composition of inventory to ship from a warehouse to a retail location) affect profits and waste. Yang and Yu (2023) study surprise clearance as a novel scheme to increase retail profit and reduce food waste. Keskin et al. (2023) quantify the value of blockchain-enabled freshness transparency by examining retail profit growth and food waste reduction brought by blockchain adoption. In addition to the retail-level waste that these studies consider, Belavina et al. (2016) examined food waste generated by households when comparing the financial and environmental performance of different revenue models offered by online grocery retailers. To explore the environmental impact of introducing an online channel, Astashkina et al. (2019) assesses food waste in the entire supply chain—at the supplier, retailer, and consumer tiers. Furthermore, Belavina (2021) studies the impact of grocery store density on retailer and household waste.

We contribute to the growing food waste literature by examining the waste generated at food service establishments. We estimate the reduction in food waste from the deployment and upgrade of a technology that digitally records instances of food items discarded in a commercial kitchen. We also shed light on the mechanisms that drive this impact.

2.2. **Perishable Inventory Management.** Our work also relates to the perishable inventory management literature. Nahmias (2011) and Karaesmen et al. (2011) provide complementary reviews of advancements in the OM research on perishable inventories, with a focus on the computation of optimal and heuristic ordering policies. Many studies also explore markdown management. For example, Li et al. (2016) and Hu et al. (2016) study joint replenishment and clearance sales of perishable goods. Going beyond the focus on optimal stocking levels in the literature, Astashkina et al. (2019) and Belavina (2021) examine the perished inventories that result from these optimal decisions and the associated carbon emissions via multi-echelon perishable inventory models with spatially distributed agents. Further progress on the impact of incorporating food waste into supply chain decisions, however, has been stalled by the unavailability of inventory-age data, in practice. As such data becomes more readily available, there will be a push toward the development of new algorithms for better management of perishables (Akkas and Gaur 2022). With the first-of-its-kind data, we contribute by providing statistics on the nature of food waste in commercial kitchens, quantifying the waste reductions on account of detailed waste information gathering and shedding light on how the information feedback on food waste can influence perishable inventory management.

2.3. **Digital Technology-enabled Interventions.** There is extensive literature studying the impact of implementing technology-enabled interventions. Pierce et al. (2015) finds that the implementation of a monitoring system (enabled by back-office technology) reduces employee theft and improves productivity in a casual restaurant setting. Staats et al. (2016) explores the effectiveness of RFID-based electronic monitoring on hand hygiene compliance in hospitals and its persistence. Anderson and Kimball (2019) document increased student learning performance due to teachers' use of performance measurement systems for diagnosing and remediating problems in students' daily learning. Soleymanian et al. (2019) study how driving behavior is affected by participating in the telematics-based usage-based insurance (UBI) program, where drivers receive real-time feedback about their driving behavior in exchange for potential discounts on future premiums from auto insurers. They also show that after UBI adoption, UBI users drive safer. Berman and Israeli (2022) demonstrate an increase in online retailers' revenues post-adoption of a descriptive analytics dashboard. Our research examines the change in food waste in commercial kitchens due to the use of Winnow systems. This system enables granular capture and feedback on the amount and the origin of the food wasted. We, thus, contribute to the OM literature that studies the impact of various technology-enabled interventions, such as monitoring and feedback, to drive operational improvements.

2.4. **AI and Productivity.** Recent years have witnessed a surge in the development of AI to assist human users across diverse domains (Furman and Seamans 2019). Existing research on AI applications revolves around two primary areas: 1) AI for automation, such as automating repetitive data entry processes, fluent text generation, and improving data accuracy (Defize et al. 2022, Babina et al. 2024); 2) AI-co-pilots for feedback or advice, such as AI-assisted medical diagnosis and other applications based on Large Language Models (LLMs) (Brynjolfsson et al. 2023, Otis et al. 2024). Numerous studies have provided valuable insights into the effectiveness of different ways to incorporate AI tools in the decision-making processes of managers across various domains, spanning project management (Beer et al. 2022), healthcare (Liu et al. 2022, Lin et al. 2023), retail (Kawaguchi 2021, Sun et al. 2022), procurement (Cui et al. 2022), idea generation (Girotra et al. 2023), transportation (Cui et al. 2023), consulting (Dell'Acqua et al. 2023), freelance work (Hui et al. 2023) and law (Cohen et al. 2023). Our research is in the same stream but focuses on a new context— the use of computer vision-based food waste information gathering in the context of production operations.

3. Institutional Context

Our study estimates the impact of granular (AI-enabled) food-waste recording on the amount of food waste generated in commercial kitchens. In particular, we consider kitchens of standalone restaurants as well as kitchens of food-service establishments in hotels, hospitals, corporate offices, resorts, casinos, cruise ships, and retail stores. Some establishments in our study have only one kitchen, while others are multi-location chains with several identical kitchens run by the same company.

A typical commercial kitchen is a fast-paced environment, and relative to mechanized production facilities that are the focus of most operations management research (for ex., those for cars, electronics, etc.), there is far less standardization, automation, and optimization. These kitchens are staffed with cooks and managers, who often have limited operations training or the time to optimize the operations; the focus is on food taste and presentation. Quality management programs that are standard in other facilities are rare. Most importantly, in a typical kitchen, there is no instrumentation of the production processes, and as such, there is very little in-process data. It is common for a kitchen offering a buffet service to have no data beyond the basic raw materials ordered on a weekly or monthly basis. Table-service kitchens may also have access to point-of-sale data, but even in these cases, the SKUs sold are often not well standardized, and there is rarely an accurate mapping between final goods and the semi-prepared items and raw materials that go into them.

Our study is built around the rollout of a new information-gathering system, developed by Winnow Solutions. The Winnow system records the food items thrown in the trashcans deployed in a kitchen. There are two variants of the system—Winnow Classic, which is essentially a digital weighing scale attached to a tablet, see Figure 3.1(a). Trashcans are placed on top of a Winnow scale, which allows for accurate measurement of the weight of food items thrown in every use of the trashcan. Users can use the attached tablet to record specific food items thrown in the trashcan and the reason why the item was thrown— unsold prepared food, inventory spoilage, cooking & handling errors, waste from preparation (such as trimmings), or food left on the consumer plates.

For recording the specific food item, users must use a menu-based system— to first identify the category of the item (desserts, vegetables, etc.) and then the specific item. The list of items is customized to each establishment, and is organized in a smart way, with the most likely items first on the list, but given the large number of potential items, the user may need to click through several times to record the item correctly, see Figure 3.2.

Winnow Vision is an upgraded version of this system. See Figure 3.1(b) that uses AI-based automatic classification of the items thrown, drastically simplifying data entry. This version includes everything in the classic system plus a small wide-angle camera that is motion-activated at the base of the tablet assembly. The camera takes a picture of the trashcan right after the food is thrown. This picture is compared with an older picture. The changes in the two pictures are then classified using a modern deep-learning-based image classification algorithm developed in consultation

(a) Winnow Classic (b) Winnow Vision

FIGURE 3.1. Winnow Systems Demonstration

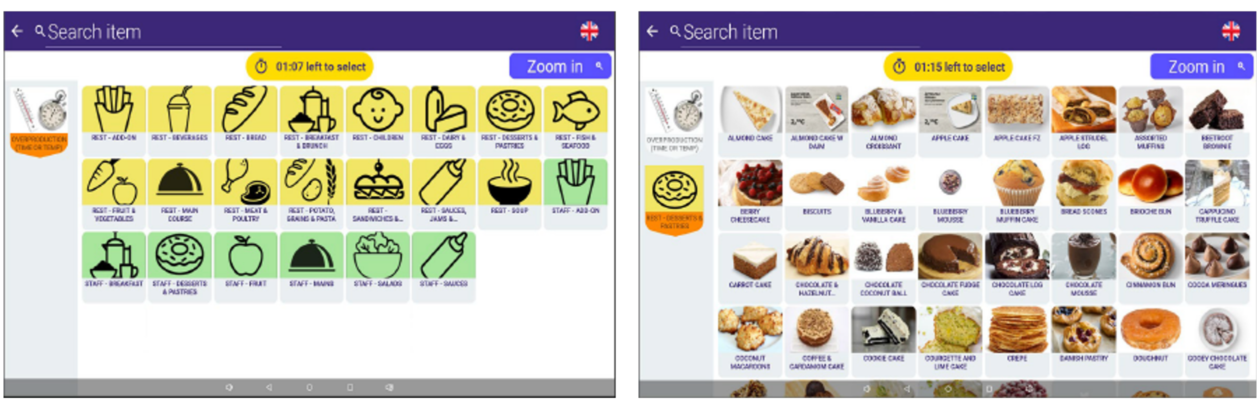


FIGURE 3.2. Two-Step (Category & Item) Data Entry Process on the Touchscreen

with the study authors. The algorithm is trained on manually-labeled food-waste images. Overall, Winnow Vision can automatically classify food items with an accuracy level surpassing humans. Data collected by Winnow is stored on Winnow's cloud platform, which kitchen managers can access. The platform shows various statistics of the items wasted. Kitchen managers also receive

Figure 3.3. Winnow Reports

daily, weekly, and monthly reports via email, and they can access additional data via the cloud platform (see Figure 3.3).

4. Theoretical Expectations

Our study estimates the effects of providing past food waste information to kitchen managers who are responsible for making decisions around raw materials ordered, items prepared for cooking (work-in-progress inventory) and the final amounts of food cooked and served. Each of these decisions can be conceptualized as a multi-product, multi-period, perishable inventory system with continuous review and non-stationary demand. Food waste information is the information on the left-over expired inventory in preceding periods, inventory that arises as a function of the quantities produced, demand, and food expiration.

Prior to the installation of Winnow, decision-makers in the system had access to (at best) limited crude data on overall final item sales. The Winnow system uses modern digital technology and AI/Image classification algorithms to provide granular product-level left-over-inventory information to the decision-makers. In this section, we leverage past work in operations, economics, and psychology to build a theoretical understanding around (a) the potential effects of providing additional information to decision-makers, (b) the potential mechanisms of action for these effects, and (c) how these effects vary between different settings.

4.1. **Effects of Tech-Enabled Information Gathering.**

Reasons to expect food waste reductions. Extensive work has shown that technology enabled information gathering can improve performance across various domains, such as utilities (Thaler and Benartzi 2004, Gaker et al. 2010), health (Staats et al. 2016, Kim et al. 2020), and transportation

(Choudhary et al. 2021). The literature identifies four broad arguments for this, that we conjecture are relevant in our setting.

First, technology adoption can remove information hurdles, complementing workers by freeing up their cognitive capacity (Brynjolfsson and Hitt 2000, Hitt and Tambe 2016); we expect similar effects with the Winnow system. Second, information saliency promotes better goal-setting and induces greater effort (Anderson and Green 2018). As is shown in Bandura (2010), goals based on one's prior performance are deemed achievable through increased self-efficacy, and this increased self-efficacy results in increased and persistent effort, which, in turn leads to better performance. We conjecture Winnow increases information saliency and should promote better goal setting and induce greater effort.

Third, feedback on prior performance also serves as a basis for evaluating one's ability to successfully perform subsequent tasks (Bandura 1991). Given that human decision-making is often biased (Tversky and Kahneman 1973), reporting prior performance is a low-cost intervention to exploit human behavioral tendencies and thus initiate better decision-making. For example, Blader et al. (2020) conducted a field experiment with a large US transportation company offering electronic onboard recorders that report drivers' performance automatically, and they found that providing feedback leads to better driving performance. Many studies in the behavioral OM literature (e.g., Schweitzer and Cachon 2000, Bolton and Katok 2008) have also shown that in NewsVendor games, when players are provided with feedback on realized demand and profitability after each round, player performance improves. Although our setting is more sophisticated than the classic lab-experiment-NewsVendor, and the information provided is more nuanced than that in the lab experiments (leftover inventory as opposed to profits), we hypothesize that the Winnow system provides important feedback on prior performance and will lead to performance improvements.

Fourth, the Winnow system can also be viewed as a monitoring tool. The mere presence of monitoring may change user behavior by making individuals feel accountable for their now observable actions. For instance, prior work has documented that simply making one's action visible to others may influence various categories of behavior ranging from voter turnout to employee theft in restaurants (see Gerber et al. 2008, Pierce et al. 2015). Additionally, monitoring may signal management's commitment to process compliance. Winnow is, in effect, an employee monitoring tool, and its adoption signals management's ambition to reduce food waste.

Overall, we expect that the Winnow system will remove information hurdles and free up cognitive capacity, increase information saliency to promote more goal setting and induce greater effort, provide inventory managers with performance feedback and signal the priority management places on food waste and holding individuals accountable for their actions related to food waste. The combination of these effects should lead to a reduction in food waste after the installation and use of the Winnow system.

Reasons to expect food waste increases. Notwithstanding the benefits of technology-driven interventions, there are some additional concerns. Information provided by new technologies can be often classified into four categories (Lismont et al. 2017): (i) descriptive (what happened), (ii) diagnostic (why did it happen), (iii) predictive (what will happen next), and (iv) prescriptive (what should be done about it). Whereas the potential benefits of predictive or prescriptive technology that uses sophisticated modeling have been well-studied (e.g., Wedel and Kannan 2016, Bradlow et al. 2017), the value of descriptive information (as is the information from the Winnow system in our study) is questionable, as receivers may misinterpret and fail to turn them into actionable insights (Berman and Israeli 2022). Kluger and DeNisi (1996) found in their meta-analysis that 38% of the studies report a negative effect of feedback information on performance. Concurring with their findings, several recent studies have demonstrated feedback's inefficacy in improving user performance. For example, Rolim et al. (2017) observed that bus drivers do worse after real-time feedback because they only focus on some safety parameters while overlooking others. There is an additional psychological concern. Technology-driven high-resolution data collection can play the role of a monitoring tool, but it can also be perceived as invasive and a signal of distrust by management, which could result in reactance or reduced compliance (Frey 1993, Bernstein 2012). This effect is likely to be quite salient in commercial kitchens. As Lemos (2019) summarizes, chefs are extremely busy running demanding commercial kitchens and do not have the time or patience to invest in the right initiatives. Also, chefs usually have worked in the industry for a long time and might feel suspicious about the outside intervention and invasive monitoring.

Overall, sound arguments and evidence from other settings indicate that leftover inventory or food waste information from prior periods should reduce food waste. At the same time, arguments and prior evidence suggest that simple descriptive information can be harmful in settings like commercial kitchens. We speculate that the positive effects dominate. In other words, we believe that *having detailed food waste information from prior periods should reduce the amount of food waste* generated, though we don't hold this opinion strongly.

From a practical commercial point of view, the effect's direction is insufficient. The real concern is more nuanced— what is the extent of the benefits of food waste monitoring systems, in commercial kitchen settings, if any? In line with this practical concern, we will design our study to measure the level of the benefits/losses of these information-gathering systems rather than just assess the direction.

4.2. **Correction of Inventory Management Biases.** The factors discussed in Section 4.1 pertain to the role of additional information in a generic decision-making task. We next delve into the specific case of inventory decisions. The value of additional information in inventory management decisions manifests itself primarily via a better understanding of the demand. Behavioral operations management researchers have identified that inventory managers often misuse demand information (Kremer et al. 2011). In particular, the literature has identified three common biases. First, inventory managers may not respond sufficiently to demand variability. That is, they follow simple heuristics or rules of thumb and produce a certain amount of food without sufficient adjustment in accordance with demand changes when making decisions (see Kahneman et al. 1982, Su 2008, Gino and Pisano 2008). We refer to this decision bias as a *static production* plan. Conversely, inventory managers may overemphasize demand signals by *demand chasing* (Schweitzer and Cachon 2000, Lau and Bearden 2013, Kirshner and Moritz 2020) such that the order quantities are adjusted towards the demand in the prior period. Along similar lines, managers may overreact to *demand changes* (Watson and Zheng 2008) such that the order quantities are increased/decreased in response to the sharp demand increase/decrease in prior periods. We describe these further in Section 8.2 and Appendix A.7.

In commercial kitchens, contemporaneous demand realizations are often not observed (for example, in buffet-type settings), are observed by different decision makers (for example, the commercial manager rather than the chef or kitchen manager who makes inventory decisions), or are observed with a significant lag. The Winnow system makes leftover inventory information very salient to the kitchen manager. This information combined with direct knowledge of the recent inventory decisions made can allow inventory managers to get a better sense of the recent demand realizations. As such, we conjecture that *the installation of the Winnow system and the information gathered by it would limit the incidence of the three behavioral biases.*

4.3. **Effect Heterogeneity.** Beyond the average effect of information, there may also be significant differences in the effects amongst different kinds of food service establishments. One potential cause of such heterogeneity is the site size. The larger sites enjoy the benefits of statistical pooling and, as a result, experience less waste per meal served. Also, they might be able to better optimize their operations. Thus, we expect *smaller sites to benefit more from Winnow.*

Service type (buffet-, table- or counter-type restaurants) may also drive the heterogeneity in the effects of Winnow. A wide variety of foods are often served at buffets, making item-level information gathering very challenging. Also, the buffets' make-to-stock strategy is much more vulnerable to inaccuracies in demand forecasts (Wu and Teng 2023), compared with table- or counter-type restaurants. Thus, we expect *buffets to benefit more from Winnow (as compared with table- or counter-type restaurants).*

We also consider how different levels of demand variability may lead to different outcomes. When the service level is sufficiently high, as typical in food service establishments, higher demand uncertainty induces higher inventories (Rubin 1980, Gerchak and Mossman 1992, Song et al. 2009) and, as a result, food waste. Thus, sites with higher demand uncertainty might expect more benefits due to Winnow adoption. Further, the deployment of Winnow systems could make demand variability management more effective by providing incremental (waste) information. This additional information could allow users to keep revising their demand forecasts and partially resolve the faced uncertainty over time (Graves et al. 1986, Heath and Jackson 1994, Chen and Lee 2009). Finally, it is likely that establishments with lower demand variability might be able to use this additional information with higher precision. As a result, the impact of demand variability on waste reduction outcomes is uncertain, and *the sites with lower or higher demand variability might see more benefits from Winnow.*

4.4. **Effect of AI/Computer Vision.** AI is revolutionizing data capture (Defize et al. 2022). Traditional data entry can be time-consuming and error-prone. Winnow Classic requires the user to identify the item thrown in the trashcan and manually indicate what is being wasted. Such manual entry is labor-intensive and demands significant time and effort. Further, humans are prone to errors. Even the most diligent data entry operators can make mistakes and miss entries; this is much harder for busy kitchen staff. AI algorithms, on the other hand, can automate repetitive data entry processes and improve data accuracy (Babina et al. 2024). The Winnow Vision system uses a modern deep-learning-based based image-classification algorithm to automate waste information gathering. The AI-enabled system recognizes the item thrown from images captured with high accuracy, delivering more accurate data with minimal effort. Automation is a massive boost to efficiency. By taking repetitive tasks off the plate, Winnow Vision users can free up time to focus on other aspects of their work, which will arguably lead to better waste categorization by origin, more engagement with the collected data, and enhanced productivity in the kitchen.

Overall, the addition of a computer vision element should provide better data quality, reduce human errors, improve waste data accuracy, and increase general comfort with the use of the Winnow information-gathering system. However, the literature also presents some competing arguments—humans can be averse to the input from algorithms (e.g., see Dietvorst et al. 2015, Dietvorst and

Bharti 2020). Cao and Zhang (2020) also provided first-hand field evidence that human workers are reluctant to adopt AI. This aversion and reluctance to AI and automation may make inventory managers distrust or ignore the data provided by the Winnow Vision system. That said, our setting is somewhat different from the cases mentioned above in that AI is used to help achieve managers' professional and prosocial goals (e.g., to run a more efficient and sustainable kitchen) rather than substitute them. As such, we expect *Winnow Vision to achieve further reductions over and above Winnow Classic's.*

5. Data

Our data spans 876 sites from 71 different food service establishments that installed the Winnow Classic system in a staggered way between May 2016 and Nov 2019.[2] A typical food service establishment operates kitchens at multiple different sites. At each site, the rollout begins with a blind period that lasts between two weeks and four months; the installed system collects baseline data for Winnow's internal use, while the site operates as usual and does not have access to this data. There is no involvement from the kitchen staff other than throwing food in the trashcan as before. The blind period is designed to ensure that the kitchen staff does not notice any changes in the operation and that the tablet and other items are disguised. After the blind period, the system is officially activated. From this point, the staff gains access to real-time feedback on the items thrown in a shift and all archival data via periodic reports and a data portal.

After using the Classic system for a while, some sites upgraded to the Winnow Vision system, eliminating the need for manual entry. In the initial days after the upgrade, photos taken by the camera were manually labeled at the backend by the Winnow team to generate site-specific training data and validate algorithm performance at the specific site. All sites eventually have a custom-trained, fully automated computer vision system that automatically identifies the specific item thrown in the trashcan.

The Winnow system logs every item thrown in trashcans at participating sites. These "waste events" constitute our primary data. Overall, there were 12,828,147 waste events in the study period. For each waste event, we have the weight of the item thrown, the exact item thrown (obtained by manual entry in the classic system and by computer vision in the vision system), and a manually entered reason for the waste— unsold prepared food, inventory spoilage, trimmings, cooking & handling errors, or plate waste.

[2]We limit our study to installations prior to the breakout of COVID-19 as COVID disrupted operations and led to downsizing and wide shutdown in many commercial food service establishments.

Variables		**Mean**	**Standard Dev.**	**Min**	**Median**	**Max**	**N**
Daily total waste (weight)	grams	386,007	648,657	10	185,840	37,820,840	404,763
# of meals served daily (covers)	count	1,939	1,727	1	1,416	24,844	404,763
Daily sales (sales)	US dollars	13,555	11,745	1.07	10,934	348,941	404,537
Demand variability (demand_cv)	coef. of variation	0.2440	0.2429	0.0011	0.1856	2.8284	397,029
Waste per cover (waste_per_cover)	grams/cover	544	1,908	0.0021	142	165,800	397,029

Table 1. Summary Statistics

We supplement this waste data with the manual collection of data on the type of food service establishment each site was associated with— hotel, restaurant, staff restaurant, quick-service restaurant, or healthcare facility. Restaurants dominate our sample (75.5%), but we also have hotels (11.1%), staff restaurants (3.5%), quick-service restaurants (1.1%), and healthcare facilities (0.2%). Finally, we observed and recorded the type of service each site offered— buffet, counter, or table service. In our sample, table service is the most common (70.2%), whereas 18.5% of the sites are buffets and 3.5% are counter services.

We aggregate the raw waste data at the day-site level. Waste in grams at site i on day t is captured by $weight_{it}$; $waste\_per\_cover_{it}$ is the normalized version of this waste measure, computed as $weight_{it}$ divided by the number of covers served at site i on day t.

We capture the variability in the site's demand by computing the coefficient of variation based on historical daily sales data:[3]

$$demand\_cv_{it} = \frac{demand\_std_{it}}{demand\_mean_{it}} \approx \sqrt{\frac{\sum_{k\in\Omega_{i,t}} \left(sales_{i,k} - \sum_{h\in\Omega_{it}} sales_{i,h}/|\Omega_{i,t}|\right)^2}{|\Omega_{i,t}|-1}} \Big/ \left(\sum_{h\in\Omega_{i,t}} sales_{i,h}\right)$$

where $\Omega_{it} = \{sales_{i,t}, sales_{i,t-1}, ..., sales_{i,t-6}\}$.

Table 1 presents the summary statistics.

Additionally, Figure 5.1 (a)-(d) depicts the distributions of the key variables. Figure 5.1(e)-(h) visualize the pre-adoption waste statistics by site type and service type, respectively. The average pre-adoption waste generated at buffets is nearly six times higher than that of counter service and almost double that of table service settings. They waste 2042, 358, and 826 grams/cover, respectively. Hotels waste more than double that of restaurants (2361 vs. 856 grams/cover), which

[3]Ideally we would construct this variability measure from historical demand data, but what is observed here is only censored demand data, i.e. sales. It is typical in industry to use sales rather than demand (see Schleifer 1995, Cachon and Terwiesch 2012), as such sales data is all that is available to the site managers (and us). With the high service levels typically provided by food-service establishments, we expect that demand censoring does not significantly alter this variable.
We also ran our analysis with demand variability variable constructed using past sales on only the same day of the week data (following Jain et al. 2013, see below): $\Omega_{it} = \{sales_{i,t}, sales_{i,(t-7*1)}, sales_{i,(t-7*2)}, sales_{i,(t-7*3)}...\}$. Our main results continue to hold under this alternative specification.

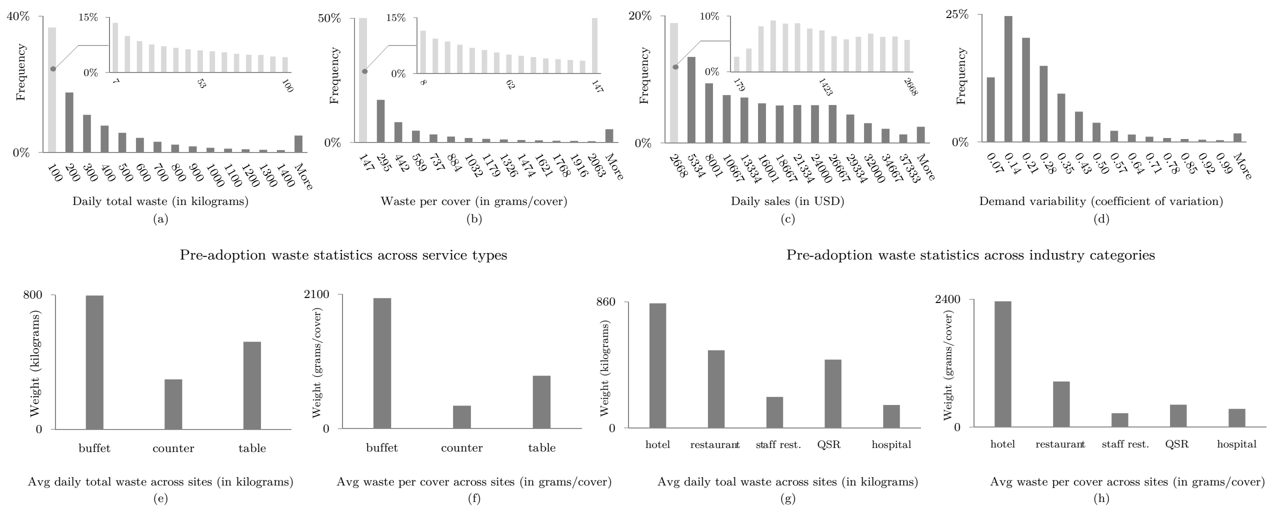


FIGURE 5.1. Distributions of Main Variables

themselves waste more than double that of quick-service, staff, and healthcare facilities (422, 260, and 342 grams/cover, respectively).

The composition of the pre-adoption waste is: unsold prepared foods (40%), raw inventory spoilage (8%), cooking & handling errors (7%), plate waste (10%), trimmings (14%), and uncategorized (21%), see Figure 5.2. That is, on average, 48% of the total daily pre-adoption waste at an individual site is due to leftover inventory (prepared or raw).

6. Empirical Strategy and Identification

6.1. **Identification Challenges and Method Selection.** To establish the causal effects of Winnow treatment based on our observational data, we have to overcome the following identification challenges. First, every site in our data in each time period is either treated or untreated, and *the counterfactual outcome in the alternate condition is not observed.* Second, sites adopted Winnow at

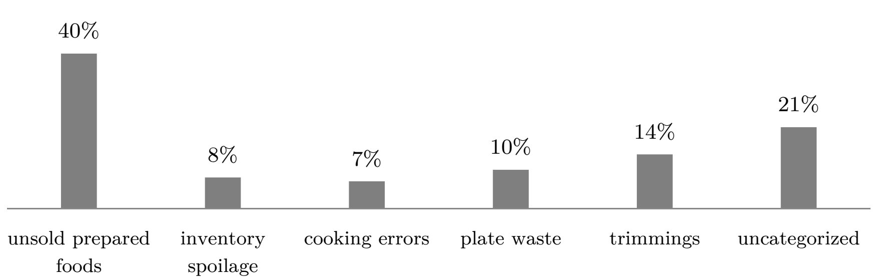


FIGURE 5.2. Pre-adoption Waste Breakdown

different times. Third, the decision of when to adopt Winnow might be endogenous (adoption may be chosen at the time when the highest gain is expected); this would bias our estimates upwards.

To address these identification challenges, we explored several methods. First, we naturally considered using a Difference-in-differences (DiD) estimator. As a *quasi-experimental* method, DiD assesses the impact of an intervention in the absence of exogenous variation by setting up comparison groups (control/treatment group) and measuring the change in an outcome between a pre-and post-intervention period when only one of the groups has access to the intervention (Bertrand et al. 2004), thus addressing challenge (1). It also allows for variation in treatment timing (see de Chaisemartin and D'Haultfoeuille 2020, Sun and Abraham 2021, Goodman-Bacon 2021, Callaway and Sant'Anna 2021), enabling us to resolve challenge (2). However, the parallel trends assumption that implies the difference between the treatment and control group is constant over time in the absence of treatment, is crucial for identification. It is equivalent to assuming that the inclusion of time-specific and unit-specific fixed effects in the model completely controls for all confounding relationships between the treatment and the outcome, other than the desired causal treatment effect. Data from observational studies, unfortunately, *rarely exhibit parallel trends* for the treatment and control group, thus leaving challenges (1) and (2) unaddressed.

To alleviate this concern, we consider more recent methods that seek to compensate for the lack of parallel trends. The Synthetic Control (SC) method "recovers" parallel trends by finding a weighted combination of untreated units such that the pre-treatment trends of this weighted combination of the untreated units are matched with that of treated units. It was first developed for a single treated unit (see Abadie 2005), and, in recent years, has been extended to a staggered rollout setting (see Ben-Michael et al. 2022). Synthetic difference-in-differences (SynthDiD), is a doubly robust estimation method that combines attractive features of both SC and DiD methods (Arkhangelsky et al. 2021). Like SC, this method reweighs and matches pre-treatment trends. Like DiD, it is invariant to additive unit-level shifts in outcomes and allows for inference with large panels even when the pre-treatment period is short. Intuitively, SynthDiD reweighs the unexposed control units to make their time trend parallel (but not necessarily identical) to that of the treated in the pre-intervention, and we then apply a DiD estimator to this reweighed panel (thus addressing challenges 1 and 2).

In addition, when it comes to challenge (3) of endogenous adoption timing, SynthDiD overcomes this as well. SynthDiD is consistent even in the presence of an unobserved correlation between treatment assignment and site-level time trends. Our concerns about timing endogeneity were further alleviated as we worked with the Winnow team to roll out the device in a deliberately random way by ensuring that adoption/upgrades did not start with more tech-savvy sites, sites with early adopter managers, or sites in more environmentally conscious locations.

In sum, the synthetic difference-in-differences method allows us to resolve the identification challenges and establish a causal effect of the adoption/upgrade of Winnow. We use it as our main empirical strategy throughout the paper.

SynthDiD allows us to overcome identification challenges and provides a causal estimate of the Winnow system's effect on kitchens that adopt it. That is, we are able to causally establish the average treatment effect on the treated (ATT). When interpreting our estimates as an average treatment effect (ATE), caution should be exercised. In any setting where all the units are treated eventually, such as ours (all sites eventually adopted Winnow), the adopters may differ from the unobserved non-adopters. This might cause a *selection bias* when interpreting our estimates as ATE (i.e., generalizing to the entire population). We first note that selection bias concerning an earlier selection into the treatment of the higher-expected-benefit sites is directly handled by SynthDiD (as we discuss in the solution to identification challenge 3).

Further, in line with the literature on technology adoption (see, for example, Berman and Israeli 2022), we argue that the average treatment effect on the treated (ATT) is a relevant and appropriate measure in our case. First, currently, there are no reliable estimates of the effect of this technology on food waste in commercial kitchens, even for the adopters. Moreover, because Winnow provides simple tracking/descriptive feedback, benefiting from it requires active engagement with the data collected and consequent data-driven decision-making (i.e., just like the adoption decision, these actions are all endogenous as well). That is, even if sites were randomly assigned to adopt the system, we do not expect to see any benefits without interest in engaging with the systems. As a result, we expect that similar sites that are interested in the adoption of Winnow would see similar outcomes and benefits in food waste reduction as those in our sample. In other words, we provide a first-of-a-kind benchmark on the benefits an adopting site might expect for the firms interested in employing such technology.

To help practitioners obtain a more precise benchmark of the effects, Section 7.3 explores effect heterogeneity with regard to various site characteristics. This is especially helpful because endogenous adoption choice might yield a higher prevalence in our dataset of sites that achieve higher benefits (firms that expect to obtain higher benefits might adopt with higher likelihood), resulting in higher *average* estimates. Our heterogeneity analysis helps sites pinpoint a better site-specific benchmark as per their specific characteristics.

6.2. **Synthetic Difference-in-Differences (SynthDiD).** Synthetic difference-in-differences was originally designed for a balanced panel of units where the treatment timing is identical for all treated units. Like the most recent studies that allow for a staggered rollout design (see de Chaisemartin and D'Haultfoeuille 2020, Callaway and Sant'Anna 2021, Ben-Michael et al. 2022), we perform SynthDiD estimation with a staggered rollout. Following the literature (Callaway and Sant'Anna 2021, Berman and Israeli 2022), we estimate cohort-level treatment effects and then aggregate them into an overall estimate. To perform the analysis, for each adoption cohort r, we construct a balanced panel in which the treatment group comprises sites that adopted Winnow[4] in period r and have outcome data available l_{min} (< 0) periods before adoption and l_{max} periods after, and the control group comprises sites that have data available for the same time frame but adopted Winnow at least l_{max} periods after cohort r. If we denote by N_r, the set of units in the balanced panel of the cohort r, by N_r^{co}, the set of units in the control group, and by N_r^{tr} the set of units in the treatment group, then for each cohort r, the SynthDiD estimation procedure solves

$$(\hat{\tau}_r, \hat{\alpha}_0, \hat{\alpha}_i, \hat{\gamma}_t) = \underset{\tau_r, \alpha_0, \alpha_i, \gamma_t}{\arg\min} \{ \sum_{i \in N_r} \sum_{t=r+l_{min}}^{r+l_{max}} (\log(Y_{it}+1) - \alpha_0 - \alpha_i - \gamma_t - AfterAdopt_{it} \cdot \tau_r)^2 \hat{\omega}_i \hat{\lambda}_t \}, \quad (6.1)$$

where Y_{it} is the food waste outcomes for site i in period t, α_0 is an intercept, α_i is unit- and γ_t is the time fixed-effect, and $AfterAdopt_{it}$ indicates whether site i adopted Winnow by time period t. Standard errors for each $\hat{\tau}_r$ are estimated using the jackknife method (algorithm 3 of Arkhangelsky et al. 2021), or the placebo method (algorithm 4 of Arkhangelsky et al. 2021) if a cohort has only one treated unit. The coefficient τ_r measures the average change in waste outcomes within the $l_{max}+1$ periods after the adoption of cohort r. Equation 6.1 estimates a two-way-fixed-effect model with the addition of unit-specific weights $\hat{\omega}_i$ and time-specific weights $\hat{\lambda}_t$. The unit weights $\hat{\omega}_i$ are selected such that pre-treatment control outcomes weighted by $\hat{\omega}_i$ have a similar trend to that of the average outcomes of the treated units, that is, for all time periods $t < r$, we have

$$\hat{\omega}_0 + \sum_{i \in N_r^{co}} \hat{\omega}_i \log(Y_{it}+1) \approx \frac{\sum_{i \in N_r^{tr}} \log(Y_{it}+1)}{|N_r^{tr}|}.$$

The time weights $\hat{\lambda}_t$ are designed so that the weighted average of historical outcomes predicts average treatment period outcomes for the same control units, up to a constant, that is, for all $i \in N_r^{co}$, we have

[4]Recall, the adoption of Winnow starts for a site once its blind period ends and the recorded waste data are accessible to the site.

$$\hat{\lambda}_0 + \sum_{t=r+l_{min}}^{r-1} \hat{\lambda}_t \log(Y_{it}+1) \approx \frac{\sum_{t=r}^{r+l_{max}} \log(Y_{it}+1)}{l_{max}+1}.$$

The unit weights $\hat{\omega}_i$ serve the same role as in the standard synthetic control method to align the pre-treatment trends in the outcomes of treated and control units. The time weights $\hat{\lambda}_t$ balance pre-treatment time periods with post-treatment ones. That is, if a specific pre-treatment period is more predictive of post-treatment outcomes, it receives a higher weight.

Given these cohort-level estimators $\hat{\tau}_r$, we can compute an overall treatment effect as a weighted average. The weights are chosen to be the proportion of treated units that belong to each cohort. As such, the staggered treatment timing SynthDiD estimator can be formulated as $\hat{\tau} = \sum_r \mu_r \hat{\tau}_r$, where $\mu_r = \frac{N_r^{tr}}{\sum_r N_r^{tr}}$ is the weight for cohort r. Practically, this estimator is simply a weighted average of cohort-specific estimated average treatment effects, where the weight applied to any individual cohort's specific estimator is equal to the proportion of treatment group observations that originate in a specific cohort. We utilize the property of influence functions for summary parameter estimators to compute the standard error of $\hat{\tau}$ (see Appendix A.2 for more technical details).

7. Impact of the automated AI-powered waste monitoring system

This section provides our estimates of the waste reductions due to (1) the adoption of a waste monitoring system and (2) the upgrade to AI-powered waste classification.

7.1. **Effect of Winnow Adoption on Total Waste.** We start by measuring the food waste reductions due to the adoption of Winnow Classic via SynthDiD with a staggered rollout. We use the later adopters' blind period to create a synthetic control for each site. During the blind period, food waste levels for the site are being recorded, but sites have no interaction with the system. When the blind period starts and how long it lasts varies among sites (two weeks to four months). To perform the analysis for each adoption cohort r, we first construct a balanced panel in which the treatment group comprises sites that adopted Winnow in period r and have waste outcome data available 15 days before adoption and 90 days after (i.e. $l_{min} = -15$, and $l_{max} = 90$), and the control group comprises sites that have data available for the same time frame, but that adopted Winnow after more than l_{max} periods after cohort r.

We drop any cohort r, for which the donor set used for the construction of the synthetic control (i.e., the set of units in the control group) is empty. As a result, we obtained 79 cohorts involving 178 sites that adopted Winnow at different times between May 2016 and Nov 2019.[5] We conduct

[5]We end up with fewer sites for measuring causal effect of Winnow adoption, because for several sites a synthetic control cannot be constructed. We provide the summary statistics of the subsample in Appendix A.1. The sites from the subsample are mostly the same as those from the whole sample, but are slightly bigger on average.

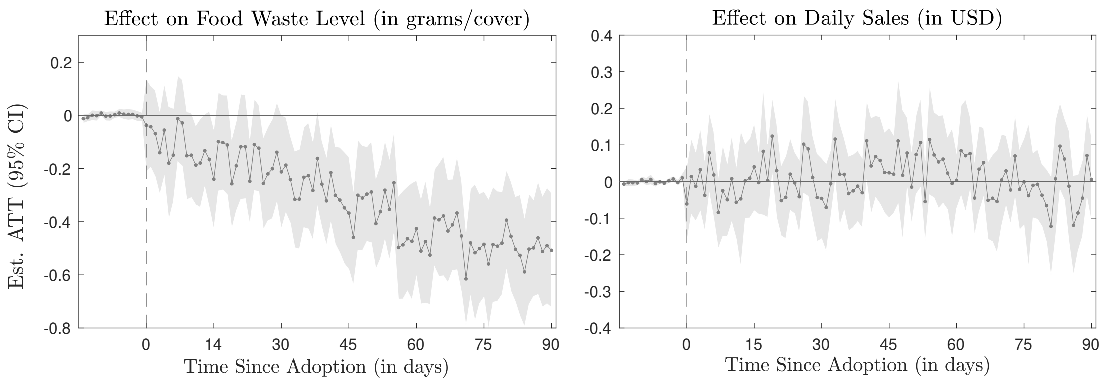


Figure 7.1. Event-study Plots: Staggered Treatment of Winnow Adoption

Effect	Outcome Variable: log(waste_per_cover)		
	1 month	2 months	3 months
	(1)	(2)	(3)
After adoption	-0.1337	-0.2673**	-0.3459***
	(0.0844)	(0.1196)	(0.1288)
†Reduction (%)	12.51%	23.46%	29.24%

Notes. The table shows N-month ATT estimates from synthetic DiD model. Row (†) reports waste reductions (%) for the N months post-adoption on average, which are transformed from each ATT estimate. Significance level: 10% (*); 5% (**); 1% (***).

Table 2. Average Effects for the First $\{1, 2, 3\}$-months Post-Adoption

synthetic DiD analysis within each cohort and aggregate cohort-level ATT estimates into an overall ATT estimate.

Figure 7.1 presents the dynamic event-study treatment effect estimates in each specific period obtained via the SynthDiD method, along with their 95% confidence intervals (CIs). Appendix A.3 describes the methodology used to construct these event-study plots.

First, note that the SynthDiD method produces a nice pre-treatment fit between the trends of the treated and the untreated. The post-adoption treatment effect estimates in Figure 7.1 (left) show that food waste level significantly decreases within the first three months after Winnow adoption. To make sure this waste reduction does not happen at the expense of sales, we run the same analysis but using the sales data instead. Figure 7.1 depicts our findings: while we do see a clear downward trend in waste (left), there is no such trend in sales (right).

Table 2 presents our estimates for the $\{1, 2, 3\}$-month[6] SynthDiD ATTs, with standard errors in parentheses. Table 2 columns (1)-(3) present the average effects for the first month, the first two

[6]We set $l_{max} = \{30, 60, 90\}$ in Equation 6.1 for estimating $\{1, 2, 3\}$-month ATT, respectively.

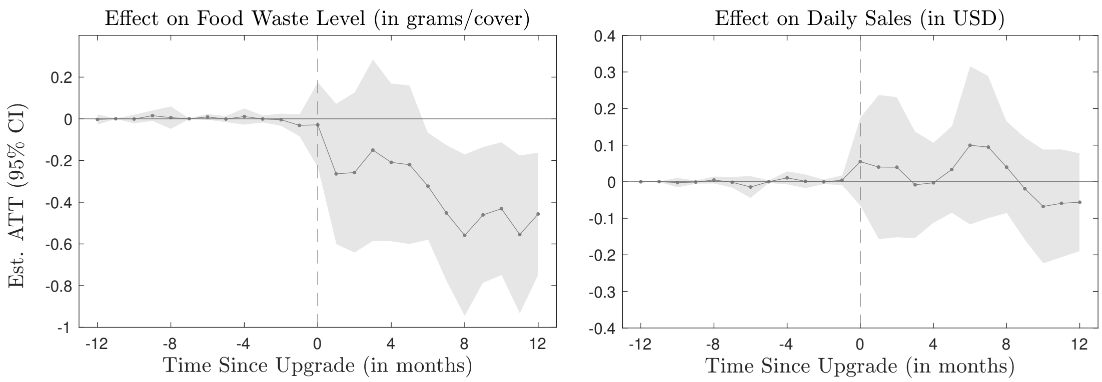


FIGURE 7.2. Event Study: Vision Upgrade Treatment

months, and the first three months post-adoption, respectively. The 2-month-ATT is estimated to be at -0.2673 (95% CI: -0.5017, -0.0329), and the 3-month ATT is valued at -0.3459 (95% CI: -0.5983, -0.0935).

Since the outcome variables are logged, an N-month ATT estimate x indicates an average decrease of $(1-\exp(x))*100\%$ in food waste level for those Winnow adopters within N-months post-adoption. That is, we see about a 23% decrease in daily food waste two months after adoption and a 29% decrease three months after adoption! To put this in perspective, the food waste reduction goal set in 2016-2020 by many nations, NGOs, and private sector firms is to reduce the waste in half by 2030 (in line with the United Nations' Sustainable Development Goal Target 12.3, United Nations 2021). For example, Hilton, Four Seasons, Sodexo, IKEA, Compass Group, and Chipotle all committed to reducing food waste by 50% by the end of 2030 (Hanson 2016, Hilton 2020, Four Seasons 2022, Compass Group 2023). That is, it is expected that it would take almost a decade to meet this goal. By adopting Winnow Classic commercial kitchens in our sample, on average, achieved a 29% waste reduction in a mere three months! As we will show later, further gains are possible with Winnow Vision; the 50% goal can likely be achieved within a year.

7.2. **Effect of Vision Upgrade on Total Waste.** The second key event in our study is a staggered upgrade of the installed Winnow Classic systems to the more sophisticated computer-vision-powered system, creating a quasi-experimental setting for estimating waste reductions on account of the upgrade to an AI-powered waste monitoring system. As before, we conduct SynthDiD analysis but now on the waste sample before and after the *vision upgrade*. The pre-upgrade lags ($l_{min} = -12$ months) and post-upgrade leads ($l_{max} = 12$ months) are used to create the panel. Figure 7.2 (a) presents the corresponding event study plot. It shows the pre-treatment fit between the treatment and control units as well as the evolution of the treatment effect in each time period. The SynthDiD method matches the treatment and control units such that the difference between them is indistinguishable from zero for all pre-treatment periods.

Table 3 presents the estimated N-month SynthDiD ATTs (standard errors are in parentheses). Columns (1)-(4) of Table 3 present the average effects for the first 3, 6, 9, and 12 months post-adoption, respectively. The 9-month's ATT is -0.2964 (95% CI: -0.5275, -0.0653), the 12-month ATT is -0.3570 (95% CI: -0.6175, -0.0965). These negative estimates suggest there is a significant additional effect of bringing AI into commercial kitchens on food waste reduction. On average, 12-month months post-adoption, an upgrade to the computer vision based Winnow Vision yields a further 30% waste reduction.

7.3. **Heterogeneous Treatment Effects: Winnow Classic.** We next explore whether the waste reduction outcomes differ across sites. In particular, we examine whether the estimated (period/site) treatment effects of Winnow Classic $\hat{\tau}_{it}$ (as described in Section A.4) differ based on site size, service type, and demand variability.

The Winnow Adoption column of Table 4 reports our estimates. Here indicator variable $Small\_size = 1$ for sites serving less than 800 covers per day and average. $Large\_size = 1$ for sites serving above 800 covers (the median value in our sample). $Buffet = 1$ if the site offers buffet service and $Table = 1$ if the site offers table service. When the site demand's coefficient of variation is below 0.5, we set the indicator variable $Low = 1$; when it is in the 0.5 to 1 range, $Medium = 1$, and when it is above 1, we set $High = 1$. To control for changes in food waste over time, we include a time trend ($Log(usage\_time)$). As expected, we see the negative and significant coefficient on $Log(use\_time)$ consistent with our findings from Section 7.1, which indicate a greater effect over time.

Effect	Outcome Variable: log(waste_per_cover)			
	3 months	6 months	9 months	12 months
	(1)	(2)	(3)	(4)
After upgrade	-0.1397	-0.2017	-0.2964**	-0.3570***
	(0.1357)	(0.1546)	(0.1179)	(0.1329)
†Reduction (%)	13.04%	18.27%	25.65%	30.02%

Notes. The table shows N-month ATT estimates from synthetic DiD model. Row (†) reports further waste reductions (%) for the N months post-upgrade on average, which are transformed from each ATT estimate. Significance level: 10% (*); 5% (**); 1% (***).

Table 3. Average Effects for the First N Months Post-upgrade

Site-Size. The negative and significant coefficient on $Small\_size$ indicates that a greater Winnow effect is associated with smaller food service sites than big ones. Three months post-adoption, smaller sites, on average, enjoy about a 20% higher reduction in food waste than bigger sites (42% vs 22%). This more pronounced effect for smaller establishments may happen because the larger sites enjoy the benefits of pooling and, as a result, experience less waste per meal served. The larger sites might also be able to better optimize their operations.

Service-Type. The negative and significant coefficient on the indicator variable $Buffet$ implies that buffet sites gain more benefits of waste reduction from Winnow adoption than table sites. Three months post-adoption, buffets, on average, enjoy a 10% higher reduction compared to table sites (36% vs 26%). This more pronounced effect for buffets is likely due to the higher value of the feedback in the make-to-forecast production systems.

Demand Variability. Because there are different mechanisms that might be in effect on the low and high sides of the variability spectrum (see Section 4.3) we divide sites into three groups (low, medium, and high). The negative and significant coefficients on the indicator variables Low and $High$ indicate that sites with low or high demand variability both benefit more from Winnow than sites with medium demand variability. High variability sites have more to gain from the feedback, while low variability sites can be more precise in the use of the signal. Three months post-adoption,

Individual Treatment Effects	Winnow Adoption	Vision Upgrade
Constant	0.6241***	0.5162
	(0.0490)	(0.5131)
Log(usage_time)	-0.1539***	-0.1958**
	(0.0103)	(0.0781)
Indicators for site size		
Small_size (=1)	-0.2206***	-0.4411***
	(0.0230)	(0.1273)
Indicators for service type		
Buffet (=1)	-0.1034***	-0.3724**
	(0.0235)	(0.1863)
Indicators for demand variability		
Low (=1)	-0.1897***	-0.2643
	(0.0307)	(0.4951)
High (=1)	-0.2165***	
	(0.0673)	
N_observations	13,713	275
N_site	177	22

*** -- 1% level, ** -- 5% level, * -- 10% level

Large_size is dropped as the reference level for site size, Table for service type, and Medium for demand variability.

Table 4. Heterogeneity in the Effect of Winnow Adoption and Vision Upgrade

on average, high-variability sites experience 45% reduction vs. 25% for medium variability sites and 42% for low variability sites.

Combinations of site characteristics. For ease of reference, Table 5 provides the breakdown of the adoption effect (3 months post-adoption) for various size-service type-demand variability combinations along with their average pre-adoption waste levels. The least favorable combination of site characteristics, large site with table service and medium demand variability sees, on average, a 7% food waste reduction three months post-adoption. While small, high-variability, buffets benefit the most, enjoying, on average, a 46% reduction.

7.4. **Heterogeneous Treatment Effects: Winnow Vision.** We also explore the heterogeneity in the effect of the Winnow Vision upgrade (see Table 4, Vision upgrade column).

Contrary to our findings for the Winnow Adoption effect, we don't see a significant difference across sites with different demand variability levels. This might be because (as we show in Section 8.1) the additional waste events captured by the vision system are largely straightforward inventory and cooking errors that have less to do with managerial responses to demand changes.

Similar to the effect of Winnow Adoption, smaller sites and buffets also enjoy the biggest benefits from upgrading to vision. One year post upgrade, smaller sites experience a higher additional reduction compared to larger sites (57% vs. 21%). Buffets experience a higher additional reduction compared to table sites (55% vs 24%). Small buffets see the highest additional gains (about 65% on average), followed by small table services (54%), large buffets (46%), and large table sites (21%).

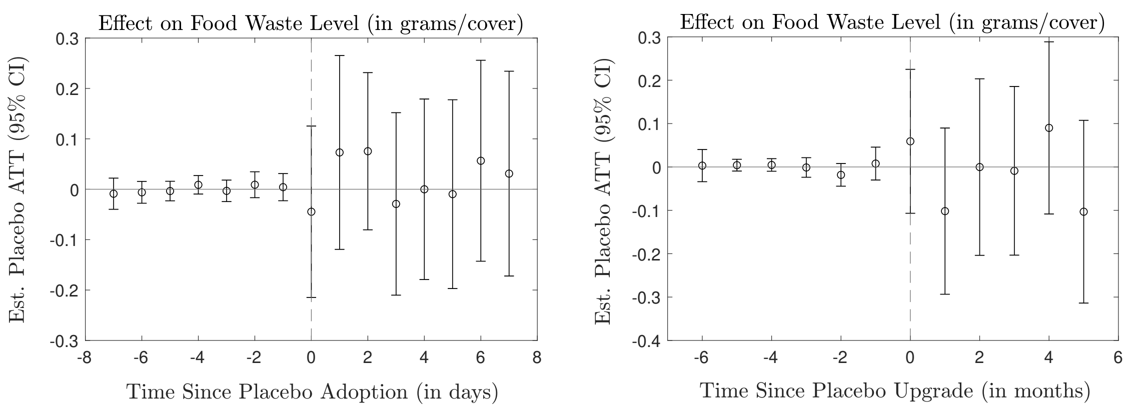
7.5. **Validation and Robustness.** We conduct several additional analyses to examine the validity and robustness of our findings.

		Small Site Size		Large Site Size	
Service type	Demand variability (Coeff. of variation)	Avg pre-adoption waste (grams/cover)	Adoption effect (% reduction)	Avg pre-adoption waste (grams/cover)	Adoption effect (% reduction)
Buffet	Low	1724	44%	1185	30%
	Medium	2207	32%	1234	16%
	High	4371	46%	2297	32%
Table	Low	1188	38%	384	23%
	Medium	1893	25%	415	7%
	High	1917	40%	803	25%

TABLE 5. Winnow Adoption Treatment Effects Heterogeneity

7.5.1. *Placebo Checks.* Following the literature (Abadie et al. 2010, Ferman and Pinto 2021, Athey and Imbens 2022), we assess the validity of our empirical approach by conducting several placebo tests. In these tests, we investigate whether we observe the effects of Winnow adoption or vision upgrade in the population that was unexposed to the treatment.

Placebo Checks in Time. Our first placebo test shifts the analysis sample back in time to the pre-treatment period. We change the treatment time (i.e., when to activate Winnow or when to upgrade to vision systems in our setting) to an earlier date and re-run the analysis considering the period from the new treatment time to the old treatment as the placebo post-treatment period. We expect to verify that the SynthDiD analysis suggests no effect for these redefined post-treatment periods, which are, in fact, all prior to the actual treatment time. The results of this placebo test using the pre-treatment periods are presented in Figure 7.3. As expected, we find no significant effects in those treated sites during either pre/post placebo adoption periods or pre/post placebo upgrade periods.



(a) Winnow Adoption (b) Vision Upgrade

FIGURE 7.3. In-time Placebo Tests

Placebo Checks in Units. Our second placebo test is conducted by reassigning treatment status to the control units (i.e., sites that did not upgrade to computer-vision-based systems in our setting), in turn, to estimate a distribution of placebo effects. This test allows us to further investigate the likelihood that we might encounter a similarly sized treatment effect just by chance. If the estimated treatment effect is unusually extreme compared to the distribution of placebo effects (i.e., via two-sided tests), then the estimated treatment effect is considered significant. Specifically, we iteratively assign the treatment to the same number of sites as occurred under the vision upgrade (i.e., 42 sites upgraded their Winnow systems) and estimate placebo effects in each iteration from the SynthDiD model. The results are presented in Figure 7.4, with the vertical lines indicating the 5th and 95th percentiles of this placebo distribution. Our estimates for the vision upgrade (depicted with the dashed line) fall well below the 5th percentiles of the distributions of placebo coefficient estimates and t-statistics, further increasing our confidence that we are estimating a true effect.

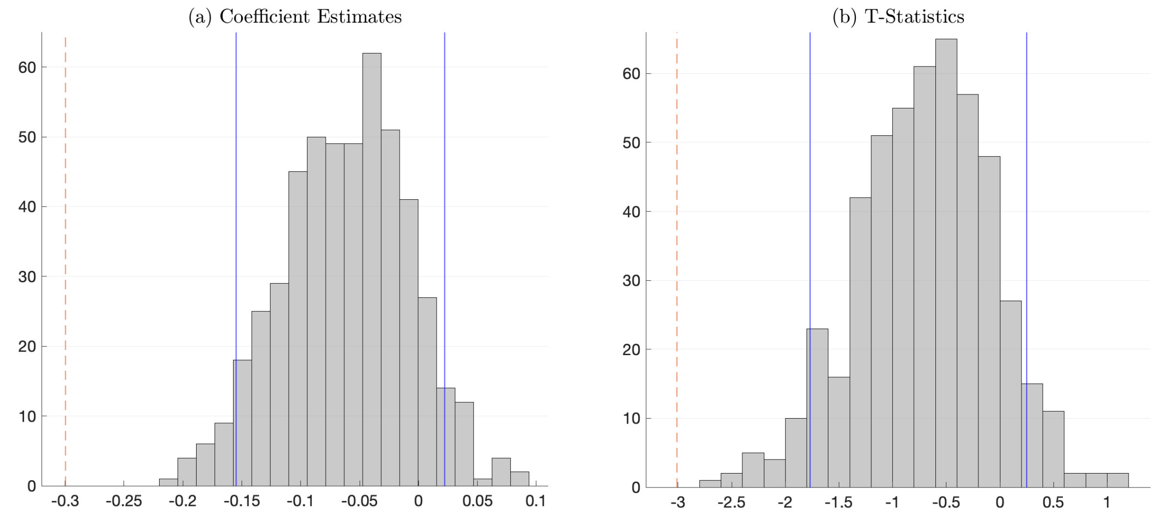


FIGURE 7.4. Distributions of Vision Upgrade Effect Estimates and T-Statistics
Notes. These two plots present the distributions of coefficient estimates and t-statistics generated from the 500 placebo simulations using the food waste data for sites that did not upgrade. The 5th and 95th percentiles are marked with solid vertical lines, while the magnitude of our estimate is depicted with a dashed line.

7.5.2. *Upgrade Selection.* As we discussed in Section 6, food service sites that expect higher benefits might choose to upgrade their classic system to Vision with a higher probability. While the Synth DiD method itself alleviates this concern, we further perform the following test. We estimate a Probit/Logit model to examine what type of sites were more likely to upgrade their Winnow system. Our random effects Probit model includes three key predictors: food waste level, budget size, and the number of kitchen staff. Intuitively, we may expect that richer and labor-intensive sites that face higher levels of food waste should be more likely to upgrade and deploy the AI-powered tool to track food waste. However, the coefficients for all the predictors in our estimated Probit model are statistically insignificant. In other words, we are unable to detect a particular selection pattern. The details of the estimation procedure and results are provided in Appendix A.5.

7.5.3. *Alternative Methods.* We also identify treatment effects using conventional methods, such as the staggered difference-in-differences (SDiD) estimator (Appendix A.6). Like Wang and Goldfarb (2017), we identify the causal effect by comparing the change in the outcomes before and after adoption for adopting sites with the change in outcomes in the same time periods for sites that have not yet adopted Winnow. As expected, the pretreatment fit between the trends of the treated and untreated sites produced by staggered DiD is not as good as that from synthetic DiD (Appendix A.6, Figure A.1). The treatment effect estimates after adoption demonstrate similar trends, as we see in our main results, though they are a bit larger in magnitude than those obtained with synthetic DiD. For example, compared with the main result that there is an average decrease of 29% in daily food waste three months after Winnow adoption, the estimates from staggered DiD indicate an average decrease of 33%.

8. Mechanisms

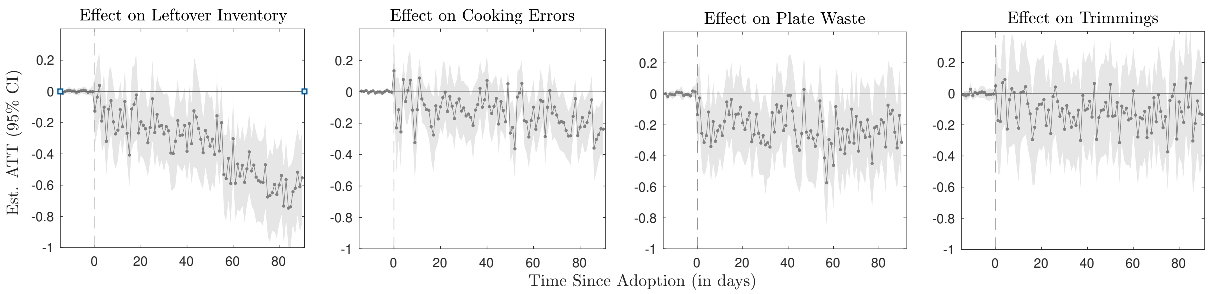
The preceding analyses identified large, significant effects of digital technology and AI-enabled waste data collection. In this section, we explore the potential mechanisms through which such systems operate to drive reductions in food waste.

8.1. **Effect by Type of Food Waste.** To unpack the origins of the waste reductions observed, we repeat our analysis separately on different types of food waste: leftover inventory,[7] waste due to cooking & handling errors, plate waste, and trimmings (as described in Section 5).

Specifically, we use our synthetic DiD approach to analyze how different types of waste change before and after their exposure to the Winnow treatments (either adoption or the AI upgrade). The application of the method is as described in Section 6.2; the only difference is that the dependent variable now becomes the amount of different types of food waste (in grams/cover) during the observation window.

Adoption Effect by Food Waste Type. The four plots in Figure 8.1 (A) show the pretreatment fit of SynthDiD between the adopted and the not-yet-adopted as well as the evolution of the Winnow adoption effect by waste type over time; here, time 0 indicates the period of adoption and other times are relative to adoption. Figure 8.1 (B) reports the estimated average effects of Winnow adoption on different types of waste (3 months post-adoption). For all types of waste, we observe that the estimated ATTs of adoption are negative. That is, all types of waste seem to experience reductions. We observe that overall waste reductions are primarily driven by a significant decrease in the leftover inventory. The leftover inventory decreases on average by about 32% (3 months post-adoption). The reductions in other types of waste are not statistically significant, perhaps due to limited statistical power. We note an average reduction of 19% in cooking errors, 16% in plate

[7]This is the sum of expired inventory and unsold prepared foods, items that can not be used or sold, due to expiration or lack of demand,



(A) SynthDiD Treatment Effects on Various Types of Waste

	Outcome Variable: log(waste_per_cover)				
	Food Waste Breakdown				
	Leftover Inventory	Cooking & Handling Errors	Plate Waste	Trimmings	Total
After Adoption	-0.3805**	-0.2081	-0.1781	-0.1683	-0.3459***
	(0.1831)	(0.1982)	(0.7107)	(0.5429)	(0.1288)
†Reduction (%)	31.65%	18.79%	16.31%	15.49%	29.24%

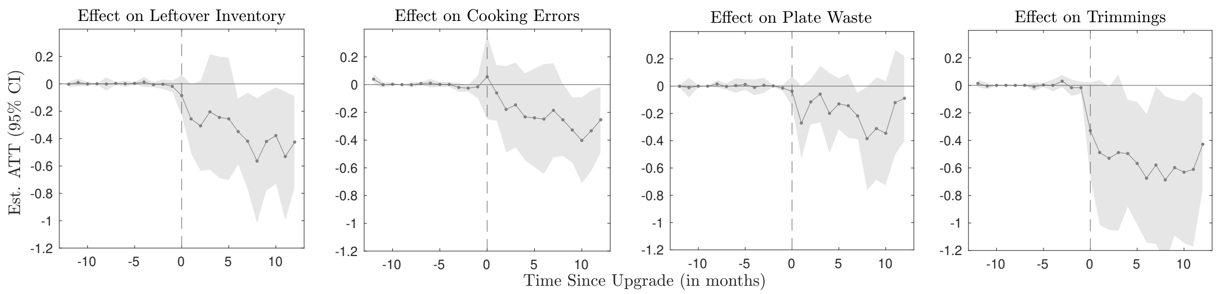
Notes. The table shows 3-month ATT estimates for different types of waste via synthetic DiD. Row (†) reports waste reductions (%) for the 3 months post-adoption on average, which are transformed from each ATT estimate. Significance level: 10% (*); 5% (**); 1% (***).

(B) Average Effects on Various Types of Waste (3 Months Post-Adoption)

FIGURE 8.1. Adoption Effect by Food Waste Type

waste, and 15% in trimmings. We suspect that in addition to better production planning, cooking and handling errors are also reduced by access to better data in the kitchen. Further, the 16% plate waste reduction suggests that chefs are able to take some actions to reduce this consumer-driven portion of food waste as well (possibly by reducing portions, changing plate sizes, etc.). Finally, it is natural to expect that fewer trimmings are created in the preparation stage if there is less overproduction and cooking and handling errors.

Vision Upgrade Effect by Food Waste Type. The four plots in Figure 8.2 demonstrate the pretreatment fit of SynthDiD between the upgraded and the not-yet-upgraded as well as the evolution of vision upgrade effect by waste type over time. Here, time 0 indicates the period of upgrade, and other times are relative to upgrade. We observe that the estimated vision upgrade ATTs for all types of waste are negative; all types of food waste are reduced after the upgrade. The table in Figure 8.2 also reports the average effects of vision upgrade on various types of waste a year post-upgrade. We document a further statistically significant decrease not only in the leftover inventory (30%) but also in the cooking errors (21%) and trimmings (43%) on account of upgrading



(A) SynthDiD Treatment Effects on Various Types of Waste

	Outcome Variable: log(waste_per_cover)				
	Food Waste Breakdown				
	Leftover Inventories	Cooking & Handling Errors	Plate Waste	Trimmings	Total
After Upgrade	-0.3546**	-0.2330*	-0.1931	-0.5646**	-0.3570***
	(0.1789)	(0.1220)	(0.1571)	(0.2766)	(0.1329)
†Reduction (%)	29.85%	20.78%	17.56%	43.14%	30.02%

Notes. The table shows 12-month ATT estimates for different types of waste via synthetic DiD. Row (†) reports further reductions (%) for the 12 months post-upgrade on average, which are transformed from each ATT estimate. Significance level: 10% (*); 5% (**); 1% (***).

(B) Average Effects on Various Types of Waste (12 Months Post-Upgrade)

FIGURE 8.2. Vision Upgrade Effect by Food Waste Type

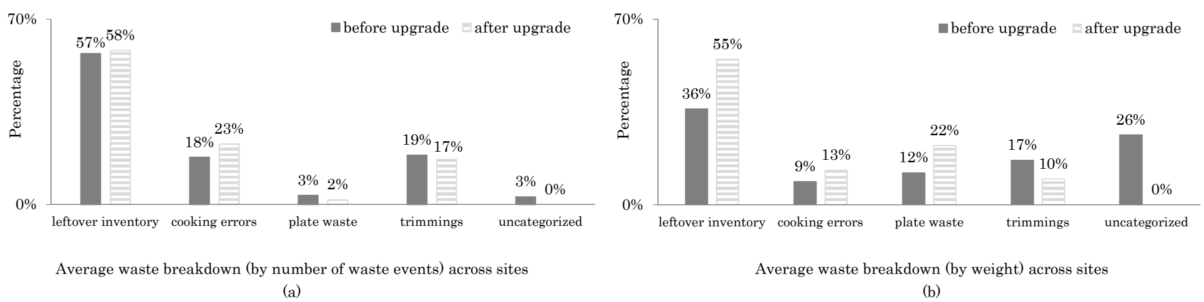


FIGURE 8.3. Total Waste Breakdown 3 Months Before and After Vision Upgrade

to the computer-vision-based system. We also see an 18% reduction in plate waste, although this reduction is not statistically significant.

We next examine the total waste breakdown before and after the vision upgrade for the upgraded sites, Figure 8.3. We observe that before the upgrade, on average, 3% of food waste events recorded were uncategorized (Figure 8.3(a)). Just these 3% of food waste events, however, accounted for about 26% of the total waste amount (by weight) (see Figure 8.3(b)). The virtually complete categorization by the vision system (Figure 8.3 a and b) prevents such large waste events from being hidden, be it massive overproduction/over-purchase or cooking errors. The ability to see and control these large waste events drives the additional food waste reduction in leftover inventory and cooking errors, and, as a result, the corresponding reduction in trimmings.

8.2. **Correction of Behavioral Biases.** Managerial biases could lead to suboptimal actions in production and inventory management. We particularly look at those behaviors that could cause higher-than-necessary food waste in the kitchen. Specifically, we explore the three main biases identified by the behavioral operations management literature:

- ▷ static production plan (i.e., relying on simple heuristics or rules of thumb),
- ▷ demand chasing (i.e., adjusting the production quantity towards the demand in the prior period),
- ▷ overreaction to demand spikes (i.e., raising/lowering the production quantity in response to the sharp demand increase/decrease in prior periods).[8]

We expect the adoption of Winnow systems to drive better decision-making for kitchen managers. By correcting the behavioral biases mentioned above, Winnow systems could reduce unnecessary food waste. Waste data collected by Winnow systems can also serve as continuous inventory feedback. Though the system does not provide recommendations (predictive or prescriptive solutions), it offers a simple way to monitor and assess the performance of production and inventory decisions, thus enabling users to correct certain wasteful behaviors in the kitchen.

Interestingly, while we cannot directly observe the changes in the site's production or inventory behaviors, we were able to infer the behavior changes by developing a pattern-matching method that uses the collected *waste data alone.* More specifically, we utilize machine learning techniques to detect these behavioral biases and then conduct a DiD-type analysis to determine how the identified incidence of these biases changes after the adoption/upgrade of Winnow. We describe the method and results from the detection of biases in Section 8.2.1, and those for establishing the causal relationship between the changes in the incidence of biases and Winnow adoption/upgrade in Section 8.2.2.

8.2.1. *Bias Detection by Machine Learning Classifiers.* Our approach to detecting biases relies on identifying patterns in the time series of waste data. Note that this problem would be trivial if we had access to the time series of waste and/or sales and production. As described above in our institutional context, as a practical matter, most sites using the Winnow system do not have

[8]Mathematically speaking, demand chasing can be expressed by $Q_t = Q_{t-1} + \alpha(D_{t-1} - Q_{t-1}) + \epsilon_t$ where $\alpha \in (0, 1]$ captures the degree of demand chasing and ϵ_t is the noise term, where Q_t be the order quantity in period t, and D_t be the demand in period t. Overreaction to demand changes can be expressed by $Q_t = Q_{t-1} + \delta(D_{t-1} - D_{t-2})^+ - \gamma(D_{t-2} - D_{t-1})^+ + \epsilon_t$ where $\delta \geq 0$ captures the degree of overreaction to demand spike and $\gamma \geq 0$ captures the degree of overreaction to demand slump.

sales data (especially the biggest culprits–buffets), nor do we have production or work-in-progress inventory data. So, we attempt to detect biases only using waste data.

There is a second challenge— the data streams recorded by Winnow are not labeled with what human mistake a kitchen manager might have made in their food production or the bias in action, and thus, we cannot train pattern-detection algorithms or time-series classifiers based directly on Winnow data; our data is unlabeled. Like in many contemporary AI applications, we instead train our models on synthetic data and validate the trained classifiers in a third-data set. In particular, we follow four steps: (1) we generate labeled synthetic waste data by simulating the dynamics of food production systems under the three behavioral biases. (2) Next, we train a classifier on this data, details below. (3) we validate the performance of this classifier in out-of-sample synthetic data and in lab experimental data from a previous behavior operations study on inventory biases. (4) finally, we use the developed trained classifier to predict the bias in operation in the Winnow data before and after the adoption/upgrade of the Winnow system. These predictions are the input to our subsequent difference-in-difference analysis on the incidence of biases. Steps (1)-(4) are described below.

(1) Generating Synthetic Training Data. Our data generation process is straightforward. We build a perishable-inventory system based on the (Q, r) model (Berk and Gürler 2008) and incorporate three biases into the model. For simulation, we calibrate the data-generating simulation with *real-world product and economic characteristics* (for ex., demand parameters, cost parameters, and product characteristics such as shelf life and cooking time). Appendix A.7 provides the full simulation algorithms.

To obtain the best possible detectors and cleanest comparisons, we identified a subset of 57 kitchens from our dataset that belong to one company, serve exactly the same products in exactly the same formats, and for which we have extensive information on food items served. We use parameters of these kitchens and products (demand parameters, cost parameters, and product characteristics such as shelf life, cooking time, etc.) to generate the synthetic data.

(2) Time-series Classifiers Developed on Synthetic Data. Our detection problem is akin to the time-series classification task, where the training data is a set of time series with class labels, and the goal is to detect the presence of a specific issue automatically. There are many algorithms specially designed for classifying time series. We start with some intuitive distance-based models like the k-nearest neighbors (KNN, Fix and Hodges 1989) algorithm, which classify time series based on similarity among them. However, one big issue is that they are unable to extract information on the relationship between variables. We then explore feature-based algorithms, for example, a time-series forest classifier that adapts random forest (RF, Breiman 2001) classifier to the time series data. They extract discriminative features of time series for classification, whose performance thus relies heavily on the quality of extracted features.[9]

Modern deep-learning algorithms, on the other hand, offer better ways for feature extraction. They automatically learn a hierarchical feature representation from the data that could preserve most of the information content of a time series. One common choice is the convolutional neural networks (CNN, Lecun et al. 1998) algorithm, where the deep layers in the network act as a set of feature extractors that are somewhat generic. A more sophisticated technique has been developed, the long short-term memory fully convolutional networks (LSTM-FCN, Karim et al. 2017) algorithm, which enhances fully convolutional networks (FCN, Long et al. 2014) by adding a long short-term memory (LSTM, Hochreiter and Schmidhuber 1997) block that is able to hold long-term temporal contextual information.

To develop bias detectors, we divided the synthesized waste time series data that are labeled with the respective behavioral biases into training and out-of-sample testing datasets and train time-series classifiers via three representative algorithms including k-nearest neighbors (KNN), random forests (RF), long short-term memory fully convolutional networks (LSTM-FCN). Appendix A.8 provides additional details.

(3) Validation of the Bias-Detection Algorithm. Out-of-sample accuracy is the main performance measure for our classification task using synthesized data. We find that the three biases can be detected from the *waste data alone* with very high accuracy via the deep-learning-based classifiers. For example, for characteristics corresponding to the most frequently wasted item in our institutional context (meatballs), we could detect whether there is a certain bias, e.g., either static production plan, demand chasing, or overreaction to demand spikes, from the synthesized waste time series data with 90% accuracy using the deep-learning-based LSTM-FCN algorithm (by comparison, KNN and RF at lower accuracy, 55% and 75% respectively).

To further validate the translation from our synthetic data to real-world data, we ran our classifier on the experimental data generated in a landmark inventory-biases study Rudi and Drake (2014). In this lab study, the waste data and the bias in operation are both available. We applied our classifier to the waste data alone and our classifier was able to detect the biases with almost full accuracy.

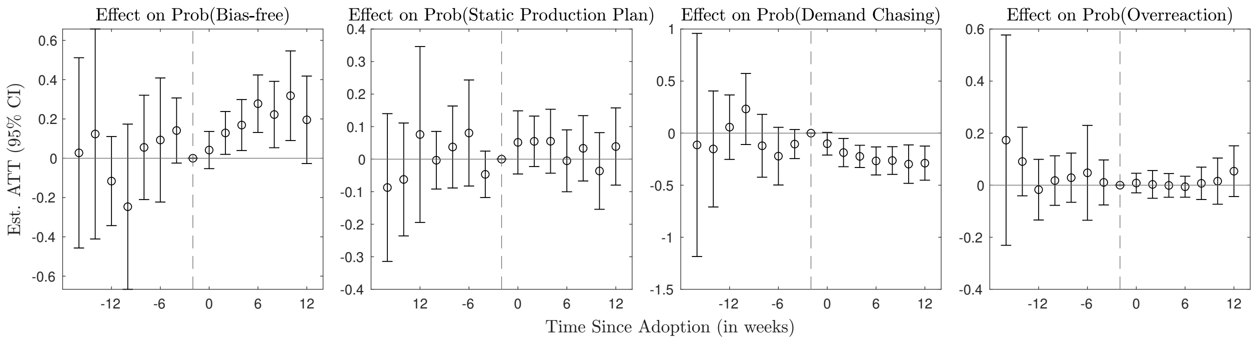
[9] To overcome this issue, there are more advanced feature-based algorithms that enable feature selection, but the effect is limited as the extracted features themselves are generally simple (i.e., so-called hand-crafted features) and cannot fully represent a time series.

(4) Bias Detection at Winnow Sites. We next deployed the validated classifiers to the 57 sites (whose economic characteristics we had used to generate synthetic data) and identified if any behavioral biases were implied by the waste generated at each site. We did this for various time windows. Take the most frequently wasted item, meatball, for example. In total, 28,000 daily observations of meatball leftovers were collected from the 57 sites.

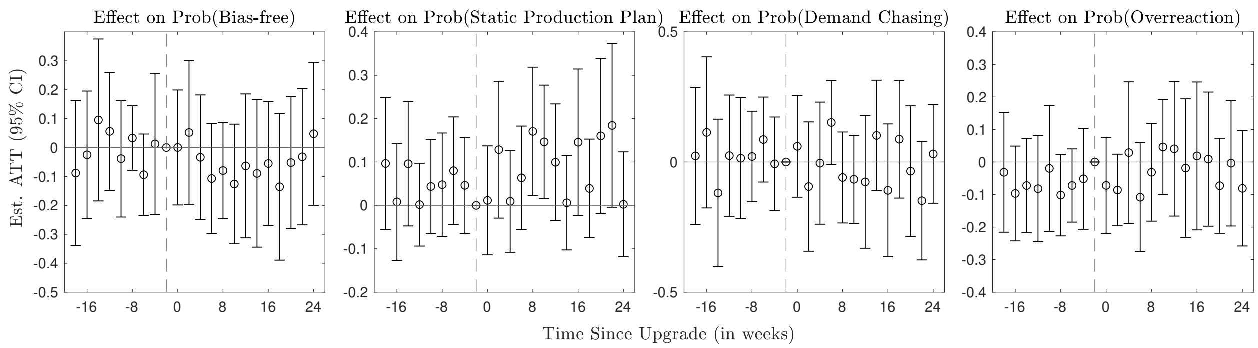
Using our LSTM-FCN multi-class classifier, we find that behavioral biases were rampant before Winnow adoption, particularly demand chasing. The average incidence of the three biases across 57 sites pre-adoption is 62% for demand chasing, 9% for static production plan, and 7% for overreaction to demand spikes. That is, only 22% of time periods were identified as having waste outcomes consistent with optimal/bias-free decision-making. We describe the results of our pre- and post-adoption comparisons next.

8.2.2. *Post-treatment Correction of Behavioral Biases.* To investigate if, post-Winnow adoption, any of the behavioral biases are changed, we investigate the probability of having a certain bias predicted by the LSTM-FCN classifier. The 57 sites adopted/upgraded Winnow in a staggered manner, offering us a quasi-experimental setting where we can apply the DiD technique to identify the causal effect of adopting Winnow Classic and upgrading to Winnow Vision on the detected bias outcomes. This analysis proceeds along the same lines as our main analysis, except we use the staggered DiD estimator, instead of the synthetic DiD estimator. The four plots in Figure 8.4(A), demonstrate the dynamic event study treatment effect estimates, along with their 95% confidence intervals (CIs). Here, time 0 indicates the period of Winnow adoption and other times are relative to the adoption. Before Winnow adoption, we do not observe any statistically significant differences in the predicted probabilities of all three behavioral biases between "treatment" and "control" groups (in our staggered treatment case, the control group constitutes sites that have not yet adopted Winnow by the time). After adoption, we observe a significant reduction in the probability of demand chasing (on average, a 23% reduction three months post-adoption) with a corresponding increase in the probability of bias-free behavior. The other two biases do not experience a significant change.

Among the 57 sites, 19 upgraded the Winnow Classic system to the computer-vision-based Winnow Vision. The four event-study plots in Figure 8.4(B), show the ATT estimates along with their 95% confidence intervals (CIs) over time. Here, time 0 indicates the period of vision upgrade. For the Vision upgraders, we do not find evidence for changes in the incidence of the three biases compared with non-upgraders. Perhaps, this is because the additional events captured by the vision system



(A) Adoption of Winnow Classic



(B) Upgrade to Winnow Vision

FIGURE 8.4. Identified Biases in Operation by LSTM-FCN Model

are largely straightforward errors (inventory and cooking) and have less to do with the managerial response to demand changes (as shown in Sections 7.2 and 7.4).

8.3. **Discussion.** Our waste breakdown analysis suggests that managing leftover inventories is the main driver of waste reductions at commercial kitchens following Winnow's implementation. The reductions may be due to the kitchen managers utilizing waste records as a form of inventory feedback and correcting wasteful behaviors in their food production. Our machine-learning-based behavioral-bias detection analysis provides further evidence. We find that the adoption of Winnow technology decreases the incidence of several common biases in operation. The behavioral changes we document are consistent with our prediction of how kitchen managers act in the presence of waste feedback and contribute to the observed reductions in leftover inventory. Overall, it appears that a benefit of using a waste measuring and monitoring system is that it allows users to evaluate their actions in the kitchen and fine-tune them.

Additionally, after the AI upgrade, we observe a further decrease not only in leftover inventories but also in other types of waste like trimmings, cooking & handling errors. We do not find evidence that supports any further change in the incidence of managerial biases. Overall, these results suggest that the additional benefits of an AI-driven system accrue through better categorization of food waste events and by preventing systematic misreporting of data related to the most egregious behaviors. We utilize machine learning techniques to infer how kitchen managers behave before and after Winnow implementation, only from waste data. There might be a gap between the detected behavioral changes and the ground truth. Given the data availability, this is the best choice we could make. If more detailed data were accessible, for example, actual production decisions, we would be able to directly analyze the changes in kitchen manager's observable actions after Winnow adoption. However, it is unlikely that such data will be available at a large scale any time soon.

9. Conclusion

What gets measured gets managed. Accurate measurement of food waste is important as data informs priorities, policies, and mitigation strategies and helps track progress. Recent technological developments in food waste capture hold promise for creating food waste solutions. Technology companies, such as our industry collaborator Winnow, have launched (AI-powered) granular food waste information gathering systems that can easily measure and stratify food waste in an automated manner, down to individual disposal transactions of single ingredients at different levels of preparation.

In this study, with the unique Winnow data, we provide the first census of the nature of commercial food waste, contributing to a better understanding of what exactly is being wasted, how much, and why. The quasi-experimental setting of the Winnow technology implementation at almost 900 commercial kitchens allows us to apply a synthetic difference-in-differences technique and identify the causal effect of Winnow adoption on food waste. We find that the adoption of Winnow systems reduces food waste, on average, by 29% three months post-adoption. We demonstrate that the effect is substantial and robust using alternative methods. In addition, we estimate that upgrading to the computer-vision-based automatic recognition system induces a further 30% average reduction in food waste level one year post-upgrade. Building on the main effects we document, we also disentangle potential avenues through which Winnow may benefit its users in reducing food waste. Our results reveal that the value of Winnow adoption mainly comes from the reduction in the leftover inventories, while the boost value of AI introduction is due to the ability to see and control additional large waste events that were previously unrecorded. By utilizing machine learning techniques, we further identify common biases in operation from waste data streams and conclude that the correction of behavioral biases in managing food production or inventories drives the value Winnow creates.

Our research has certain limitations, which create exciting opportunities for future research. First, due to data unavailability before the installation of Winnow, we were only able to leverage the limited blind period of those later adopters for constructing synthetic control for each early adopter. The effect of Winnow adoption we study is thus restricted to relatively short-term effects. Collecting historic organic waste volumes from other sources should afford excellent opportunities to better quantify a longer-term effect and validate the Winnow value we have measured. Second, we cannot directly examine how users adjust their production inventory management after receiving waste feedback from Winnow, as the actual production and inventory decisions are not observed in our setting. Detailed records of these decisions would offer future researchers a different way to uncover possible economic and behavioral mechanisms that underlie the identified effects. Third, our data were truncated by the spread of COVID-19. It remains interesting to explore how Winnow works during times of random shocks versus regular times. Finally, Winnow currently does not provide recommendations and leaves users to generate their own insights from the data. The lack of good predictive and prescriptive solutions to perishable inventory problems suggests future research avenues, such as the development of AI-based prescriptive inventory management systems.

References

A. Abadie. Semiparametric difference-in-differences estimators. *Review of Economic Studies*, 72(1): 1–19, 2005.

A. Abadie, A. Diamond, and J. Hainmueller. Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American Statistical Association*, 105(490):493–505, 2010.

A. Akkas. Shelf space selection to control product expiration. *Production and Operations Management*, 28(9):2184–2201, 2019.

A. Akkas and V. Gaur. Om forum-reducing food waste: An operations management research agenda. *Manufacturing & Service Operations Management*, 24(3):1261–1275, 2022.

A. Akkas, V. Gaur, and D. Simchi-Levi. Drivers of product expiration in consumer packaged goods retailing. *Management Science*, 65(5):2179–2195, 2018.

A. Akkas and D. Honhon. Shipment policies for products with fixed shelf lives: Impact on profits and waste. *Manufacturing & Service Operations Management*, 24(3):1611–1629, 2022.

A. Anderson and E. Green. Personal bests as reference points. *PNAS*, 115(8):1772–1776, 2018.

S. Anderson and A. Kimball. Evidence for the feedback role of performance measurement systems. *Management Science*, 65(9):4385–4406, 2019.

D. Arkhangelsky, S. Athey, D. Hirshberg, G. Imbens, and S. Wager. Synthetic difference-in-differences. *American Economic Review*, 111(12):4088–4118, 2021.

E. Astashkina, E. Belavina, and S. Marinesi. The environmental impact of the advent of online grocery retailing. Available at `https://dx.doi.org/10.2139/ssrn.3358664`, 2019.

S. Athey and G. Imbens. Design-based analysis in difference-in-differences settings with staggered adoption. *Journal of Econometrics*, 226(1):62–79, 2022.

V. Athmanathan. Food waste: A rising hurdle in the hospitality industry. Available at `https://idsnext.com/blog/food-waste-a-rising-hurdle-in-the-hospitality-industry/`, 2021.

T. Babina, A. Fedyk, A. He, and J. Hodson. Artificial intelligence, firm growth, and product innovation. *Journal of Financial Economics*, 151:103745, 2024.

A. Bandura. Social cognitive theory of self-regulation. *Organizational Behavior and Human Decision Processes*, 50(2):248–287, 1991.

A. Bandura. Self-efficacy. In I. Weiner and W. Craighead, editors, *The Corsini Encyclopedia of Psychology*, pages 1–3. John Wiley & Sons, Ltd, Hoboken, NJ, 2010.

R. Beer, A. Qi, and I. RÃos. Behavioral externalities of process automation. Available at `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4295527`, 2022.

E. Belavina. Grocery store density and food waste. *Manufacturing & Service Operations Management*, 23(1):1–18, 2021.

E. Belavina, K. Girotra, and A. Kabra. Online grocery retail: Revenue models and environmental impact. *Management Science*, 63(6):1781–1799, 2016.

E. Ben-Michael, A. Feller, and J. Rothstein. Synthetic controls with staggered adoption. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 84(2):351–381, 2022.

E. Berk and U. Gürler. Analysis of the (q, r) inventory model for perishables with positive lead times and lost sales. *Operations Research*, 56(5):1238 – 1246, 2008.

R. Berman and A. Israeli. The value of descriptive analytics: Evidence from online retailers. *Marketing Science*, 41(6):1074–1096, 2022.

E. Bernstein. The transparency paradox: A role for privacy in organizational learning and operational control. *Administrative Science Quarterly*, 57(2):181–216, 2012.

M. Bertrand, E. Duflo, and S. Mullainathan. How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics*, 119(1):249–275, 2004.

S. Blader, C. Gartendberg, and A. Prat. The contingent effect ofmanagement practices. *The Review of Economic Studies*, 87(2):721–749, 2020.

G. Bolton and E. Katok. Learning by doing in the newsvendor problem: A laboratory investigation of the role of experience and feedback. *Manufacturing & Service Operations Management*, 10(3): 519–538, 2008.

K. Borusyak, X. Jaravel, and J. Spiess. Revisiting event study designs: Robust and efficient estimation. Available at `https://arxiv.org/abs/2108.12419`, 2024.

E. Bradlow, M. Gangwar, P. Kopalle, and S. Voleti. The role of big data and predictive analytics in retailing. *Journal of Retailing*, 93(1):79–95, 2017.

L. Breiman. Random forests. *Machine Learning*, 45:5–32, 2001.

E. Brynjolfsson and L. Hitt. Beyond computation: Information technology, organizational transformation and business performance. *Journal of Economic Perspectives*, 14(4):23–48, 2000.

E. Brynjolfsson, D. Li, and L. Raymond. Generative ai at work. Available at `https://www.nber.org/papers/w31161`, 2023.

G. Cachon and C. Terwiesch. *Matching Supply with Demand: An Introduction to Operations Management, 3rd ed.* The McGraw-Hill Companies, 2012.

B. Callaway and P. Sant'Anna. Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230, 2021.

X. Cao and D. Zhang. The impact of forced intervention on ai adoption. Available at `https://dx.doi.org/10.2139/ssrn.3640862`, 2020.

L. Chen and H. Lee. Information sharing and order variability control under a generalized demand model. *Management Science*, 55(5):781–797, 2009.

V. Choudhary, M. Shunko, and S. Netessine. Does immediate feedback make you not try as hard? a study on automotive telematics. *Manufacturing & Service Operations Management*, 23(4): 835–853, 2021.

M. Cohen, S. Dahan, W. Khern-am nuai, H. Shimao, and J. Touboul. The use of ai in legal systems: determining independent contractor vs. employee status. *Artificial Intelligence and Law*, 2023.

Compass Group. Group sustainability report 2023. Available at `https://www.compass-group.com/content/dam/compass-group/corporate/sustainability/updates-2024/cgplc_sustainability_report_2023.pdf`, 2023.

R. Cui, M. Li, and S. Zhang. Ai and procurement. *Manufacturing & Service Operations Management*, 24(2):691–706, 2022.

Y. Cui, Z. Jiang, and Q. Li. Unlocking the value of real-time ai assistance: Who benefits, and why? Available at `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4497014`, 2023.

C. de Chaisemartin and X. D'Haultfoeuille. Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–2996, 2020.

D. Defize, K. Bitaraf, and N. Vermeer. Augmented data management: beyond the hype. Available at `https://www2.deloitte.com/nl/nl/pages/enterprise-technology-and-performance/articles/augmented-data-management-beyond-the-hype.html`, 2022.

F. Dell'Acqua, E. McFowland III, E. Mollick, H. Lifshitz-Assaf, K. Kellogg, S. Rajendran, L. Krayer, F. Candelon, and K. Lakhani. Navigating the jagged technological frontier: Field experimental evidence of the effects of ai on knowledge worker productivity and quality. Available at `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4573321`, 2023.

B. Dietvorst and S. Bharti. People reject algorithms in uncertain decision domains because they have diminishing sensitivity to forecasting error. *Psychological Science*, 31(10):1302–1314, 2020.

B. Dietvorst, J. P. Simmons, and C. Massey. Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144(1):114–126, 2015.

T. Erickson and T. Whited. Two-step gmm estimation of the errors-in-variables model using high-order moments. *Economic Theory*, 18(3):776–799, 2002.

H. Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P. Muller. Deep learning for time series classification: a review. *Data Mining and Knowledge Discovery*, 33:917–963, 2019.

B. Ferman and C. Pinto. Synthetic controls with imperfect pretreatment fit. *Quantitative Economics*, 12(2021):1197–1221, 2021.

E. Fix and J. Hodges. Discriminatory analysis. nonparametric discrimination: Consistency properties. *International Statistical Review / Revue Internationale de Statistique*, 57(3):238–247, 1989.

Food and Agriculture Organization. Food wastage footprint and climate change. Available at `https://www.fao.org/fileadmin/templates/nr/sustainability_pathways/docs/FWF_and_climate_change.pdf`, 2020.

Four Seasons. Four seasons for good: 2022 environmental, social and governance (esg) report. Available at `https://www.fourseasons.com/content/dam/fourseasons/images/web/PDFs/2022-ESG-Report.pdf`, 2022.

B. Frey. Shirking or work morale? *The European Economic Review*, 37(8):1523–1532, 1993.

J. Furman and R. Seamans. Ai and the economy. *Innovation Policy and the Economy*, 19:161–191, 2019.

D. Gaker, Y. Zheng, and J. Walker. Experimental economics in transportation: Focus on social influences and provision of information. *Transportation Research Record*, 2156(1):47–55, 2010.

A. Gerber, D. Green, and C. Larimer. Social pressure and voter turnout: Evidence from a large-scale field experiment. *The American Political Science Review*, 102(1):33–48, 2008.

Y. Gerchak and D. Mossman. On the effect of demand randomness on inventories and costs. *Operations Research*, 40(4):804–807, 1992.

F. Gino and G. Pisano. Toward a theory of behavioral operations. *Manufacturing & Service Operations Management*, 10(6):676–691, 2008.

K. Girotra, L. Meincke, C. Terwiesch, and K. Ulrich. Ideas are dimes a dozen: Large language models for idea generation in innovation. Available at `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4526071`, 2023.

A. Goodman-Bacon. Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277, 2021.

S. Graves, H. Meal, S. Dasu, and Y. Qiu. Two-stage production planning in a dynamic environment. In S. Axsater, C. Schneeweiss, and E. Silver, editors, *Multi-Stage Production Planning and Control. Lecture Notes in Economics and Mathematical Systems*, volume 266, pages 9–43. Springer-Verlag, Berlin, 1986.

C. Hanson. 2030 champions brings together business and government to cut us food waste. Available at `https://www.wri.org/insights/2030-champions-brings-together-business-and-government-cut-us-food-waste`, 2016.

D. Heath and P. Jackson. Modelling the evolution of demand forecasts with application to safety stock analysis in production/distribution systems. *IIE Transactions*, 26(3):17–30, 1994.

Hilton. Reduce and reuse: Tackling the global food waste challenge. Available at `https://stories.hilton.com/emea/releases/reduce-and-reuse-tackling-the-global-food-waste-challenge`, 2020.

L. Hitt and P. Tambe. Health care information technology, work organization, and nursing home performance. *ILR Review*, 69(4):834–859, 2016.

S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997.

P. Hu, S. Shum, and M. Yu. Joint inventory and markdown management for perishable goods with strategic consumer behavior. *Operations Research*, 64(1):118–134, 2016.

X. Hui, O. Reshef, and L. Zhou. The short-term effects of generative artificial intelligence on employment: Evidence from an online labor market. Available at `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4527336`, 2023.

N. Jain, K. Girotra, and S. Netessine. Managing global sourcing: Inventory performance. *Management Science*, 60(5):1202–1222, 2013.

J. Kahn. Influence functions for fun and profits. Available at `http://j-kahn.com/files/influencefunctions.pdf`, 2022.

D. Kahneman, P. Slovic, and A. Tversky. The simulation heuristics. In D. Kahneman, P. Slovic, and A. Tversky, editors, *Judgment under Uncertainty: Heuristics and Biases*. Cambridge University Press, Cambridge, 1982.

I. Karaesmen, A. Scheller-Wolf, and B. Deniz. Managing perishable and aging inventories: Review and future research directions. In K. Kempf, P. Keskinocak, and R. Uzsoy, editors, *Planning Production and Inventories in the Extended Enterprise: A State of the Art Handbook, Volume 1*, volume 151 of *International Series in Operations Research and Management Science*, pages 393–436. Springer US, New York, 2011.

F. Karim, S. Majumdar, H. Darabi, and S. Chen. Lstm fully convolutional networks for time series classification. Available at `https://arxiv.org/abs/1709.05206`, 2017.

K. Kawaguchi. When will workers follow an algorithm? a field experiment with a retail business. *Management Science*, 67(3):1670–1695, 2021.

N. Keskin, C. Li, and J. Song. The blockchain newsvendor: Value of freshness transparency and smart contracts. Available at `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3915358`, 2023.

S. Kim, J. Tong, and C. Peden. Admission control biases in hospital unit capacity management: How occupancy information hurdles and decision noise impact utilization. *Management Science*, 66(11):5151–5170, 2020.

S. Kirshner and B. Moritz. Comment on "newsvendor demand chasing revisited". *Management Science*, 66(12):6062–6064, 2020.

A. Kluger and A. DeNisi. The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2): 254–284, 1996.

M. Kremer, B. Moritz, and E. Siemsen. Demand forecasting behavior: System neglect and change detection. *Management Science*, 57(10):1827–1843, 2011.

N. Lau and J. Bearden. Newsvendor demand chasing revisited. *Management Science*, 59(5):1245–1249, 2013.

Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

L. Lemos. Innovative food waste technology that supports commercial kitchens. Available at https://blog.winnowsolutions.com/food-waste-technology-supports-kitchens, 2019.

Q. Li, P. Yu, and X. Wu. Managing perishable inventories in retailing: Replenishment, clearance sales, and segregation. *Operations Research*, 64(6):1270–1284, 2016.

W. Lin, S. Kim, and J. Tong. What drives algorithm use? an empirical analysis of algorithm use in type 1 diabetes self-management. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3891832, 2023.

J. Lismont, J. Vanthienen, B. Baesens, and W. Lemahieu. Defining analytics maturity indicators: A survey approach. *International Journal of Information Management*, 37(3):114–124, 2017.

X. Liu, A. Susarla, and R. Padman. Ai for social good: A human-in-the-loop augmented intelligence video understandability assessment to promote health literacy. Working paper, 2022.

J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. Available at https://arxiv.org/abs/1411.4038, 2014.

S. Nahmias. Perishable inventory systems. In *International Series in Operations Research and Management Science*, volume 160. Springer US, New York, 2011.

N. Otis, R. Clarke, S. Delecourt, D. Holtz, and R. Koning. The uneven impact of generative ai on entrepreneurial performance. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4671369, 2024.

L. Pierce, D. Snow, and A. McAfee. Cleaning house: The impact of information technology monitoring on employee theft and productivity. *Management Science*, 61(10):2299–2319, 2015.

Z. Porreca. Synthetic difference-in-differences estimation with staggered treatment timing. *Economics Letters*, 220(2022):110874, 2022.

C. Rolim, P. Baptista, G. Duarte, T. Farias, and J. Pereira. Impacts of real-time feedback on driving behaviour: A case-study of bus passenger drivers. *European Journal of Transport and Infrastructure Research*, 17(3):346–359, 2017.

L. Rubin. Aggregate inventory behavior: Response to uncertainty and interest rates. *Journal of Post Keynesian Economics*, 2(2):201–211, 1980.

N. Rudi and D. Drake. Observation bias: The impact of demand censoring on newsvendor level and adjustment behavior. *Management Science*, 60(5):1334–1345, 2014.

R. Sanders. Dynamic pricing and organic waste bans: A study of grocery retailers' incentives to reduce food waste. Available at https://dx.doi.org/10.2139/ssrn.2994426, 2020.

A. Schleifer. Managing global sourcing: Inventory performance. *Harvard Business School Teaching Note*, pages 895–057, 1995.

M. Schweitzer and G. Cachon. Decision bias in the newsvendor problem with a known demand distribution: Experimental evidence. *Management Science*, 46(3):404–420, 2000.

M. Soleymanian, C. Weinberg, and T. Zhu. Sensor data and behavioral tracking: Does usage-based auto insurance benefit drivers? *Marketing Science*, 38(1):21–43, 2019.

J. Song, H. Zhang, Y. Hou, and M. Wang. The effect of lead time and demand uncertainties in (r, q) inventory systems. *Operations Research*, 58(1):68–80, 2009.

B. Staats, H. Dai, D. Hofmann, and K. Milkman. Motivating process compliance through individual electronic monitoring: An empirical examination of hand hygiene in healthcare. *Management Science*, 63(5):1271–1656, 2016.

X. Su. Bounded rationality in newsvendor models. *Manufacturing & Service Operations Management*, 10(4):566–589, 2008.

J. Sun, D. Zhang, H. Hu, and J. Mieghem. Predicting human discretion to adjust algorithmic prescription: A large-scale field experiment in warehouse operations. *Management Science*, 68 (2):846–865, 2022.

L. Sun and S. Abraham. Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):175–199, 2021.

R. Thaler and S. Benartzi. Save more Tomorrom$^{\text{TM}}$: Using behavioral economics to increase employee saving. *Journal of Political Economy*, 112(1):164–187, 2004.

A. Tversky and D. Kahneman. Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5(2):207–232, 1973.

United Nations. Sustainable development goal target 12.3. Available at `https://sdg12hub.org/sdg-12-hub/see-progress-on-sdg-12-by-target/123-food-loss-wastef`, 2021.

K. Wang and A. Goldfarb. Can offline stores drive online sales? *Journal of Marketing Research*, 54 (5):706–719, 2017.

N. Watson and Y. Zheng. Over-reaction to demand changes due to subjective and quantitative forecasting. Available at `https://www.hbs.edu/ris/Publication%20Files/09-021.pdf`, 2008.

M. Wedel and P. Kannan. Marketing analytics for data-rich environments. *Journal of Marketing*, 80(6):97–121, 2016.

Winnow Solutions. Why are we the company to meet the food waste challenge? Available at `https://www.winnowsolutions.com/company`, 2024.

World Resources Institute. 2030 champions brings together business and government to cut us food waste. Available at `https://www.wri.org/insights/2030-champions-brings-together-business-and-government-cut-us-food-waste`, 2016.

C. Wu and C. Teng. Reducing food waste in buffet restaurants: A corporate management approach. *Foods*, 12(1):162, 2023.

L. Yang and M. Yu. To good to do: Combating food waste with surprise clearance. Available at `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4573386`, 2023.

Appendix A. Supplemental Materials

A.1. **Summary Statistics of Subsample.** Comparing with Table 1 in Section 5 for the full sample, we observe that statistics of the main variables basically, do not change much in the subsample. Also, the subsample sites vary by their service type and industry category, staying representative for the full sample sites.

Variables		**Mean**	**Standard Dev.**	**Min**	**Median**	**Max**	**N**
Daily total waste (weight)	grams	437,575	692,639	10	218,340	36,435,990	116,109
# of meals served daily (covers)	count	1,556	1,478	1	932	22,808	116,109
Daily sales (sales)	US dollars	12,120	10,602	1.11	9,230	319,350	116,094
Demand variability (demand_cv)	coef. of variation	0.2784	0.2240	0.0011	0.2252	2.8119	114,572
Waste per cover (waste_per_cover)	grams/cover	642	1,532	0.0026	237	155,168	116,109

Notes. In our main treatment effect analysis, we only keep 178 eligible sites and drop those whose donor pool for synthetic control construction is empty. The table reports summary statistics of the subsample.

Table 6. Subsample Summary Statistics

A.2. **Calculation of Estimated Standard Error for $\hat{\tau}$.** Let $N = \sum_r N_r^{tr}$ be the number of all the treatment sites where N_r^{tr} is the set of treatment sites in cohort r. Following Kahn (2022), it is a property of the influence functions of such estimators of summary parameters that the following holds:

$$\frac{1}{\sqrt{N}}(\hat{\tau} - \tau) = \sum_{i=1}^{N} \psi_\tau(x_i) + o_p(1)$$

where $\psi_\tau(x_i)$ is the influence function for the i-th observation (that tells the effect of a change in one observation on an estimator) and the summary parameter τ. As such, the variance of our summary parameter $\hat{\tau}$ can be computed through the procedure described in Erickson and Whited (2002) and Kahn (2022). The main idea is to calculate empirical equivalents of the influence functions for each estimate, and stack them into a single matrix, $\mathbf{\Psi}$, in which the rows correspond to each observation and the columns to each estimator. For the cohort-level estimators $\hat{\tau}_r$, $r = 1, ..., M$, and observations $i = 1, ..., N$, we create a matrix (i.e., in our setting $M = 79$, $N = 178$)

$$\mathbf{\Psi} = [\psi_{\hat{\tau}_1}, ..., \psi_{\hat{\tau}_M}]_{N \times M}$$

where every element of $\mathbf{\Psi}$, $[\mathbf{\Psi}]_{ij}$ is equal to $\psi_{\hat{\tau}_j}(x_i)$. The variance-covariance matrix for the individual cohort estimators can be thus calculated

$$\hat{\boldsymbol{V}} = \frac{1}{N^2}(\mathbf{\Psi}^T \mathbf{\Psi}).$$

Then using the weight vector $\boldsymbol{\mu} = [\mu_1, ..., \mu_M]^T$, an estimator for the variance of the aggregated summary parameter can be computed as

$$\hat{V}_{\hat{\tau}} = \boldsymbol{\mu}^T \hat{\boldsymbol{V}} \boldsymbol{\mu}.$$

A.3. **Constructing the SynthDiD Event Study Plot.** Synthetic DiD presented in Section 6.2 is designed to minimize the mean squared error of a target estimated ATT and not to separately measure effects in specific time periods. The event study plot we construct is thus used to illustrate the effects in each time period. To compute the treatment effects pre-adoption, we compute for each adoption cohort r and each time period t between $r + l_{min}$ and r–1,

$$\hat{\tau}_{r(t-r)} = (\frac{\sum_{i \in N_r^{tr}} \log(Y_{it}+1)}{|N_r^{tr}|} - (\hat{\omega}_0 + \sum_{i \in N_r^{co}} \hat{\omega}_i \log(Y_{it}+1))) \cdot \hat{\lambda}_t.$$

Because the weights $\hat{\lambda}_t$ for time periods before adoption sum up to one, summing up $\hat{\tau}_{r(t-r)}$ yields values that are approximately zero, which shows a good fit between the treatment outcomes and the synthetic control pre-adoption. The standard errors for these values are computed using the jackknife method (or the placebo method in the case where there is only one treated unit). The values for each cohort are then averaged, and the standard errors are aggregated appropriately. Computing the cohort-level effects post-adoption is done in a similar manner. For each adoption cohort r and each time period t between r and $r + l_{max}$, we compute

$$\hat{\tau}_{r(t-r)} = (\frac{\sum_{i \in N_r^{tr}} \log(Y_{it}+1)}{|N_r^{tr}|} - (\hat{\omega}_0 + \sum_{i \in N_r^{co}} \hat{\omega}_i \log(Y_{it}+1))),$$

where we do not weigh the estimated effect by $\hat{\lambda}_t$ as the weight is simply $\frac{1}{l_{max}+1}$ pre-adoption. Averaging the effects within cohorts and the resulting averages across cohorts then produces the ATT reported by SynthDiD. The standard error for the aggregate ATT is computed using the Jackknife method (or the placebo method as before) for the aggregated value, which takes into account the potential serial correlation between effects across time.

A.4. **Computing Individual Treatment Effects.** To compute individual treatment effects, for each individual site i that adopted Winnow in period r and have outcome data available $l_{min} < 0$ periods before and $l_{max} + 1$ periods after, we construct a control group that comprises sites that have data available for the same time frame but that adopted Winnow after more than l_{max} periods after r. Let N_{r_i} denote the set of all units in the balanced panel of the site i who adopted Winnow in period r and $N_{r_i}^{co}$ denotes the set of control units. The Synthetic DiD estimation procedure solves

$$(\hat{\tau}_i, \hat{\alpha}_0, \hat{\alpha}_j, \hat{\gamma}_t) = \underset{\tau_i,\alpha_0,\alpha_j,\gamma_t}{\arg\min} \{ \sum_{j \in N_{r_i}} \sum_{t=r+l_{min}}^{r+l_{max}} (\log(Y_{jt}+1) - \alpha_0 - \alpha_j - \gamma_t - AfterAdopt_{jt} \cdot \tau_i)^2 \hat{\omega}_j \hat{\lambda}_t \},$$

where $\hat{\tau}_i$ measures the average ATT of site i within the $l_{max}+1$ periods after adoption.

Then following the way we compute dynamic event study treatment effect effects in Appendix A.3, for each adopting site i and each time period t between $r + l_{min}$ and r–1, we obtain

$$\hat{\tau}_{it} = (log(Y_{it}+1) - (\hat{\omega}_0 + \sum_{j \in N^{co}_{r_i}} \hat{\omega}_j \log(Y_{jt}+1))) \cdot \hat{\lambda}_t.$$

For each adopting site i and each time period t between r and $r + l_{max}$, we compute

$$\hat{\tau}_{it} = log(Y_{it}+1) - (\hat{\omega}_0 + \sum_{j \in N^{co}_{r_i}} \hat{\omega}_j \log(Y_{jt}+1)).$$

Thus, $\hat{\tau}_{it}$ measures the individual treatment effect for site i in period t, which we will use as the outcome variable in the heterogeneity analysis of Section 7.3.

A.5. **Vision Upgrade Choice.** We estimate a random effects probit model for the decision to upgrade the Winnow Classic to Winnow Vision. We estimate the following model:

$$Pr(Upgrade_{it} = 1) = \Phi(\alpha + \beta X_{it} + \mu_i)$$

where $Upgrade_{it}$ indicates whether site i upgraded the system in period t. Observations after the upgrade are not used to estimate the above equation because the upgrade decision is made once. X_{it} indicates variables influencing the selection decision $Upgrade_{it}$, which could be a mixture of time-variant variables, time-invariant variables, and time dummies. Due to a large number of sites, and to address the incidental parameter problem with site-fixed effects in probit models, we also include μ_i as site random effects instead of site-fixed effects.

In our study, variables of interest include $waste\_per\_cover_{it}$ that measures the lagged waste level at site i in period t, $sales_{it}$ that is used as a proxy variable for the lagged budget size of site i in period t, $staff\_num_i$ that indicates the labor intensity of site i. Table 7 reports the estimation results.

A.6. **Staggered DiD Estimation.** We analyze the impact of Winnow adoption on food waste using classic staggered difference-in-differences (SDiD). We estimate the following SDiD model using OLS:

$$log(Y_{it}+1) = \alpha_i + \gamma_t + AfterAdopt_{it} \cdot \tau + \epsilon_{it}$$

Probit Models for Upgrade Decision		
	Model 1	Model 2
Constant	-9.5868***	-5.4415
	(2.1397)	(4.7475)
Log(lag waste per cover)	0.0693	0.2071
	(0.0922)	(0.3339)
Log(lag sales)	0.4433	0.0937
	(0.3569)	(0.4711)
Number of staff		0.0139
		(0.0117)
Site random effects	Yes	Yes
N_observations	1,643	123
N_site	178	25
*** -- 1% level, ** -- 5% level, * -- 10% level		
The records of staff number are missing for some sites.		

Table 7. Probit Models for the Upgrade Decision

where Y_{it} is the daily food waste level (in grams/cover) for site i on day t, and $AfterAdopt_{it}$ indicates whether site i adopted Winnow by day t. We control for individual fixed effects α_i and time fixed effects γ_t. Two-way clustering of standard errors by site and day, ϵ_{it}, is used to address serial correlation (Bertrand et al. 2004). The coefficient τ measures the change in the daily food waste level after Winnow adoption. Like Wang and Goldfarb (2017), it is identified by comparing the change in food waste before and after adoption for adopting sites with the change in food waste in the same time periods for sites that have not yet adopted Winnow.

The identification assumption in our SDiD analysis is the parallel trends assumption, i.e., there were no differential trends in food waste before adoption between sites that adopted Winnow and those that did not. We thus estimate the OLS model below to statistically test the identification assumption following Borusyak et al. (2024):

$$log(Y_{it} + 1) = \alpha_i + \gamma_t + \sum_{k \neq l_{min}, -1} \tau_k \cdot D_{it}^k + \epsilon_{it}$$

where D_{it}^k indicates whether the time $t$equals k periods relative to adoption for the site i and τ_k measures the effect of Winnow adoption in k periods relative to adoption (i.e., k period specific event-study estimator). The minimum lag l_{min} and the lag of $k = -1$ are excluded for identification purposes. The baseline of the comparison is one period before adoption such that the value at $k = -1$ is set to zero.

Figure A.1 plots the coefficient estimates $\hat{\tau}_k$, along with their 95% confidence intervals (CIs) for the daily indicators 14 days before adoption and 90 days after adoption. Time 0 is the day of adoption.

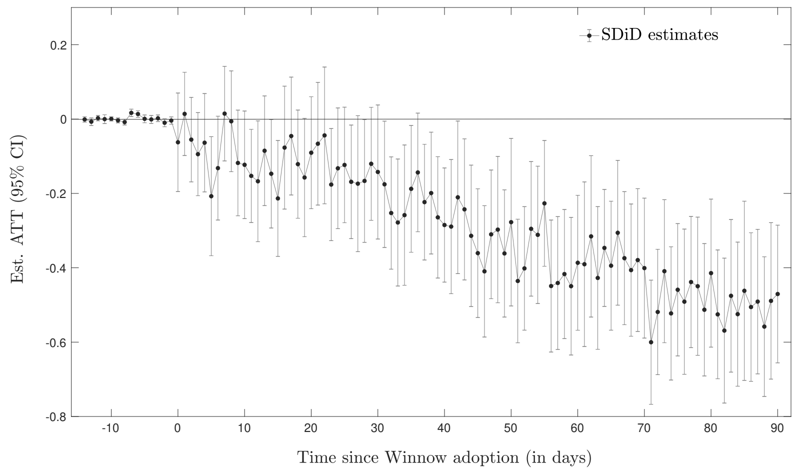


FIGURE A.1. Effect of Winnow Adoption on Food Waste Level via Staggered DiD

As the figure reveals, although the coefficients are statistically indistinguishable from zero except six and seven periods before adoption (which may raise a concern about the validity of the SDiD analysis), there is no increasing or decreasing trend in the coefficients in the pre-adoption period. The treatment effect estimates post-adoption share similar trends with those estimated via synthetic DiD.

A.7. **Synthetic Data Generation.** To train an AI-based system designed for managerial bias detection and evaluate its effectiveness. We collect synthesized data by simulating the dynamics of real-world food production environments according to the perishable (Q, r) model. Here are detailed steps for simulating calibrated perishable (Q, r) models under different experimental setups, where either bias-free or biased production decisions induced by various managerial biases are captured.

Demand Generation. Unit demands are generated according to a Poisson process with rate $\widetilde{\lambda}$

$$\widetilde{\lambda} = Y \cdot \overline{\lambda},$$

where $\overline{\lambda}$ is the long-term average demand rate, e.g., 10 arrivals per hour, and Y is any useful information observed by the decision maker, e.g., weather conditions.

Let's initialize the system with Q_0 fully fresh items on hand. For any store i, we simulate the Poisson arrival process within every single day (e.g., assuming 12 operating hours per day) and aggregate the arrivals as the demand in period $t = 1, ..., T$, i.e.,

$$d_{it} = N_{it}(12), \text{ for } i = 1, ...N, t = 1, ...T,$$

where $N_{it}(t)$ is the counting process of the Poisson arrivals in $(0, t]$ with rate $y_{it} \cdot \overline{\lambda}$.

Production Generation. The optimal production quantity and reorder point in period t at store i are obtained by numerically solving the following optimization problem

$$(q_{it}^{PF*}, r_{it}^{PF*}) = \arg\min_{(Q,r)} TC = \frac{K + cQ + \int_{z=L}^{\tau} (h\mathbb{E}[OH_{it}(Q,r)|Z=z] + p\mathbb{E}[P_{it}(Q,r)|Z=z] + \pi\mathbb{E}[LS_{it}(Q,r)|Z=z])dF_{it}(z)}{\int_{z=L}^{\tau} \mathbb{E}[CL_{it}(Q,r)|Z=z]dF_{it}(z)},$$

where $F_{it}(z)$ is the stationary distribution of the effective shelf life Z at steady state.

The approximately optimal solution to the optimization problem in period t, $(q_{it}^{PF*}, r_{it}^{PF*})$, can be obtained by applying exhaustive search algorithm (see Algorithm 1 below) on the sample average total cost rate $\widehat{TC}_{it}$.

Waste Generation. By aggregating the perishing items within every single day, we obtain the waste data streams for training and testing purpose.

Incorporate Different Managerial Biases into Perishable (Q, r) Model. The above captures the case of a bias-free decision. Next, there are several ways to add different kinds of managerial biases to the bias-free decision.

▷ **Static production plan**

The decision maker at store i may simply implement a static production plan based on their experience or intuition when they make Q decisions. This can be captured by generating the demand data streams according to the Poisson arrival process with time-invariant rate

$$\widetilde{\lambda}_{it} = \overline{y} \cdot \overline{\lambda}, \text{ for } i = 1, ...N, t = 1, ...T,$$

where $\overline{y}$ can be any constant term.

▷ **Demand chasing**

The decision maker at store i may adjust the Q decision towards the demand in the prior period. Mathematically speaking, we modify Q_{it} by

$$Q_{it} = Q_{i,t-1} + \alpha(\overline{d}_{i,t-1} - Q_{i,t-1}) + \epsilon_{it}, \ t \geq 2$$

where $\overline{d}_{i,t-1} = d_{i,t-1}/M_{i,t-1}$ is the average demand per cycle with $d_{i,t-1}$ being the Poisson arrivals counted within period $t-1$ at store i and M_{t-1} being the number of embedded cycle within period $t-1$ at store i, the noise in the production quantity is an iid random variable $\varepsilon_{it} \sim N(0, \sigma^2)$, and $\alpha \in [0, 1]$ captures the degree of demand chasing. We initialize q_{i1} from the bias-free perishable (Q, r) model ("good behavior"), i.e., $q_{i1} = q_{i1}^{PF*}$ where q_{i1}^{PF*} is the optimal production quantity.

Algorithm 1 Monte Carlo Simulation of Perishable (Q, r) Model

Simulate sufficient demand arrivals according to the Poisson process at rate λ.
Create a sequence of arrival times $A = [A_0, A_1, A_2, ...]$.
Create an array of zeros of size N for $\{T_n : n = 0, 1, ..., N\}$, the sequence of time epochs at which the inventory level hits Q.
Create an array of τ's of size N for $\{Z_n : n = 0, 1, ..., N\}$, the sequence of effective shelf lives of items on hand at T_n.
Create arrays of zeros for OH (total stocking time), LS (total number of lost sales), P (total number of perishing items), TC (total cost rate) within $[T_{n-1}, T_n]$.
function Q_R_OPTIMIZE(Q, r)
  while $i \leq N - 1$ **do**
    $j \leftarrow \arg\min_k A_k \geq T_i$
    $M_1 \leftarrow \arg\min_k A_k > T_i + Z_i$
    $M_2 \leftarrow \arg\min_k A_k > T_i + L$
    $M_3 \leftarrow \arg\min_k A_k > T[i] + L + A_{j+Q-r} - A_{j+Q}$
    $M_4 \leftarrow \arg\min_k A_k > T[i] + L - Z[i] + A_{j+Q-r} - A_j$
    if $A_{j+Q} - A_{j+Q-r} > L$ and $A_{j+Q} - A_j < Z[i]$ **then** ▷ Compute the effective shelf life
      $Z[i+1] \leftarrow \tau - (A_{j+Q} - A_{j+Q-r}) + L$
    else if $A_{j+Q} - A_{j+Q-r} > L$ and $A_{j+Q} - A_j > Z[i]$ and $A_{j+Q-r} - A_j < Z[i] - L$ **then**
      $Z[i+1] \leftarrow \tau - (Z[i] - A_{j+Q-r} + A_j)$
    else
      $Z[i+1] \leftarrow \tau$
    end if
    if $A_{j+Q-r} - A_j > Z[i]$ **then**
      $T[i+1] \leftarrow T[i] + L + Z[i]$ ▷ Compute the time epoch at which inventory hits Q
      $OH[i+1] \leftarrow \sum_{k=1}^{M_1-j}(A_{j+k} - A_j) + Z[i] * (Q - (M_1 - j))$ ▷ Compute stocking time
      $LS[i+1] \leftarrow M_2 - j$ ▷ Compute total number of lost sales
      $P[i+1] \leftarrow max(0, Q - (M_1 - j))$ ▷ Compute total number of perishing items
      $TC[i+1] \leftarrow (K + cQ + h * OH[i+1] + p * P[i+1] + \pi * LS[i+1])/(T[i+1] - T[i])$
    else if $A_{j+Q-r} + L > A_{j+Q}$ and $A_{j+Q} - A_j < Z[i]$ **then**
      $T[i+1] \leftarrow T[i] + L + A_{j+Q-r} - A_j$
      $OH[i+1] \leftarrow \sum_{k=1}^{M_1-j}(A_{j+k} - A_j)$
      $LS[i+1] \leftarrow M_3 - j$
      $P[i+1] \leftarrow max(0, Q - (M_1 - j))$
      $TC[i+1] \leftarrow (K + cQ + h * OH[i+1] + p * P[i+1] + \pi * LS[i+1])/(T[i+1] - T[i])$
    else if $A_{j+Q-r} + L > A_j + Z[i]$ and $A_{j+Q-r} - A_j < Z[i]$ and $A_{j+Q} - A_j > Z[i]$ **then**
      $T[i+1] \leftarrow T[i] + L + A_{j+Q-r} - A_j$
      $OH[i+1] \leftarrow \sum_{k=1}^{M_1-j}(A_{j+k} - A_j) + Z[i] * (Q - (M_1 - j))$
      $LS[i+1] \leftarrow M_4 - j$
      $P[i+1] \leftarrow max(0, Q - (M_1 - j))$
      $TC[i+1] \leftarrow (K + cQ + h * OH[i+1] + p * P[i+1] + \pi * LS[i+1])/(T[i+1] - T[i])$
    else if $A_{j+Q-r} + L < A_{j+Q}$ and $A_{j+Q} - A_j < Z[i]$ **then**
      $T[i+1] \leftarrow T[i] + L + A_{j+Q} - A_j$
      $OH[i+1] \leftarrow \sum_{k=1}^{M_1-j}(A_{j+k} - A_j) + Q * (A_{j+Q} - A_{j+Q-r} - L)$
      $LS[i+1] \leftarrow 0$
      $P[i+1] \leftarrow max(0, Q - (M_1 - j))$
      $TC[i+1] \leftarrow (K + cQ + h * OH[i+1] + p * P[i+1] + \pi * LS[i+1])/(T[i+1] - T[i])$
    else
      $T[i+1] \leftarrow T[i] + Z[i]$
      $OH[i+1] \leftarrow \sum_{k=1}^{M_1-j}(A_{j+k} - A_j) + Z[i] * (Q - (M_1 - j)) + Q * (Z[i] - (A_{j+Q} - A_{j+Q-r}) - L)$
      $LS[i+1] \leftarrow 0$
      $P[i+1] \leftarrow max(0, Q - (M_1 - j))$
      $TC[i+1] \leftarrow (K + cQ + h * OH[i+1] + p * P[i+1] + \pi * LS[i+1])/(T[i+1] - T[i])$
    end if
  end while
  return $\frac{1}{N}\sum_i TC[i]$
end function
Define vectors of possible Q and r values, $sample\_Q$ and $sample\_r$.
Create an empty matrix for evaluated samples, $sample\_eval$.
Initialize the best approximate solution $best\_ix\_Q \leftarrow 0$, $best\_ix\_r \leftarrow 0$.
while $i \leq len(sample\_Q)$ **do**
  while $j \leq len(sample\_r)$ **do**
    $sample\_eval[i, j] \leftarrow$ Q_R_OPTIMIZE$(sample_Q[i], sample_r[j])$
    if $sample\_eval[i, j] < sample\_eval[best\_ix\_Q, best\_ix\_r]$ **then**
      $(best\_ix\_Q, best\_ix\_r) \leftarrow (i, j)$
    end if
  end while
end while
Print the best approximate solution $(best\_ix\_Q, best\_ix\_r)$.

▷ **Overreaction to demand changes**

The decision maker at store i may overreact to demand changes in prior periods. Mathematically speaking, we modify Q_{it} by

$$Q_{it} = Q_{i,t-1} + \delta(\overline{d}_{i,t-1} - \overline{d}_{i,t-2})^+ - \gamma(\overline{d}_{i,t-2} - \overline{d}_{i,t-1})^+ + \epsilon_{it},\ t \geq 3$$

where $\overline{d}_{i,t-1}$ is the average demand per cycle within period $t-1$ at store i, the noise in the production quantity is an iid random variable $\varepsilon_{it} \sim N(0, \sigma^2)$, $\delta \geq 0$ is the degree of overreacting to the demand spike, and $\gamma \geq 0$ is the degree of overreacting to the demand slump. We initialize q_{i1} and q_{i2} from the bias-free perishable (Q, r) model ("good behavior").

A.8. Machine Learning Based Time-series Classifiers.

A.8.1. *Classifier Development.* Define a synthesized dataset as $D_{syn} = \{(X_1, Y_1), ..., (X_N, Y_N)\}$, which is a collection of pairs (X_i, Y_i) where X_i is a waste time series that is two-week long and Y_i is the assigned label (i.e., one of the three main behavioral biases or neither of them). Each of the waste time series in our study is two weeks long.[10] Our task is to develop classifiers that could predict at high accuracy which bias is labeled given any new waste time series as input.

We consider k-nearest neighbors (KNN) and random forests (RF) as benchmark classifiers and use long short-term memory fully convolutional networks (LSTM-FCN) as our main classifier.

Consider an L-layer deep neural network, where each layer l_i is a representation of the input domain, taking the output of its previous layer l_{i-1} as input and applying a non-linearity (such as the Sigmoid function) to compute its own output. The behavior of these non-linear transformations is controlled by a set of parameters θ_i for each layer (i.e., weights). Hence, mathematically, given an input x, a neural network performs the following computations to predict the class:

$$f_L(\theta_L, x) = f_{L-1}(\theta_{L-1}, f_{L-2}(\theta_{L-2}, ..., f_1(\theta_1, x)))$$

where f_i is the non-linearity applied at layer l_i and f_L is the final output (i.e., the results of convolutions) in the form of a probability distribution over the class variable in the dataset. During training, the network is presented with a certain number of known input outputs. Referring to Fawaz et al. (2019), first, the weights are initialized randomly. Second, a forward pass through the model is applied, that is, computing the output given an input x. The output is a vector whose components are the estimated probabilities of x belonging to each class, and the label assigned corresponds to

[10] For most sites the pre-adoption period is at least two-week long. Each waste time series in the synthetic dataset is set to be two-week long, so that the classifiers developed on the synthetic data can be applied on the real waste data pre- and post- Winnow adoption.

the class with maximum probability. The prediction loss of the model is thus computed using a cost function, such as the negative log-likelihood. Third, the weights are updated in a backward pass to propagate the error via gradient descent. By iteratively taking a forward pass followed by back-propagation, the model's parameters are updated in a way that minimizes the loss on the training data. During testing, the model is tested on the rest of D_{syn}. The accuracy measure is thus computed by comparing the predicted labels by the model with the true labels in the testing dataset.

A.8.2. *Results.* KNN, RF, and LSTM-FCN are all implemented in PyTorch. We develop these classifiers using a synthetic dataset of size 20,000 (5,000 for each class), in which 80% of the data are used for training, and the remaining are used for validation. The validation set is used to evaluate the performance of classification models for various combinations of hyperparameters and prevent overfitting.

The KNN model (with the number of neighbors set as $k = 5$) and the RF model (with the number of trees in the forest set as 500) yield 55% and 75% classification accuracy, respectively. The LSTM-FCN model turns out 90% accuracy. We refer to the one proposed in Karim et al. (2017) as its model architecture, where there are three convolution layers, each of which has 64 filters and an LSTM layer with 8 cells followed by a dropout layer. During the training process, cross-entropy loss function and Adam optimizer are used. The model is finally tuned with hyperparameters: learning rate = 0.0001, mini-batch size = 128, weight initialization = Kaiming initialization, activation = Rectified linear unit (ReLU), Dropout rate = 0.25.