STOCKHOLM SCHOOL OF ECONOMICS
Department of Economics
5350 Master's thesis in economics
Academic year 2019–2020

The Effect of Self Service Technologies on Store Performance in Retail

Valerio Romano (41453)

Abstract:

The purpose of this case study is to examine the effects of the introduction of self service technologies (SSTs) on store performance. The data are collected from a Swedish multinational company and studied through difference in difference regressions. "Store performance" is assessed by studying 8 metrics selected based on the literature as well as the needs of the company. These measures are idle time, conversion rates and productivity. My main findings are that in the stores in the treatment group the introduction of SSTs increased the time to complete a transaction by an average of 8.61 seconds, increased the average time to scan an item by 5 seconds, increased the *Service Gap* by 35 seconds and decreased the *Number of Daily Queues* by 45 and the *Proportion of Daily Queues* by 3.4 %. These numbers sum up to be non-trivial when aggregating over the many transactions made each day. The SSTs were not found to influence conversion rates and productivity. It is possible that the effects related to idle time at least partially cancelled out, yielding no significant effect on productivity and conversion rates.

Keywords: self service, speed of transaction, queueing time, conversion rates, productivity

JEL: L250, L810

Supervisor:	Matilda Orth
Date submitted:	18/05/2020
Date examined:	27/05/2020
Discussant:	Ismael Moreno Martinez, Markus Jury
Examiner:	Magnus Johannesson

Table of Contents

1. Introduction

1.1 Background

The creation and introduction of innovative technologies is extremely important for economic development. As Syverson states (2011), research in industrial organization has linked productivity levels to the adoption of different kinds of technologies such as IT as well as technology spillovers between industries. As productivity rises, more and better products are produced and consumed, providing profits to companies, salaries to workers and consumer surplus. Van Biesebroeck (2003) studies how auto assembly plants introduced "lean" technologies and determines the impact on productivity (thus also discussing the use of capital): he finds that introducing capital augmented technologies was the primary driver of labour productivity growth in the industry in the late 1980s. It is thus of no surprise that companies invest large amounts of capital to own the best and newest technologies in the hope of improving the efficiency of their operations and outcompeting their rivals. Other key factors in determining productivity are the allocation and characteristics of labour and capital: for the same technological level, companies can have different labour quality and productivities (Syverson 2011). Ilmakunnas, Maliranta & Vainiomäki (2004) for example show that productivity increases with workers' education and age. The balance between labour and capital will often change when new technologies are introduced: Self Service Technologies (SSTs) for example replace labour with capital in the checkout process.

All industries can benefit from better technology, but not all industries do so in the same way or are equally important for the economy. The service industry has been growing dramatically in recent decades. The World Bank for example calculates that in 2017 the service sector represented 65% of world GDP, increasing from 54% in 1995. Similarly, the percentage of worldwide employment in the service sector increased from 36% to 48% over the same period. The development of this sector is now widely recognized to be crucial to the success of an economy, and even within companies *service* has become of great importance, as it is now necessary to examine the entire life cycle of whatever is sold, even if it is a product (Lee, Ribeiro, Ohlson, Roig 2006). This rise in the importance of services and the introduction of new technologies seem to be tightly linked. For example, the growth of IT in the 90s has created an entire range of new business to business industries related to IT services. To quote Karmarkar, U. (2004), "the primary change driver behind the service revolution is technology". It is thus important to examine how innovations can affect businesses in this area. In retail, SSTs seem some of the most promising technologies introduced in recent years which affect how the customer is being served. By now, SSTs permeate many retail sub-sectors such as supermarkets and clothing stores.

The first forms of SSTs were automated vending machines, which were first introduced in the 1880s in London. At the time, essentially every sale of goods was done between a human employee and a customer, and it would take another half a century until similar innovations would be invented (in 1947 the first self service gas station was created). The 1960s were the key decade for SSTs. In 1964, Herb Timms designed a system that enabled an attendant inside a refilling store to activate the pumps outside: this allowed self service in petrol stations to become widespread in the US. In the same years, ATMs were also invented and introduced in Japan, Sweden, the UK and the US. From the petrol and financial industries, kiosks would then become widespread in parking, hotels, restaurants, the transportation industry (airplane and train tickets) and finally the retail sector.

SSTs were first introduced in retail in supermarkets in the late 1990s and early 2000s, but soon after other kinds of retailers such as houseware (IKEA) and clothing stores adopted the technology. Even in retail, the actual machines employed differ widely depending on the application and the industry. For example, some self-checkouts have buttons to select items while others employ touch-screens, some provide a pistol to scan items, some have a fixed scanning area while others do not require any kind of scanning. This heterogeneity makes conclusions on SSTs hard to generalize. Nonetheless, a literature is growing on self service and its applications. In the marketing domain, perceived service quality of SSTs has been investigated in recent years by Dabholkar (1996), Dabholkar, Bobbitt & Lee (2003) & Anselmsson (2001). Meanwhile, Basker, Foster & Klimek (2017) have been examining how SSTs affect productivity.

1.2 The effect of SSTs on store performance

The purpose of this study is to examine the effect of automated checkouts on store performance in a multinational Swedish retailer. My goal is to understand how the retailer's business processes and outcomes are affected by the introduction of the technology from the stores' perspective. All the measures I studied were thus explicitly suggested to me by the company which provided me the data because they were believed to be of interest and possibly affected by the introduction of SSTs. As a result, this study will be of high interest to them. Moreover, other retailers planning to introduce automated checkouts will also find this useful. Finally, this case study is also of interest to economists, as it describes how a new technology affected store performance in retail using detailed data. According to Basker (2012), many authors have speculated that the rise in product variety and the growth of stores observed in past decades was made possible by the introduction of innovations such as barcode scanners and IT (see, e.g., Holmes 2001; Basker, Klimek, & Van 2012). As a result, my study will add to this literature by describing the specific case of SSTs.

Jointly, the measures I have chosen provide a good understanding of how the store's processes and performance have changed after the introduction. These changes are varied, and since "store performance" is a broad concept, I have examined it from several perspectives. Thanks to the data available, I will be able to assess how introducing the new technology affected the speed of checkout, the queuing time, the conversion rate of customers and the labour productivity of employees. Productivity measures directly how efficient the store is, and profits are increasing in the company's productivity level (Syverson 2011). To improve a stores' performance, it is thus crucial to understand which technological interventions can raise its productivity. Measures of idle time are also important because they describe how the efficiency of the checkout process changed after the introduction; the checkout is of course the part of the customer journey which is most affected by SSTs. Generally speaking, customers prefer to check out quickly (Dabholkar, Bobbitt & Lee. 2003) and avoid waiting in line (Johnston 2003). Bottlenecks in operations are thus considered to be an issue that needs to be solved (Schmenner & Swink 1998). As a result, the performance and profitability of a store is likely to improve if customers are provided a higher quality service and can complete their transaction quickly, increasing the likelihood that they will return for future purchases. Finally, conversion rates measure how effective the store is at convincing potential customers to make a purchase. By definition, higher conversion rates imply higher sales and a higher degree of customer satisfaction (all else equal). This in turn will increase store revenue, profitability and performance. All the variables will be studied with a difference-in-difference approach using data from before and after the introduction for a treatment and a control group.

It is commonly thought that self-scanners can bring considerable advantages to the retailers that employ them, which explains their fast adoption in retail in recent years. They often require less space than regular cashiers, less personnel, less waiting time and are expected to produce higher customer engagement, as customers who would prefer not to interact with human cashiers flock to them. Disadvantages include higher shoplifting and the challenges arising from substituting human labour with machines. However, while much has been written on these technologies, there are wide gaps in the literature caused by their recent introduction in certain industries as well as the availability of data. There is therefore open space to study how these technologies affect businesses, and a key contribution of my study will be to fill part of this gap. Some studies have been made on the impact of self service (Basker 2017) and related technologies (Basker 2012) on productivity. To my knowledge though, none of these studies on SSTs were performed in retail and instead focus on gas stations. Most of the existing literature on SSTs is instead focused on service quality. Many of these papers try to understand which features of the technology attract customers (Davis 1989; Dabholkar 1996; Marzocchi & Zammit 2006), and sometimes either assume or expect that automated checkouts increase the speed of checkout without performing an empirical analysis to examine whether this is actually the case.

2. Literature Review

2.1 Service quality and Idle time (Speed of checkout, Queueing)

The literature on the effects of SSTs on service quality is rather comprehensive. Many surveys study which factors increase service quality of SSTs and lead consumers to use or avoid them. In particular, three frequently cited papers propose models that describe what will determine the adoption of self service technologies by clients. Davis (1989) proposes *the Technology Acceptance Model* (TAM), while Dabholkar (1996) and Dabholkar, Bobbitt & Lee (2003) create and test the *Attribute-Based Model* and the *Overall-Affect Model*. All three models are supported by survey results and have considerable overlap with regards to what variables they test. The *Technology Acceptance Model* states that the usefulness and ease of use of a technology will determine its adoption. The *Overall Affect Model* instead considers the attitude of consumers towards using technological products and the need for interaction with service employees as its key independent variables.

In particular, the *Attribute-Based-Model* is built on the *Overall Affect Model* and seems to be the most dominant in the literature. It focuses on speed of delivery, enjoyment, control, reliability and ease-of-use as key factors in determining the perception of service quality and thus the adoption of the technology. Speed, control and enjoyment also appear to be drivers of satisfaction according to Anselmsson (2001). Speed or "perceived time taken" seems to be especially important (Bateson 1985, Fernandes & Pedroso 2016). It is commonly accepted in this literature that SSTs have a positive or null effect on the speed of checkout and that a greater time to checkout has a negative effect on service quality and store performance.

Queuing times are also generally not considered to be positive for a stores' performance: as part of their *Theory of Swift, Even flow*, Schmenner & Swink (1998) propose the *Law of bottlenecks*, which states that eliminating or managing bottlenecks improves the productivity of operations. Reducing queuing could thus positively affect the store's productivity, as it may be possible to complete more transactions faster with the same number of employees. There may be exceptions to this statement: Johnson & Jones (2003) for example explain that in exceptional circumstances queuing time can be used by customers to gain information and improve their overall experience as well as productivity. In a McDonalds for example, waiting in line gives customers time to choose the menu and select the offering most suited for them. In this retailers' case however, customers are not gaining information but simply waiting idly in line to the great concern of the management which wished to decrease idle times. As a result, I will assume that longer queuing times are a detriment to the stores' performance in line with the dominant position of the literature.

Nonetheless, very few papers have been written on the actual effect of self-scanning technologies on the speed of checkout. This can be partly explained by the recent introduction of self-scanning technology, and partly by the difficulty of accessing databases where the actual speed of checkouts can be measured with quantitative tools. Store specific characteristics also play a role: the quality of the interface, the specific industry where SSTs are adopted (eg supermarkets, clothing stores, banks) and the actual implementation of the technology will affect the final speed of checkout and make identification in a regression difficult. To my knowledge, only Anselmsson (2001) performs a similar study on the effect of SSTs on speed of checkout.

2.2 Productivity

Productivity is examined in many domains and industries unrelated to SSTs, and it is worth mentioning both some key findings that seem to hold generally as well as those papers that deal with self service and related technologies specifically. Two key papers which surveyed the literature are Bartelsman & Doms (2000) and Syverson (2011). The first identifies several key factors that drive productivity: managerial ability, technology, human capital, international exposure and regulation. All these have been shown to affect productivity, and my focus will of course be on technology. However, the relative importance of these factors is not fully understood, and it is worth noting the role of the others as well to avoid confounding effects in the analysis. Bartelsman subsequently underlines three stylized facts relevant to this paper. First of all, the dispersion across firms of levels of productivity is very high. Secondly, highly productive firms in one period are likely to retain some of that productivity in the following periods. Finally, resource reallocation is responsible for a considerable part of aggregate productivity growth. Syverson (2011) builds on the work of Bartelsman & Doms (2000) and reviews the literature once again, identifying both firm-level and environment-level factors that affect productivity.

Firm-level factors (Syverson, 2011)

Managerial productivity seems to affect productivity through many small complementary practices that managers execute in their daily work, which are insignificant in isolation but add up and reinforce each other. Ichniowski, Shaw, & Prennushi (1997) focus on incentive pay, teams, flexible job assignments, employment security, and training. Hamilton, Nickerson & Owan (2003) find that the introduction of teams in a garment plant increased worker productivity by an average of 14%. Bloom & Van Reenen (2007) perform a much more comprehensive study of managerial productivity find strong correlations between higher-quality management practices with several measures of productivity, including labor productivity and TFP. Moreover, the quality and characteristics of both labour and capital can differ between stores and have a similar effect. With regards to labor, Maliranta & Vainiomäki (2004) find for example that workers which are older and more highly educated are more productive. IT and R&D also affect productivity, although it is sometimes hard to determine in which direction the causality is flowing (whether more R&D increases productivity or more productive firms perform more research). In a case study, Brynjolfsson, McAfee & Sorell. (2008) show that IT increases the speed at which productive practices in one line of business can be replicated in other ones by the same firm. Learning by doing is also an important determinant of productivity. In manufacturing, Benkard (2000) shows how the number of labour hours needed by the airplane manufacturer Lockeed to assemble their aircrafts halved by the 30th plane and halved again by the 100th. The first few units off the line required more than one million person hours (equivalent to three shifts a day of 2,500 workers each for fifty workdays). This was cut in half by the 30th plane and halved again by the 100th. Product innovation can also increase either the number of units produced per input or raise the quality of existing products, thus increasing productivity. On this topic, Bernard, Redding & Schott (2010) show that there is a positive correlation between the number of products a firm produces and TFP. Finally, firm structure is also considered as a possible determinant of productivity.

Environment-level factors (Syverson, 2011)

On the environment level, Syverson (2011) lists spillover effects, competition and regulation as playing an important role in determining productivity. Productive practices of other firms can have spillover effects and raise the productivity of all firms. Moretti (2004) for example finds higher spillovers in plants that are closer from a geographic and technological point of view. While not controlling for this factor may introduce bias in my analysis, due to data limitations I am forced to assume that spillover effects do not differ systematically between the stores examined. If this bias exists, I would expect stores in larger cities with more competitors (eg. Stockholm) to be more productive than those located in less populated areas. This reasoning extends to the role of competition, which I expect to be higher in larger cities. These speculations however are not testable as I do not have access to data on competitors in the area. Regulation also plays a role: Basker (2015) finds that the introduction of scanners in the US lowered prices in cities with looser regulatory regimes by an average of 1.6%, but had no effect on prices in cities with stricter regimes.

The introduction of scanners

The effects of the introduction of scanners in the 70s was studied thoroughly by the literature. Foster, Haltiwanger & Krizan (2006) found that virtually all productivity growth in retailing in the 1990s was due to store entry (chain expansion) and exit. However, it seems that technological innovation also played a role in the growth in productivity. Basker (2012) concluded through a difference-in-differences specification that in the 70s and early 80s in the U.S. productivity increased by an average of 4.5 percent in stores that installed barcode scanners before 1982. The introduction of scanners had small short-run gains in productivity relative to the fixed costs of introduction, but these increased over time. This is in line with Shaw (1977), which concludes that out of 50 firms which adopted scanners early and before 1976 "23 firms … [reported] improved speed due to scanning at the checkout, while 12 … [claimed] unchanged or reduced productivity,… the remaining 15 were indeterminate". Basker (2012) lists three reasons for this gradual increase in productivity. The first is technological improvement (later scanners were able to read smaller barcodes and commit less reading errors). The second is the diffusion of complementary technologies such as barcodes. The third is improvements in employee training. These three variables will all be taken in consideration in my analysis.

The effects of the introduction of scanners on price decreases also varied based on store size. Later (smaller) adopters contributed much less to price decreases than early adopters, in spite of the technological innovation mentioned above. Moreover, Basker (2015) points out that introducing scanners increased the speed of checkout, improved accuracy and caused other productivity enhancing behaviours. For example, stores could change prices more quickly because employees no longer had to apply stickers or memorize prices (now stored in the system and printed on the receipts).

Automated checkouts

The main difference between the introduction of scanners and automated checkouts is that, while the introduction of scanners did not significantly alter the behaviour of employees, automated checkouts completely replace them and require them to perform different tasks. As mentioned above, scanners also changed the tasks performed by employees, but to a much lesser degree. Basker & Klimek (2017) estimate customer productivity in self service gasoline stations and show how ignoring it in measuring employee productivity will bias estimates upwards. While true productivity may rise by introducing self service technologies by increasing the speed of transactions, measured productivity will rise disproportionately more. A graphical representation of this situation is shown in the appendix.

It is also worth mentioning though Johnston, Porter, Cobbold & Dolamore (2000), which in an industry report on productivity in Australia during the 1990s find that the introduction of single operator console controlled self service stations increased labour productivity in petrol stations.

3. Data and descriptive statistics

3.1 Information on the data

The data used in this thesis is obtained from the internal records of a multinational company. The main database employed (*Receipts*) contains receipt-level data on all transactions made in 14 stores in Sweden (3,656,477 observations). Information from this database is used to study the speed of checkout, conversion rates, queuing time and productivity. It includes the beginning and ending time for every transaction, the number of items per transaction, the revenue from the transaction and whether the checkout was completed with automatic or with regular cashiers.

I was also provided with several accompanying databases with store-level information. *Store-level-data* (29,691 observations) had daily data on net sales, the number of employee hours worked, the number of items sold and the number of customers in the store. This last information was gathered by a sensor which counted how many customers entered and exited the store.

Finally, the database *Storeinfo* contained descriptive, time-invariant data on the Swedish stores. It included information on the location of the store, the opening date, the opening hours, the number of floors and the store area (in square meters).

All databases collected data from the 1st of July 2019 to the 29th of February 2020. They also all had as a variable the stores' identification number based on the store's location code, which was used to match and combine them. Moreover, the information on the date of introduction of the SSTs was obtained from the company's management team and manually used to create the relevant variables in all databases.

3.2 Variables

In the receipts database, I use the two variables *Beginning time of transaction* and *Ending time of transaction* to create the time variables *weekday, hour*, *month*, *day-of-the-month* and *day-of-the-year*. These are used to analyse seasonalities. I also construct the variable *speed*, which is the time difference in seconds between the ending and the beginning times of the transaction. *Net sales* contains data on the total revenue resulting from a transaction (after discounts), including in the case that a product was returned to the store and refunded. The database also contains the variables *Number of items sold* as well as the identifier *Automated till,* which takes a value of 1 if a transaction had been made by an SST and 0 otherwise.

The key dummy variable of interest is *SST,* constructed at the store level*:*

$$SST = \begin{cases} 1, & For\ stores\ in\ the\ treatment\ group\ that\ had\ already\ introduced\ SSTs \\ 0, & Otherwise \end{cases}$$

The treatment group includes stores which introduced SSTs and did not take part in renovations during that time. The control group includes the subset of stores which did not introduce SSTs and are comparable to the treatment group. The SSTs were introduced in stores in the treatment groups on different dates: the indicator variable will take this into account and start taking values of 1 from different dates for each store.

The databases *Store-level-data* and *Storeinfo* contain many variables at the store level. *Store-level-data* contains daily data on *Net Sales* (actual sales after discounts), *Number of Items Sold*, *Number of daily Visitors* and *Employee hours* (the sum of all worked hours in a store every day). The *Conversion rate* was calculated as the daily number of items sold over the daily number of customers. In *Storeinfo*, *Closing date* describes the date at which the store was closed (2030-01-01 is coded if the store is still open). *Monday opening hours*, *Tuesday opening hours*, *Wednesday opening hours* and so on are the opening hours of the store for every day of the week. *Concept area* is the size of the total area accessible to customers, while *Store area* describes the total area of the store. *Number of floors* is the number of sales floors.

At the store level, I also created several variables. Firstly, I rename for convenience the unique store identifier in the analysis as Store 1, Store 2 and so on. *Tillratio* was calculated as the number of manual cashiers over the total number of cashiers in a store. It describes the degree of automation of the store: it is equal to 1 for stores that only have manual checkouts, and 0 for stores without human cashiers. *Numberhours* describes the total number of opening hours of a store over the course of a week: it was obtained by summing the number of opening hours every day.

3.3 Measuring store performance

Store performance is a broad term with many possible meanings. My task is to understand the positive and negative consequences of the introduction of the technology for the company, and to do so I will use a range of different measures of store performance. Store performance is thus used as an umbrella term to capture many of the key aspects that are affected by the introduction of the technology and which are likely to affect long-term profitability. The choice of measures to describe store performance is guided by the company and economic theory. Several of my dependent variables are of interest for different branches of the literature, such as *service quality* and *productivity*. They can be classified in three groups: idle time, conversion rates and productivity. I will now present them in turn.

Idle time (speed and queuing time)

Outcomes of great importance that change when SSTs are introduced are related to the idle time and the process of checking out itself. It is of course of value for the business if the time to checkout and the queueing time decrease as this would allow cashiers to service more customers faster and provide them with a service of higher quality. Through quality and the number of transactions, introducing SSTs could thus impact revenue and profitability. It was thus not surprising that these outcomes are of interest to the management of the company, which want to know if the idle time of customers is reduced.

Idle time is difficult to measure, but two ways to study it are analysing the speed of checkout and queuing times. The speed of each individual transaction affects idle time. In this study, the speed of checkout is defined as the time (in seconds) to complete a transaction. The *service quality* literature focuses on this measure, showing evidence on how speed of checkout and service quality are positively related (see Dabholkar, 1996, 2003, Anselmsson, 2001, Fernandes & Pedroso, 2016). In this literature, increasing the speed of checkout is shown to positively affect service quality, consumer satisfaction and in turn profitability. However, most of these papers study the effect of speed on service quality and customer satisfaction through surveys asking customers' opinions without taking part in an empirical analysis to measure the *actual* effect on speed of the introduction of SSTs. My results will therefore be of great interest to this literature, especially with regards to this last assumption.

Moreover, I will also determine the effect of the introduction on the speed to scan an individual item by studying *Speed per item.* This measure is obtained by simply dividing *Speed* by the *Number of items per transaction,* and the advantage of using it is controlling for the size of the transaction.

The second measure of idle time relates to queuing. The company which provided me the data is very interested in knowing whether queueing times are affected by SSTs because it can affect service quality, the number of transactions and profitability. Queuing times are generally not considered to be positive for a stores' performance: as part of their *Theory of Swift, Even flow*, Schmenner & Swink (1998) propose the *Law of bottlenecks*, which states that eliminating or managing bottlenecks improves the productivity of operations. Reducing queuing could thus positively affect the store's productivity, as it may be possible to complete more transactions faster with the same number of employees.

Queuing times are studied by using three measures: the *Service Gap*, the number of daily queues and the proportion of daily transactions which are queues. The *Service Gap* is the time that passes between the end of one transaction and the beginning of the next. This measure requires less manipulation to create (and therefore fewer limiting assumptions), but it harder to interpret. I then define a dummy variable for what a queue is: it takes a value of 1 if the service gap is less than 30 seconds and 0 otherwise. I will also use an alternative definition of queue where the dummy variable takes a value of 1 only if the service gap is less than 10 seconds. These two numbers are arbitrary, as establishing when a queue occurs is of course difficult. Usually, some time passes naturally between one transaction and the next: customers might need to move their items or end a conversation, the employee might look at some of the items before he begins the transaction through the machine etc… There may be also be circumstances in which a queue does not form when the service gap is 0 or it forms even though the service gap lasts more than 30 seconds. For example, a customer could reach the queue just as the previous customer is leaving: in this case the service gap is 0, but no queue had been formed. It would also be possible that an employee is temporarily unable to begin a transaction and thus customers would wait in a queue for more than 30 seconds. In summary, it is preferable to use two definitions of *Queue* to improve the quality of my analysis.

Finally, I calculate the *Proportion of Queues* to the total *Number of Transactions* through the dummies described above. These last two measures are more complex but offer a clearer interpretation. I will thus analyse all 3 and examine my results in order to obtain a better understanding of the effect of the introduction of SSTs on queuing.

Overall, idle time will be considered to have a negative impact on store performance, as slower speed of checkout translates in worse service quality and consumer satisfaction and longer queuing times produce bottlenecks which decrease productivity. Over the hundreds of thousands of transactions that occur every year in the company's many stores, even a small average change in these two measures could have a considerable effect on aggregate.

List of measures:

$$Speed = \ Ending\ time\ of\ transaction - Beginning\ time\ of\ transaction$$

$$Speed\ by\ Item = \ \frac{Speed}{Number\ of\ items\ per\ transaction}$$

$$Service\ Gap = \ Beginning\ time\ of\ transaction_{a+1} - Ending\ time\ of\ transaction_{a}$$

$$Number\ of\ daily\ queues$$

$$Proportion\ of\ daily\ queues = \frac{Number\ of\ daily\ queues}{Total\ Number\ of\ Daily\ Transactions}$$

Conversion rates

Whether introducing SSTs affects conversion rates is an interesting question. As for the previous variables, conversion rates also capture store performance: when they rise, they imply higher sales for the same number of visitors and greater efficiency in convincing customers to perform a purchase. The company's management was especially hopeful that this would be the case and requested that I perform this analysis.

The literature also suggests that some of the factors behind the choice of checkouts over human cashiers include the willingness to interact with employees and whether customers had a positive or negative attitude towards technology. Dabholkar (1996) finds support for these hypotheses (see H7, H8, pp. 36, 44), as does Dabholkar, Bobbitt & Lee (2003) (see H6a, H6b, pp. 82). If a store offers both automated and human transactions as an option, this could result in greater value for customers by providing a wider variety of services. As a result, it is possible that introducing SSTs would provide more value to consumers and increase conversion rates by attracting customers which would prefer not to interact with employees when they shop. While I cannot directly measure if this is true, I will investigate whether there is any evidence that *Conversion Rates* are affected by the introduction of SSTs. The conversion rate is defined as the ratio of the *Number of Items Sold* over the *Total Number of Daily Visitors*. *Conversion Rates* are sometimes calculated as the *Number of Transactions* over the *Total Number of Daily Visitors*: the appendix (section 6.3) contains an empirical analysis on this alternative measure of the *Conversion rate* as well.

List of measures:

$$Conversion\ rate = \frac{Number\ of\ Items\ Sold}{Number\ of\ Daily\ Visitors}$$

Productivity

Finally, productivity is a key measure of performance. Productivity measures the efficiency in production, and it is of interest to economists, policy makers as well as any company which would wish to introduce SSTs. The literature on productivity has still not determined all factors that explain productivity, but it has shown that technological innovations explain part of the variation. Moreover, it has also shown that many of the key determinants of productivity are small changes in the businesses' operations which, jointly have an effect on the business as a whole. This is especially true for management practices (Syverson, 2011). It is thus important for me to study how the businesses operations have been affected by the introduction, and it is possible that changes to the speed of checkout and the queueing time could affect productivity.

Productivity measures take the form of the ratio between an output and an input. With regards to the output, in the literature two kinds of measures are usually adopted: measures of quantity and of revenue. Measures of quantity are rare to obtain and generally preferable, partly because they do not contain information on prices. Measures of revenue however can also be used, as the data on revenue is easier to obtain but there is a key problem to solve: revenue could vary due to changes in prices caused by market power, market conditions or competitor behaviour, all factors unrelated to productivity. It is thus considered preferable to use quantity measures, although it is usually very difficult to obtain them. One factor that makes it harder to use quantity measures in the service industries is the lack of a common unit of measure for quantity.

With regards to the input instead, it is possible to create a measure of labour productivity using the number of hours worked by employees or the number of full-time adjusted employees. The alternative is to study capital productivity or obtain a measure of Total Factor Productivity (TFP) that controls for both capital and labour. Keeping this in mind, I have focused on labour productivity, and since I was kindly provided data on the price of every item sold, it was possible for me to create both measures of revenue and of quantity; I was especially lucky to be able to produce the latter. The measures I use are therefore (a) the *Number of Receipts* over *Employee Hours* and (b) *Net Sales* over *Employee Hours*.

List of measures:

$$Productivity\ (a) = \frac{Number\ of\ Receipts}{Employee\ Hours}$$

$$Productivity\ (b) = \frac{Net\ Sales}{Employee\ Hours}$$

3.4 Descriptive statistics

I will now proceed to describe the raw data provided to me from all databases. The following tables include the summary statistics on the data from all 14 stores for the time period from the 1st of July 2019 to the 29th of February 2020. With the exception of table 4, they were calculated from the database *as is*, without any kind of filtering.

The summary statistics of the database *Receipts* are shown in table 1. *Beginning Time of Transaction and Ending Time of Transaction* contain information on the hour, minute and second in which a transaction began / ended. Together with *Number of Items Sold*, these two variables were used to create *Speed*, *Speed per item* and the *Service Gap.* The sample selected is unfiltered at the transaction level: every row represents a transaction.

Table 1: Summary statistics of *Receipts* (receipt-level data).

	Number of items sold	**Net Sales (SEK)**	**Automated till**	Average price per item (SEK)
Mean	2.39	293.6	0.1	134.39
Median	2	199	0	109.3
Min	1	0.5	0	0.25
Max	231	22788	1	8554.75
Range	230	22787.5	1	8554.50
Standard deviation	2.10	286.03	0.32	94.88
Skewness	4.33	4.45	2.35	3.67
Kurtosis	84.60	82.42	3.55	50.19
5th percentile	1	49.9	0	37.48
25th percentile	1	109.8	0	77.56
75th percentile	3	377	0	169
95th percentile	6	804	1	299
Missing values	208102	208122	1820223	208122
N	3,656,477	3,656,477	3,656,477	3,656,477

Note: The *Number of items sold* is calculated for every transaction, as is *Net sales*. Automated Till is a dummy which takes a value of 1 if a transaction was made with an SST. *Average price per item* is the only constructed variable in this table: it is used to construct *Productivity measure (a)* and as a control for *Productivity measure (b)*. It is shown here as it was never filtered.

Number of items sold has a mean of 2.39 and median of 2, showing that few items are purchased in most transactions. *Net Sales* instead has a mean of 293.6 and a median of 199 and *Average price per item* has a mean of 134.99 and a median of 109.3. The means are greater than the median for all variables and the skewness are positive, suggesting that they are positively skewed. The kurtosis for all three variables is also very high and much greater than 3 (the kurtosis of a standard normal distribution), suggesting the presence of many outliers.

Table 2 shows the data in *Store-level-data.* The *Number of daily Visitors* and the *Number of Items Sold* are used to create *Conversion Rate,* while *Employee hours* is used to create both productivity measures. The sample selected is unfiltered and the data is aggregated at the daily level.

Table 2: Summary statistics of *Store-level-data* (store-level data).

	Net Sales (SEK)	Number of items sold	Number of transactions	Number of visitors	Employee hours
Mean	138371.8	1152	1122	1927	64.98
Median	106943.2	912	974	1474	48.75
Min	-875.7	-1	298	0	0
Max	1570630.2	12580	5436	20416	3323.5
Range	1571505.9	12581	5138	20416	3323.5
Standard deviation	114712.8	935.56	599.79	1613.27	76.92
Skewness	2.857386	2.58	2.11	2.63	3.67
Kurtosis	13.86199	11.36	6.98	11.10	73.67
5th percentile	33155.3	197	500.65	443	0
25th percentile	66857.87	566	729	882	0
75th percentile	170653.59	1453.25	1310	2428	94.25
95th percentile	351027.93	2855	2276	4958.85	204.34
N	29,691	29,691	29,691	29,691	29,691

Note: *Net Sales* and *Number of Items Sold* represent the sum of all values on the same day of the respective variables in table 1. The *Number of Daily Transactions* and the *Number of Daily Visitors* are displayed, as are the hours worked every day in each store (*Employee Hours*).

The means and medians of the variables are 138371.8 and 106943.2 for *Net Sales*, 1152 and 912 for *Number of Items Sold*, 1122 and 974 for *Number of Items Sold,* 1927 and 1474 for *Number of Visitors* and 64.98 and 48.75 for *Employee Hours*. The *Number of Visitors* is of course higher than the *Number of Transactions.* Negative values of *Net Sales* reflect returned items for which the store needs to pay back a transaction's value to a customer, while negative *Number of items sold* reflect coding errors. *Net Sales*, *Number of transactions, Number of items sold, Number of Daily Visitors* and *Employee Hours* are all positively skewed (skewness >0) and with a high number of outliers (kurtosis >3, although the values are not as extreme as with the variables in *Receipts)*. I notice that many values of *Employee Hours* are 0 (until the 25th percentile), which again suggests coding errors or missing values. Aside from these considerations though, there are no missing values in the *Store-level-data* database*.*

Stores

I now examine the stores themselves by studying store characteristics. Below is reported the whole database *Storeinfo*, which summarises this information:

Table 3: S*toreinfo*

Store	Opening date	No of floors	Store area (Sqm)	Concept area (Sqm)	Numberhours	Introduction date SST	Tillratio
Store 1	2012-10-25	2	2659	1718	70	2019-11-19	0.76
Store 2	2015-11-26	2	3433	2018	77	2019-11-26	0.84
Store 3	2018-10-25	1	1544	1035	58	2019-11-12	0.57
Store 4	1995-03-23	2	2445	1630	65	2019-11-28	0.75
Store 5	1974-01-01	2	1864	1174	66	2019-11-26	0.78
Store 6	2010-09-23	2	1705	1005	66		1
Store 7	2009-02-26	2	1764	1174	70		1
Store 8	1974-01-01	2	3192	1751	78.5		1
Store 9	1989-09-21	2	2240	1536	66		1
Store 10	2006-10-25	2	2631	1413	66		1
Store 11	2011-09-22	2	2303	1508	66		1
Store 12	2006-04-09	2	1494	1062	66		1
Store 13	2013-09-06	2	2330	1465	66		1
Store 14	1990-01-01	2	3446	1478	66		1

Note: The opening date of all stores is displayed, as is the *Number of Floors,* the total *Store Area*, the area accessible to customers (*Concept Area*), the total number of opening hours in a week (*Numberhours)* and the date in which the SSTs were introduced (for stores 1-5). All checkouts were introduced in November 2019, but on different dates. *Tillratio* is the share of human cashiers over the overall number of cashiers, and it is equal to 1 for stores that only have human checkouts. This controls for the fact that the 5 stores had differing numbers of manual and automated checkouts.

All 14 stores are similar. They have a similar *Number of Floors* (2)*, Store Area* (between 1494 and 3446 Sqm)*, Concept Area* (between 1035 and 2018 Sqm) *and Numberhours* (between 58 and 78.5). Moreover, all stores had been open for at least 5 years before the data was collected, with the exception of Store 3. Store 3 is also the outlier with respect to the other variables: it has only one floor, the smallest *Concept Area*, the lowest number of opening hours and the highest proportion of automated checkouts (since it has the lowest *Tillratio*).

Store performance measures

Several of my dependent variables are created from the existing ones (as well as *Average price per item*). In the table below, their summary statistics are presented. Constructing these variables required filtering, especially with regards to the ratios *Speed per item, Conversion rate* and the two measures of productivity which would take on infinite values if the denominator is 0. The unfiltered table that summarizes the descriptive statistics of my dependent variables is in the appendix: table 4 instead describes these variables after the filtering. The aggregation level of these variables is either at the transaction level (*Speed, Speed per item, Service gap)* or at the store level (*Number of queues, Proportion of queues, Conversion rate, Productivity measure (a) and Productivity measure (b).*

Table 4: Store performance measures (filtered)

	Speed	Speed per Item	Service gap	Number of queues	Proportion of queues	Conversion rate	Productivity (a)	Productivity (b)
Mean	50.98	26.35	160.6	482.7	0.42	0.57	2.53	7.41
Median	36	20	38	401	0.42	0.57	2.45	7.32
Standard deviation	50.80	25.28	541.51	320.20	0.08	0.10	0.46	0.48
Skewness	3.84	6.63	8.92	1.76	-0.18	0.08	0.35	0.46
Kurtosis	25.43	96.64	102.01	3.96	-0.27	0.06	-0.46	-0.47
Min	0	0	1	53	0.16	0.27	1.11	6.14
5th percentile	11	7.66	8	142	0.28	0.39	1.86	6.75
25th percentile	22	13	17	263.75	0.36	0.50	2.19	7.05
75th percentile	61	31	97	595	0.48	0.63	2.87	7.73
95th percentile	141	65	582	1160	0.56	0.73	3.37	8.29
Max	799	799	9998	2199	0.66	0.98	4.09	9.03
Missing values	0	208075	0	0	0	0	0	0
N	2,179,258	2,179,258	2,159,010	1,936	1,936	1,904	1,923	1,923

Note: The 8 outcome variables shown above are described in detail in section 3.3.

The mean of *Speed* (50.98) is twice than that of *Speed per item* (26.35): this is consistent with the average *Number of Items Sold*, 2.39. The mean (160.6) of the *Service Gap* is much higher than that of the two previous variables, but the median (38) is similar, suggesting the presence of more outliers: this particular distribution is very skewed, even after filtering. The *Number of Queues, Proportion of Queues, Conversion rate* and the two productivity measures instead do not seem to be as skewed. Nonetheless, the skewness of *Number of Queues* and the two productivity measures are still greater than 0 (respectively, they are 1.76, 0.35 and 0.46).

The filtering has a small effect on the descriptive statistics of *Speed*, no effect on *Number of queues* and *Proportion of queues* and a dramatic effect on *Conversion rate, Productivity measure (a), Productivity measure (b)* and the *Service gap* (see appendix).

Study of the dataset

I will now study in greater depth the *store performance measures* in table 4 with the goal of noticing trends and characteristics of the database which will be relevant in constructing variables. All graphs and tables shown use the unfiltered variables unless explicitly stated.

Speed

The speed of checkout was obtained by subtracting the *Ending time of transaction* to the *Beginning time of transaction* for every transaction. The variable *Speed* is thus defined as the number of seconds it takes for a customer to complete a transaction, regardless of the number of items. The density and distribution over time of the speed of checkout is displayed in figure 1 below (unfiltered sample):

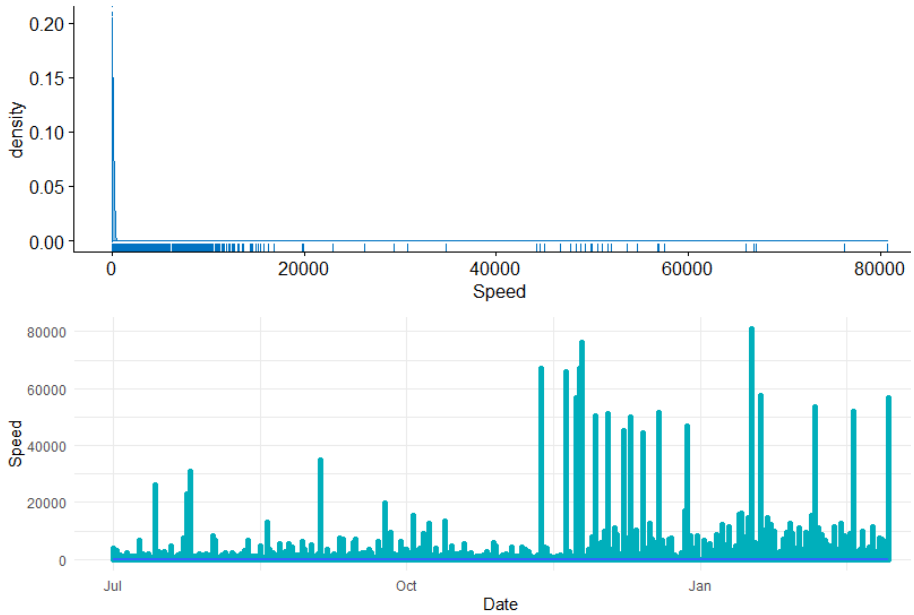


Figure 1: Speed of checkout, density and distribution over time

I notice that *Speed* tends to be rather constant, but towards the end of the year and in early 2020 many outliers appear and the average speed of checkout seems to increase. There is also an extreme amount of outliers, to the point that the box-plot (in the appendix) is not very informative. Many of these outliers are observations of Store 3 (see the density of *Speed* for this store in the appendix). If I instead remove all observations for which the speed of checkout was more than 800 seconds for Stores 1-14, I obtain in figure 2:

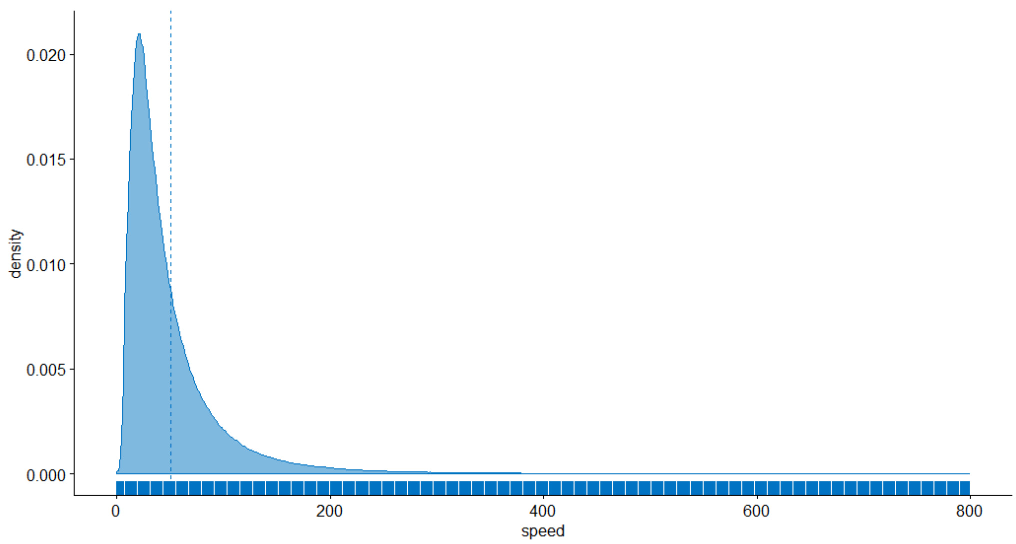


Figure 2: Speed of checkout, filtered

A threshold of 800 seconds seems reasonable, as it is unlikely that a transaction would last longer than 13 minutes. I will thus use this cut-off for outliers in my models.

Speed per Item

The density and distribution over time of this variable (displayed below in figure 3, unfiltered sample) show that it has a very similar behaviour to *Speed*.

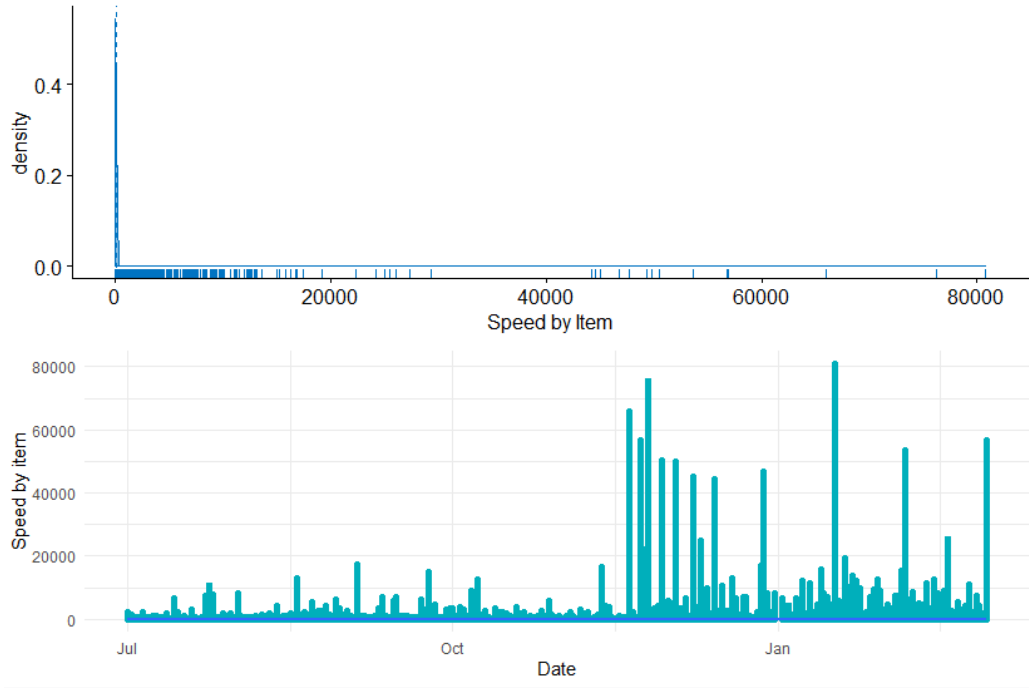


Figure 3: Speed per item, density and distribution over time

Service gap

No data is available to directly measure queueing times. However, it is possible to construct this variable using the proxy *Service Gap*. The *Service Gap* is the amount of time between when a transaction ends and the following starts (at the same cashier). This variable captures whether there is a queue or not: if there is a short time between one transaction and the next then that indicates that a customer was queueing behind them. It can be coded by subtracting the *Beginning Time of Transaction* with the *Ending Time of Transaction* of the previous transaction. [1]

[1] In order to do this, it is first necessary to sort the *Receipts* database by *Store*, *Sales date*, *Number* of *checkout* and *Beginning Time of Transaction*. Then, I lagged the *Beginning time of transaction* and subtracted it to *Ending time of transaction* in the previous period. As I subtract, it is also necessary to control for changes in *stores*, *sales data* and *Number of Checkout* through if

Table 5: Constructing *Service Gap*

	Store	**Beginning time of transaction**	**Ending time of transaction**	**Service gap**
2020-02-29	Store 1	2020-02-29	2019-07-01	= 2020-02-29 - 2019-07-01
2019-07-01	Store 2			

The density and distribution over time of the *Service gap* can be visualized below in figure 4 (unfiltered sample):

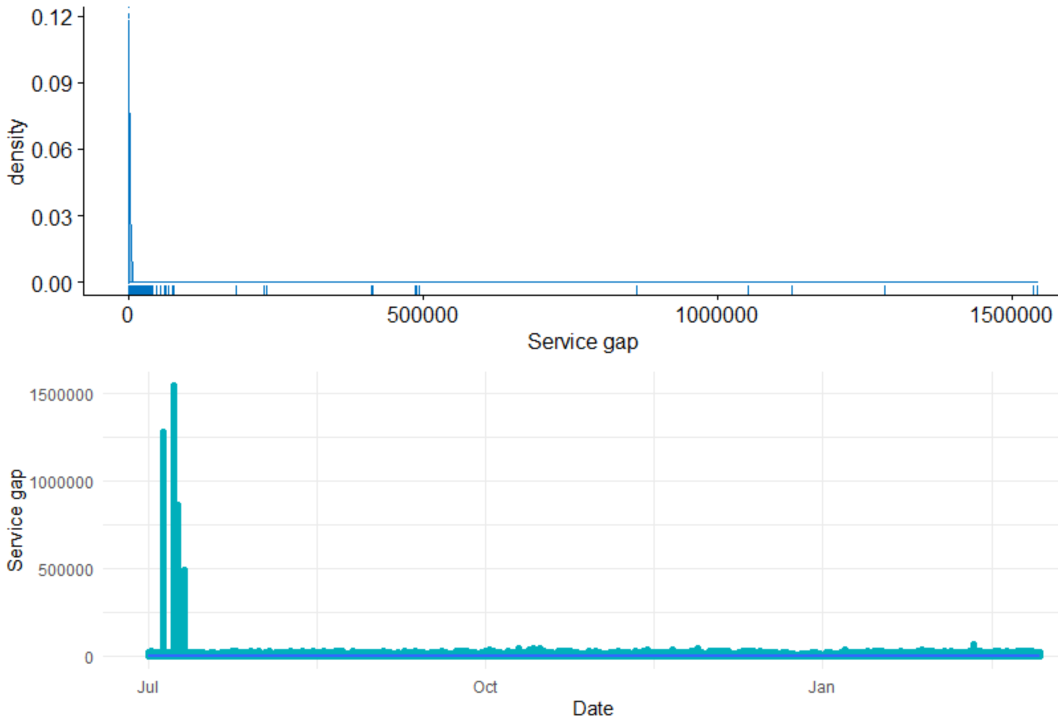


Figure 4: Service gap, density and distribution over time

Creating this variable produces many unintended outliers. Some of these are negative values which are clearly erroneous (all service gaps must be larger than zero), but larger positive outliers are harder to select. It is safe to remove all values larger than 45000 seconds (12.5 hours). After this preliminary step though, I chose to filter out values such that the *Service gap* < 10000 and > 0. I use this continuous measure of the *Service gap* where a shorter time between transactions indicates queues.

In addition, I will also create the dummy variable *Queue* which takes a value of 1 if the time between transactions is less than 30 seconds. *Queue* will instead obtain a value of 0 in all other circumstances, including when the *Service Gap* is 31 and when it is 20000.

$$Queue = \begin{cases} 1, & If\ the\ Service\ Gap\ is\ smaller\ than\ 30\ seconds\ (there\ is\ a\ queue) \\ 0, & If\ the\ Service\ Gap\ is\ larger\ than\ 30\ seconds\ (there\ is\ no\ queue) \end{cases}$$

statements: otherwise, the *Beginning Time of Transaction* of Store 2 on day 1 would be subtracted to the *Ending time of Transaction* of store 1 on day n.

Queue is used to create the variables *Number of queues* and *Proportion of queues*.

Number of queues (daily)

The *Number of Queues* is calculated as the daily number of queues (the situations in which the dummy *queue* is equal to 1). The density and distribution over time of the *Number of Queues* can be visualized below in figure 5:

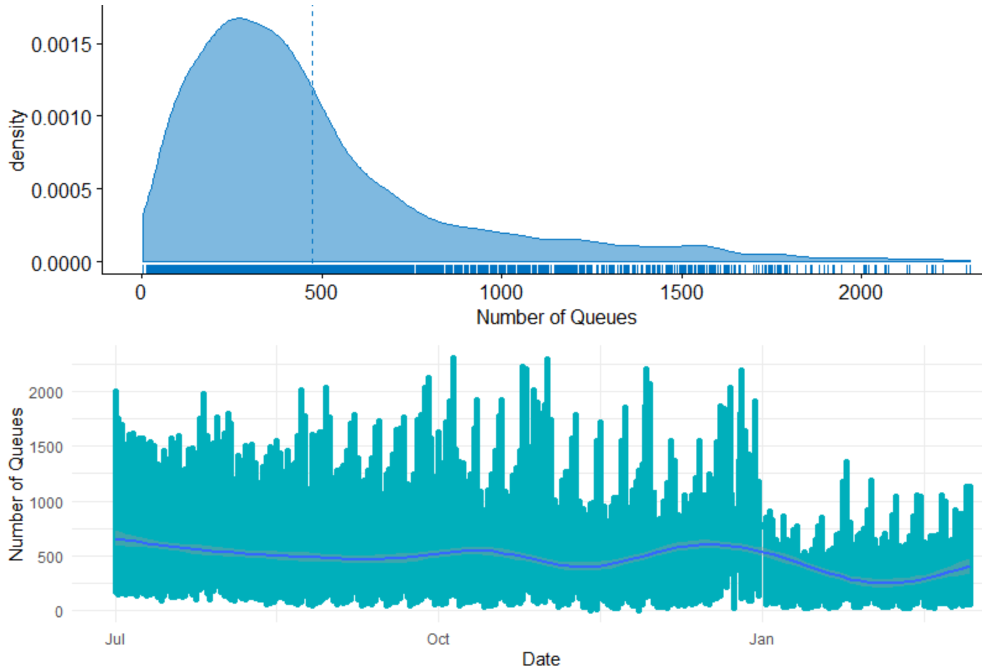


Figure 5: Number of daily queues, density and distribution over time

The graph shows that the *Number of Queues* is approximately constant over the whole sample. The *Service Gap* was filtered beforehand such that it was greater than 0 and smaller than 10000. However, no additional filtering was made.

Proportion of queues (daily)

The *Proportion of Queues* is calculated as the daily proportion of the total transactions which had a queue (the situations in which the dummy *Queue* is equal to 1). The density and distribution over time of the *Proportion of Queues* can be visualized below in figure 6:

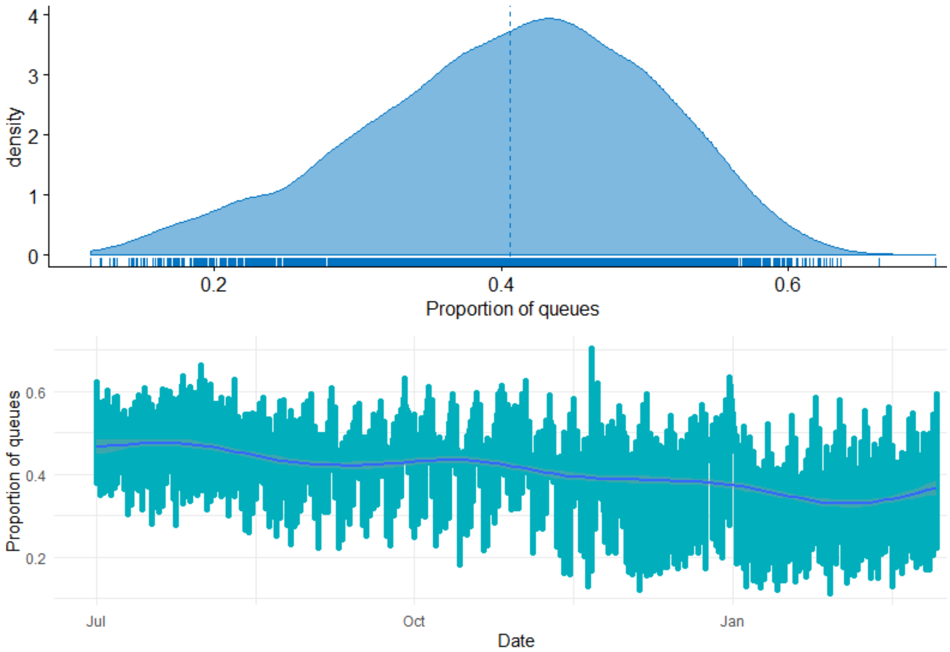


Figure 6: Proportion of queues, density and distribution over time

The graph shows that the proportion of queues over the whole sample falls somewhat over our sample, to then slightly increase in January 2020. The service gap was filtered beforehand such that it was greater than 0 and smaller than 10000. However, no additional filtering was made.

Conversion rate

The (daily) *Conversion Rate* is calculated as the $Number\ of\ Items\ Sold$ / *Number of Daily Visitors*. The *Number of Visitors* (*Daily)* has no missing values and only one value equal to 0. Any value equal to zero will cause the conversion rate to be infinite. As a result, when I calculate the conversion rate, I notice some outliers that do not fall in the range of (0:1). The density and distribution over time of the variable is below in figure 7 (unfiltered sample):

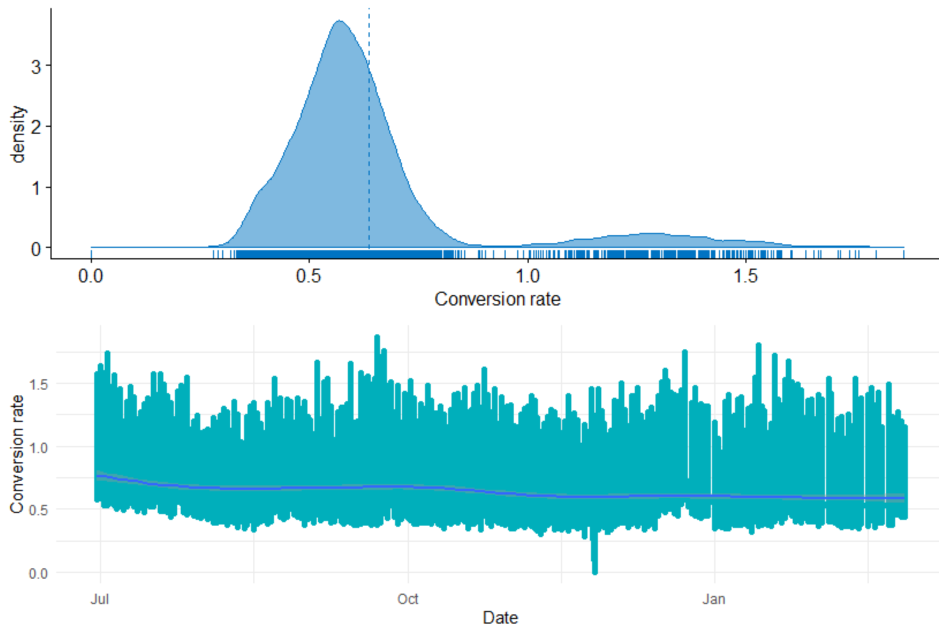


Figure 7: Conversion rate, density and distribution over time (unfiltered)

When I break down the sample store by store, I find that the outliers are distributed over stores 7, 8, 9 and especially 4 (with high values between November and early December, the graph is in the appendix). When all values outside of the unit interval are removed are removed, I obtain the following graphs (figure 8):

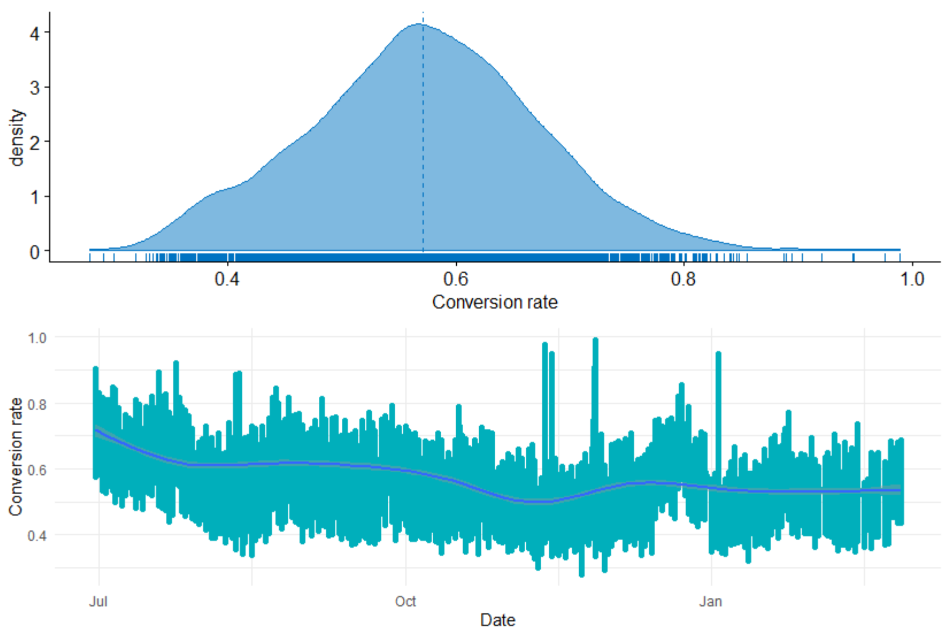
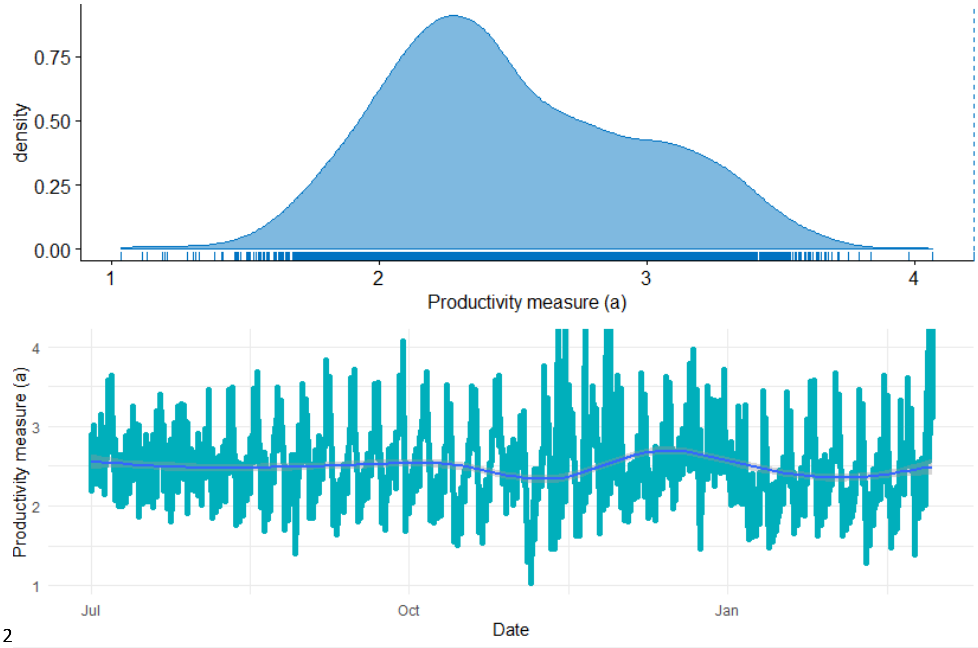


Figure 8: Conversion rate, density and distribution over time (filtered)

The conversion rate seems to be slightly decreasing over time and taking values from 20% to 100%. It also appears that the observations follow a normal distribution, with a mean of 0.6.

Productivity measure (a)

Productivity measure (a) is defined as the *Number of receipts* divided by the *Employee hours*. The density and distribution over time of the variable can be visualized below in figure 9 (unfiltered sample):



[2]

Figure 9: Productivity measure (a), density and distribution over time

Productivity measure (b)

Productivity measure (b) is calculated as follows:

$$Productivity\ (b) = \frac{Net\ Sales}{employee\ hours}$$

This measure has the advantage of being a quantity measure of productivity, which is generally rare and hard to obtain. The density and distribution over time of the variable can be visualized below in figure 10 (unfiltered sample):

[2] The graphs show some extreme values where the productivity measure is above 4. Moreover, R warns that 29 non-finite values were removed from the sample. The reason for this is the lack of filtering of *Employee hours*: the variable takes values of 0 and thus produces infinite productivity values (as *Employee hours* is at the denominator). These values are concentrated in Store 3 (see *Employee hours* in this section).

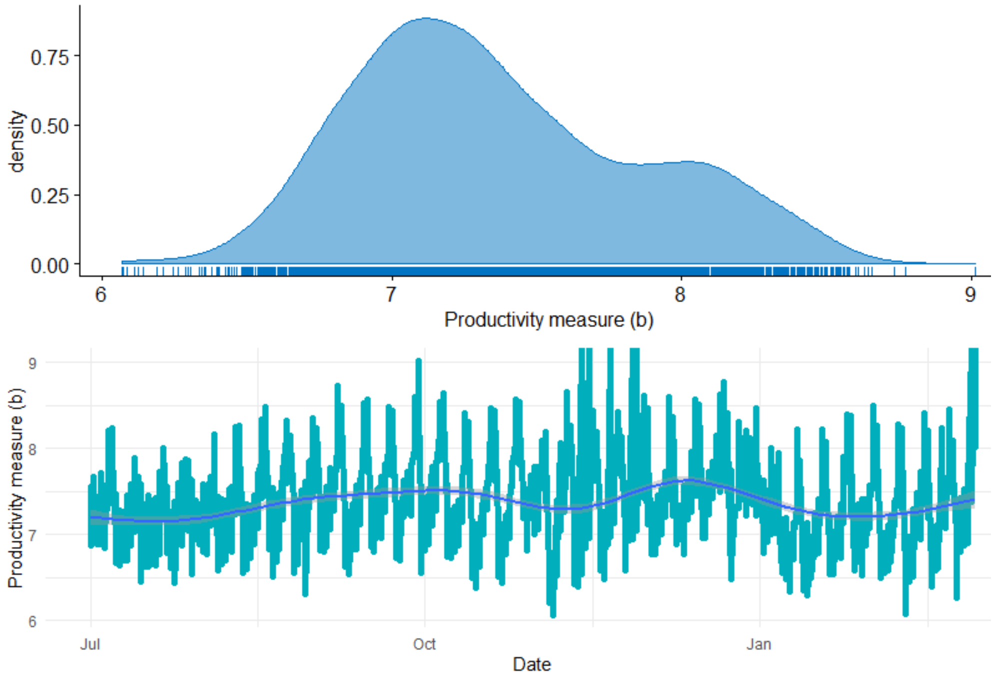


Figure 10: Productivity measure (b), density and distribution over time

This productivity measure is very similar to *productivity measure (a)* in its density as well as its distribution over time (the absolute value of productivity is slightly higher for productivity measure (b)). The considerations made on values of *Employee hours* equal to 0 for *productivity measure (a)* still apply. Moreover, this variable also displays the oscillating behaviour of *productivity measure (a).*

3.5 Store selection: The choice of comparable stores

A key consideration to make in choosing a treatment and control group in a difference-in-difference analysis is whether the stores in the two groups are comparable. To examine whether this is the case, it is worth studying store characteristics such as the *Opening Date*, the *Number of Floors*, the *Store area*, the *Concept area, Numberhours*, the mean number of visitors and the *Average Price Per Item*. Of course, all stores in the treatment group need to have introduced SSTs without undergoing other major renovations at the same time, while stores in the control group need to be entirely manual. On top of this, the dependent variables studied need to exhibit similar behaviour in the control and treatment groups pre-treatment. I have compared pre-treatment averages in this section, while (more importantly) their trends over time are studied in the sections on *descriptive patterns* for each dependent variable (section 4.1.1-4.4.3).

The treatment group can only be comprised by a subset of Stores 1-5: the reason for this is that these stores introduced SSTs without undergoing major renovations. The control group will therefore have to be selected from the remaining stores.

With the possible exception of store 3, stores 1-14 are rather similar: as shown previously, they had all been open for at least 4 years, they had 2 floors, a *Concept area* between 1000 and 2100 and *Numberhours* between 65 and 80. Moreover, with the possible exception of Store 3, the

Introduction date and the *Tillratio* for Stores 1-5 are also very similar. Finally, the proportion of transactions carried out using the SSTs in each of the 5 stores was rather similar. In these stores, SSTs were used in 104,311 transactions out of 564,521, i.e. 18.4% of the total.

I now examine the *Number of daily Visitors* (from *Store-level-data*). The averages for Stores 6-14 are below:

Table 6: Number of visitors

Store	Number of Visitors
Store 6	2232
Store 7	3147
Store 8	6379
Store 9	3211
Store 10	10214
Store 11	4480
Store 12	2403
Store 13	4013
Store 14	4217

The mean *Number of Daily Visitors* for stores 1-5 is 4284. When I examine Stores 6-14 instead (in table 5), I see that Store 10 had a very high *Number of Daily Visitors* (10214) compared to the rest of the stores in both the Treatment and Control groups, almost twice the amount of any other store.

Moreover, when I study Stores 9 and 14 in greater detail, I find that they also use automatic checkouts although, according to the company, renovations took place when they were being introduced. As a result, these stores cannot be included in the treatment or control groups. Another issue arises with Store 8: as it had no data before the 13th of November 2019, it does not have a large quantity of pre-treatment observations.

After all these considerations, I have chosen to exclude Stores 3, 8, 9, 10 and 13 from the analysis. As a result, Stores 1, 2, 4 and 5 will form the Treatment group, while Stores 7, 11, 12 and 13 will form the Control group.

3.6 Filtering

To facilitate the analysis, some of the data had to be filtered away. First of all, the variable *Speed* contained numerous outliers, which reached values of tens of thousands of seconds to complete a transaction. These values are of course unreasonable, and as a result the database was filtered by removing all values of *Speed* above 800.

The only filtering the three queuing variables went through came from selecting only the values of the *Service gap* which were greater than 0 and smaller than 10000.

By definition, *Conversion Rates* are a ratio which takes values between 0 and 1. As a result, all conversion rates which did not belong to this interval were dropped.

Productivity measures also took infinite values due to the denominator *Employee hours* taking values of 0. I thus removed all values of hours = 0.

The only other selection that took place involved choosing the treatment and control groups. In this regard, Store 3 was removed as it contained most outliers of *Speed* and its store characteristics differed from those of all other stores (the *Opening Date*, *Number of Floors*, *Numberhours*…). Finally, Store 8 was also dropped as it did not include data from the 13th of November 2019 and presented most extreme values in *Employee hours* and *Conversion rates*.

Other variables

In the appendix, I examine visually the behaviour of the treatment and control groups of other variables in the database to determine whether they changed at the time of the introduction of SSTs. The variables which did not vary are *Number of items sold (per transaction)*, *Net sales (per transaction), Number of Transactions (daily)* and *Average price (daily)*. I instead find that the *Net sales (daily), Number of items sold (daily)* and *Number of Visitors (daily)* increased faster in the treatment group than in the control group. The number of *Employee hours* in the control group instead rose faster than in the treatment group.

3.7 Differences in means (Treatment group and control group)

I now proceed with the comparison of means before and after the introduction of SSTs for the treatment and the control group. The table below displays this information for all my dependent variables after they had been appropriately filtered (as described in the *filtering* section):

Table 7: Mean comparison for treatment and control groups

	Treatment group		Control group	
	Pre-treatment	Post-treatment	Pre-treatment	Post-treatment
Speed	49.40	54.81	53.19	49.90
Speed per item	24.11	29.03	25.81	25.55
Service Gap	159.3775	203.4012	165.44	163.60
Number of queues	629	512	378	389
Proportion of queues	0.4814	0.3529	0.4275	0.4063
Conversion rates	0.5731	0.5279	0.6107	0.5415
Productivity (a)	2.482	2.548	2.534	2.582
Productivity (b)	7.346	7.413	7.396	7.465

This initial mean comparison suggests a possible increase in *Speed* and *Productivity (a)* and a decrease in *Number of queues* and *proportion of queues*. After the introduction, *conversion rates* fall in both groups, but they fall faster in the control group (by 7%) than the treatment group (by

5%). *Productivity (b)* seems to behave in the same way in both sets of stores (both increase by 0.07).

4 Empirical analysis

In this study, my focus will be on store performance. To examine it, I will use several measures of store performance:

$$Store\ performance \begin{cases} Speed\ of\ checkout \\ Speed\ per\ item \\ Service\ gap \\ Number\ of\ queues \\ Proportion\ of\ queues \\ Conversion\ rates \\ Productivity \end{cases}$$

These measures will be studied in different specifications. I will begin with a univariate regression (1) which does not include controls. I will then add time fixed effects δ_t in specification (2) and store fixed effects α_i in specification (3). Store fixed effects include information on store characteristics that are stable over time. For example, it may capture average employee ability, managerial skill, and capital which can affect many of my dependent variables. The final specification (4) will include all relevant measures.

$$Store\ performance_{it} = \alpha_i + \delta_t + \beta\ SST_{it}\ +\ \varepsilon_{it}$$

Where *i* captures the store or transaction and *t* time. After this first set of models, I will run a second set of models to control for seasonal effects instead of just using monthly effects. The additional variables I used to account for seasonalities include *weekday, month, day-of-the-month* and *day-of-the-year:* they were introduced as dummy variables following the approach of Wooldridge, 2001. I also attempted using a dummy variable to control for all national holidays in Sweden, but I did not display it in my regressions as it was always shown to be insignificant.

In specification (1.1), I include *weekday* effects, in specification (1.2) I add *day-of-the-month* effects and in specification (1.3) I control for *day-of-the-year* effects. Finally, specification (1.4) includes all controls of my final model specification (4) as well as all the time effects described above.

The reason behind my attention to seasonal effects is that the introduction of the technology happened shortly before the holiday season.

For all models with time or store fixed effects, I performed F-tests to determine whether introducing them in my specification was an improvement over the OLS: the values of these tests are included in the text, and more information on every test is in the appendix.

For certain dependent variables, I performed my analysis two times narrowing down the sample. This was done every time that the parallel trends assumption seemed to be violated or no effect could be found.

Identifying assumptions

The identifying assumptions of my difference in difference specifications are the following:

a) Absence of shocks the time of intervention which affected the outcome variable
b) The composition of the treatment and control groups remained stable over time.
c) The treatment and control groups exhibited parallel trends before the intervention
d) There is no omitted variable bias (caused, for example, by seasonalities)

Assumptions a) and b) will be discussed jointly for all outcome variables as they depend on external factors which affect most or all of the outcome variables. c) and d) instead tend to be more variable-specific.

Evidence for the validity of assumption a) was shown in the appendix and section 3.4 (*Descriptive Statistics*) by visually inspecting which variables changed at the time of treatment. Some variables changed more in the treatment group than in the control group after the introduction of SSTs (*Net sales (daily), Number of items sold (daily), Number of Visitors (daily)* and *Employee hours)*, but these changes all occurred during the month of December and after all checkouts had been introduced: it is therefore more likely that they are a result of seasonal effects or of the introduction of the technology. The exception is *Employee Hours*, which increased faster in the control group than in the treatment group and cannot be explained by seasonalities. This might indicate that introducing SSTs had labour saving effects which could influence productivity. As I cannot verify if other unobserved shocks occurred at the time of introduction, I must assume that assumption a) holds.

Assumption b) also seems to hold. The selected stores had all been open for several years prior to the introduction and did not undergo major changes during the time of the analysis. Moreover, I individually checked that all of them were active and had transactions regularly over the time period covered. Details are present in section 3.4 (*Descriptive Statistics*).

Assumption c) will be examined in this section by visually studying the behaviour over time of each dependent variable for the treatment and control groups.

In all my dependent variables I found that seasonalities had to be accounted for, and in many cases the choice of time controls greatly affected my coefficients. I thus made a thorough analysis of how seasonalities could affect my specification to ensure the validity of assumption d).

4.1 Speed of checkout

4.1.1 Descriptive patterns

Figure 11 shows the behaviour of the average number of seconds to complete a transaction over time for the control and treatment groups. The vertical lines represent the first date in which the SSTs were introduced in the first store (the time of the first treatment on November 12, 2019): [3]

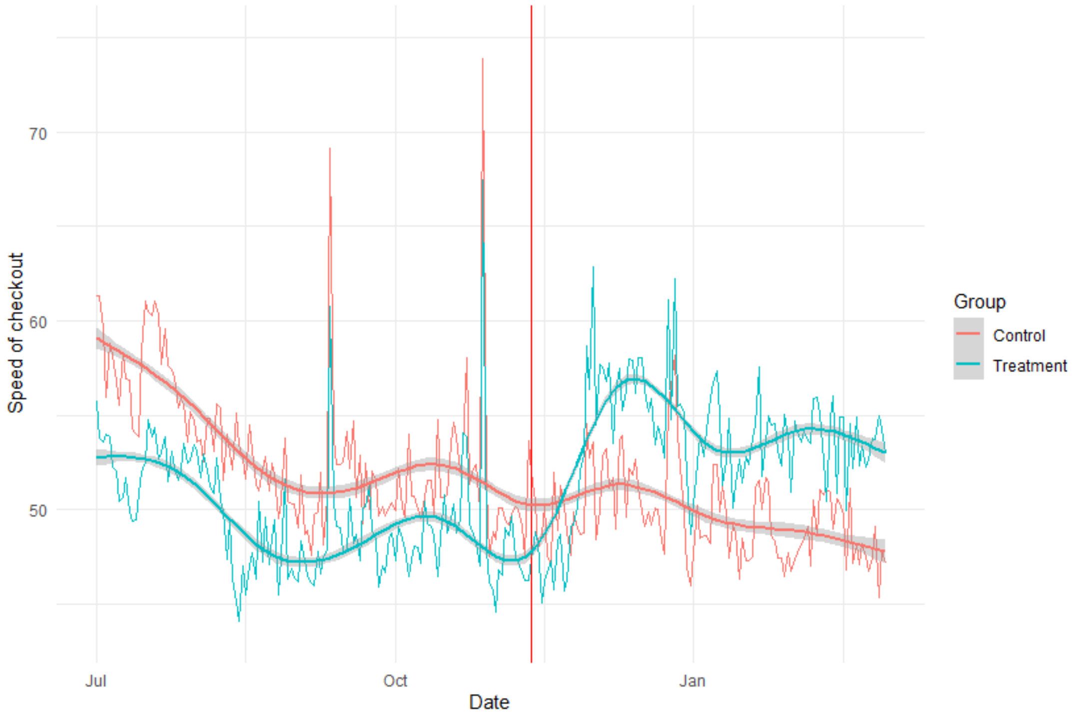


Figure 11: Speed of checkout, distribution over time for treatment group and control group

We can thus see an increase of approximately 10 seconds in the time necessary to perform a transaction only in the stores that had introduced SSTs. This increase occurs during the time of introduction of the technology. Stores I the treatment group were slightly faster than their counterparts before the technology was introduced but became slower afterwards. In summary, the behaviour over time of the control and treatment group shown by the trendlines indicate that the identifying assumption c) (parallel trends) holds.

[3] *Speed* was constructed as the difference between *Beginning time of transaction* and *Ending time of transaction. Speed per item* was easily obtained by dividing *Speed* by the *Number of Items* purchased in every transaction. The only filtering made was removing extreme values larger than 800 seconds: this last transformation is unlikely to greatly affect estimates. See section 3 for further details.

I will now compute the averages before and after the introduction of SSTs for each store for the whole period. I will also compare the *Speed* of manual cashiers and SSTs cashiers within the same store.

Table 8: *Speed*, treatment and control groups

Average Speed of all checkouts		
	Treatment group	
	Pre-treatment	Post-treatment
Store 1	52.5	59
Store 2	49.4	55
Store 4	48.4	58.9
Store 5	46.4	46.7
General	49.4	54.81

Average Speed of all checkouts		
	Control group	
	Pre-treatment	Post-treatment
Store 7	51.1	48
Store 11	53.3	50.7
Store 12	51.6	47.1
Store 13	55.9	52.5
General	53.19	49.9

When I split the statistics by store, I find that in every store *Speed* increased after the introduction in the treatment group and fell in the group. The increase in the treatment group (between 0.3 and 10.5 seconds) seems to be larger than the fall in the control group (between 2.6 and 3.5 seconds). The absolute values of the stores are very similar.

The mean comparison for automated and manual checkouts is displayed below.

Table 9: *Speed*, manual cashiers and SSTs

Average Speed of all checkouts			
	Treatment group		
	Manual (Pre-T.)	SST	Manual (Post-T.)
Store 1	51.3	105	52.6
Store 2	49.5	111	50.0
Store 4	50.8	84.4	48.9
Store 5	43.9	91.7	46.4
General	48.8	100	49.73

Average Speed of all checkouts	
Control group (all checkouts)	
	Manual (Post-T.)
Store 7	48.2
Store 11	50.8
Store 12	47.2
Store 13	52.6
General	50

The tables show that the *Speed* of manual transactions is approximately half of its automated counterpart (50 instead of 100 seconds). This difference is drastic and partially hidden by previous mean comparisons because automated transactions only represent 18.4% of total transactions. Meanwhile, the variation in *Speed* for manual checkouts is very small for all stores and time periods. I thus expect to see a significant positive effect of *SST* on the time to complete a transaction in the treatment group.

4.1.2 Regression analysis

I estimate four initial regression specifications using *speed* as my outcome variable. *Speed* is the number of seconds to complete a transaction and *SST* is an indicator for a transaction occurring in a store which had already introduced self-scanners. Specifications (1) to (3) are as described in the beginning of the section *Empirical Analysis.* In specification (4), I include all controls from models (1) to (3) (time and store fixed effects).

I can assess whether introducing monthly fixed effects and store fixed effects improves my models by employing F-tests, which have as a null hypothesis H0 that the observed and unobserved effects are equal to zero. If the null hypothesis is rejected, the monthly fixed effects or the store fixed effects model is an improvement over the OLS. The p-value of both F-tests is approximately zero, which indicates that the store fixed effects model and the time fixed effects model are an improvement over model (1).

Table 10: Speed of checkout

	OLS (1)	Month FE (2)	Store FE (3)	Final (4)
SST	4.14969*** (0.08051)	5.30486*** (0.1061)	5.39469*** (0.09078)	8.59567*** (0.13402)
Adjusted R^2	0.001217	0.5068	0.5078	0.5086
Time FE	No	Yes	No	Yes
Store FE	No	No	Yes	Yes
N	2,179,258	2,179,258	2,179,258	2,179,258

The standard errors are in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

I also control for seasonal effects as described in the beginning of the section *Empirical Analysis*.

Table 11: Speed of checkout (seasonalities)

	Weekday (1.1)	Day-of-month (1.2)	Day-of-year (1.3)	All time effects (1.4)
SST	4.18675 *** (0.08063)	4.14993*** (0.08098)	4.85084*** (0.10956)	8.21383*** (0.14104)
Adjusted R^2	0.5063	0.00118	0.00089	0.5096
Store FE	No	No	No	Yes
N	2,179,258	2,179,258	2,179,258	2,179,258

The standard errors are in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The F-tests for all these specifications are in the appendix: they all suggest including the time fixed effects.

All coefficients are summarized in Tables 10 and 11. The coefficient of the OLS model (1) is 4.14, while those of models (4) and (1.4) were respectively 8.59 and 8.21. When we compare the coefficients of specifications (1.1) (4.18), (1.2) (4.14) and (1.3) (4.85) we also see that they did not significantly change from the coefficient of the baseline OLS model (1). Although all the F-tests suggested to include as time effects weekday, day-of-month and day-of-year in the final specification, introducing them in specification (1.4) did not significantly alter the coefficient of interest of my specification (4). It thus seems that the month fixed effects included in specifications (2) and (4) already control for seasonal effects appropriately. This reflects in the low R-squared of some of these specifications (namely, (1.2), (1.3)), while it seems that weekday effects have a higher explanatory power. The time effects in (1.1), (1.2) and (1.3) were therefore omitted from the final model to avoid losing an excessive number of degrees of freedom (there are 15 controls in (4) and 289 in (1.4), more details are in the appendix). These additional regressions however show that seasonalities were properly accounted for in my regression. Moreover, all the models are not overfitted. A good rule of thumb is to have a maximum 1 regressor for every 10-20 observations. Model (1.4) has the most regressors (289) and thus requires 5780 observations (289*20).

In conclusion, the final specification (4) indicates that the time to complete a transaction increased by 8.59 seconds on average in stores that introduced SSTs. This result is highly significant with a p-value of 0. A change in the time of transaction of merely 8.59 seconds might seem trivial, but it is not. The average time to complete a transaction in the 8 stores of the treatment and control group before the introduction of checkouts is 50.66 seconds. These 8.59 additional seconds thus represent a 16.9% increase on average in the time to complete a transaction, which is not at all insignificant.

On top of this, the sheer volume of daily transactions occurring leads to a considerable increase in the amount of idle time for customers. The 8 stores in my models recorded 2,179,174 transactions over 243 days. If all transactions had been performed by SSTs, an additional 5205 hours would have been necessary for customers to checkout, resulting in an average of 2.7 hours of customer time lost in each store every day. Because only a small proportion of transactions was carried out by SSTs in our sample (4.4% if we include the control group), the actual time lost in all 8 stores over 98 days is of approximately 230 hours. However, if the adoption rate was higher, the average time to checkout could increase further, even doubling: mean comparison had previously showed that the average *Speed* in the treatment group for SSTs was of 100 seconds, while it was 48.8 for manual checkouts.

4.1.3 Descriptive patterns (Speed per item)

Figure 12 shows the behaviour of the average number of seconds to scan an item for the control and treatment groups. The vertical lines represent the first date in which the SSTs were introduced in the first store (the time of the first treatment on November 12, 2019):

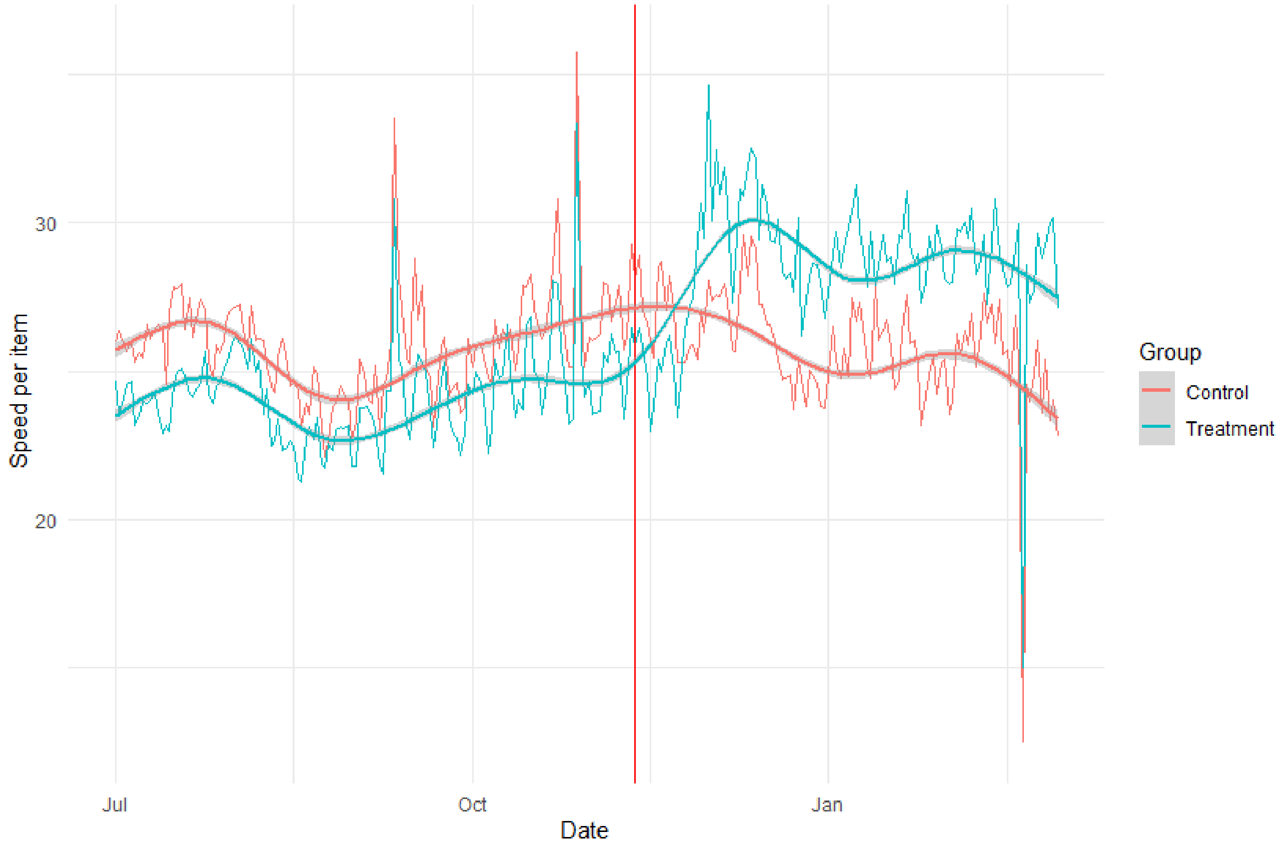


Figure 12: Speed per item, distribution over time for treatment group and control group

I notice an average increase of approximately 6 seconds in the time necessary to scan an item only in the stores that had introduced SSTs. This increase occurs during the time of introduction of the technology. Stores in the treatment group were slightly faster than their counterparts before the technology was introduced but became slower afterwards. Moreover, the behaviour over time of the control and treatment group before the time of introduction shown by the trendlines indicates that the identifying assumption c) (parallel trends) holds.

I will now compute the averages before and after the introduction of SSTs for each store for the whole period. I will also compare the *Speed per Item* of manual cashiers and SSTs cashiers within the same store.

Table 12: *Speed per item*, treatment and control groups

Average Speed per item		
	Treatment group	
	Pre-treatment	Post-treatment
Store 1	24.4	29.6
Store 2	23.4	29.1
Store 4	23.6	31.1
Store 5	25.4	26.9
General	24.11	29.03

Average Speed per item		
	Control group	
	Pre-treatment	Post-treatment
Store 7	25.5	25.3
Store 11	24.7	24.7
Store 12	25.7	24.9
Store 13	27.3	27.1
General	25.81	25.55

Table 13: *Speed per item*, manual cashiers and SSTs

Average Speed per item			
	Treatment group		
	Manual (Pre-T.)	SST	Manual (Post-T.)
Store 1	24.4	48.1	25.3
Store 2	23.4	49	24.4
Store 4	23.6	45.7	24.6
Store 5	25.4	46.3	24.4
General	24.1	47.6	24.7

Average Speed per item	
Control group (all checkouts)	
	Manual (Post-T.)
Store 7	25.4
Store 11	24.7
Store 12	25.4
Store 13	27.2
General	25.71

I find that in every store *Speed per item* increased after the introduction in the treatment group and fell (or remained the same) in the group. The increase in the treatment group (between 1.5 and 6.5 seconds) was larger than the fall in the control group (between 0 and 0.8 seconds). The tables also show that the *Speed per item* of manual transactions is also approximately half (25 seconds) of its automated counterpart (47.6 seconds). I thus expect to see a significant positive effect of *SST* on the time to scan an item in the treatment group.

4.1.4 Regression analysis

I estimate 4 initial regression specifications using *Speed per item* as my outcome variable. *Speed per item* is the average number of seconds to scan an item and *SST* is an indicator for a transaction occurring in a store which had already introduced self-scanners. Specifications (1) to (3) are as described in the beginning of the section *Empirical Analysis.* In specification (4), I include all controls from models (1) to (3) (time and store fixed effects).

I can assess whether introducing monthly fixed effects and store fixed effects improves my models by employing F-tests, which have as a null hypothesis H0 that the observed and unobserved effects are equal to zero. If the null hypothesis is rejected, the monthly fixed effects or the store fixed effects model is an improvement over the OLS. The p-value of both F-tests is approximately 0, which indicates that the store fixed effects model and the time fixed effects model are an improvement over model (1).

Table 14: Speed per item

	OLS (1)	Month FE (2)	Store FE (3)	Final (4)
SST	4.07018***	3.57987***	4.90409***	5.09570***
	(0.03921)	(0.05165)	(0.04423)	(0.06531)
Adjusted R^2	0.005198	0.5365	0.5369	0.5372
Time FE	No	Yes	No	Yes
Store FE	No	No	Yes	Yes
N	2,179,258	2,179,258	2,179,258	2,179,258

The standard errors are in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

I also control for seasonal effects as described in the beginning of the section *Empirical Analysis*.

Table 15: Speed per item (seasonalities)

	Weekday (1.1)	Day-of-month (1.2)	Day-of-year (1.3)	All time effects (1.4)
SST	4.11780***	4.12166***	3.48412***	5.09829***
	(0.03925)	(0.03944)	(0.05333)	(0.0687)
Adjusted R^2	0.5365	0.5364	0.5384	0.5392
Store FE	No	No	No	Yes
N	2,179,258	2,179,258	2,179,258	2,179,258

The standard errors are in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The F-tests for all these specifications are in the appendix: they all suggest including the time fixed effects.

All coefficients are summarized in Tables 14 and 15. The coefficient of the OLS model (1) is 4.07, while those of models (4) and (1.4) both indicated that the time to scan an item increased by 5.09 seconds on average after the introduction of SSTs. As the average time to complete a transaction in the 8 stores of the treatment and control group before the introduction of checkouts is 25.13 seconds, this represents an increase of 20%. This result is highly significant with a p-value of 0. Although all the F-tests suggested to include as time effects weekday, day-of-month and day-of-year in the final specification, introducing them in specification (1.4) did not significantly alter the coefficient of interest of my specification (4). When we compare specifications (1.1), (1.2), (1.3) we also see that their coefficients did not significantly change from the coefficient of the baseline OLS model (1) (4.07). It thus seems that the month fixed effects included in specifications (2) and (4) already control for seasonal effects appropriately. This reflects in the low R-squared of some of these specifications (namely, (1.2), (1.3)), while it seems that weekday effects have a higher explanatory power. The time effects in (1.1), (1.2) and (1.3) were therefore omitted from the final model to avoid losing an excessive number of degrees of freedom (there are 15 controls in (4) and 289 in (1.4), more details are in the appendix). These additional regressions however show that seasonalities were properly accounted for in my regression. Moreover, all the models are not overfitted due to the large sample size (2,179,258).

4.2 Queuing

Analysing variations in the speed of checkout is not sufficient to understand how business operations have been affected by the introduction of SSTs. Although automated scanners increased the average time to complete of transaction in stores in the treatment group by 16.9%, it may still be preferable to introduce them for other reasons. For example, it's certainly possible that although the service is slower queuing times would fall because more machines are available with the same number of employees. I will therefore investigate the behaviour of queuing time.

4.2.1 Descriptive patterns (Service gap)

Figure 13 shows the average *Service gap* in the treatment and control groups over time. The vertical line shows the first date in which the SSTs were introduced in the first store (the time of the first treatment on November 12, 2019):

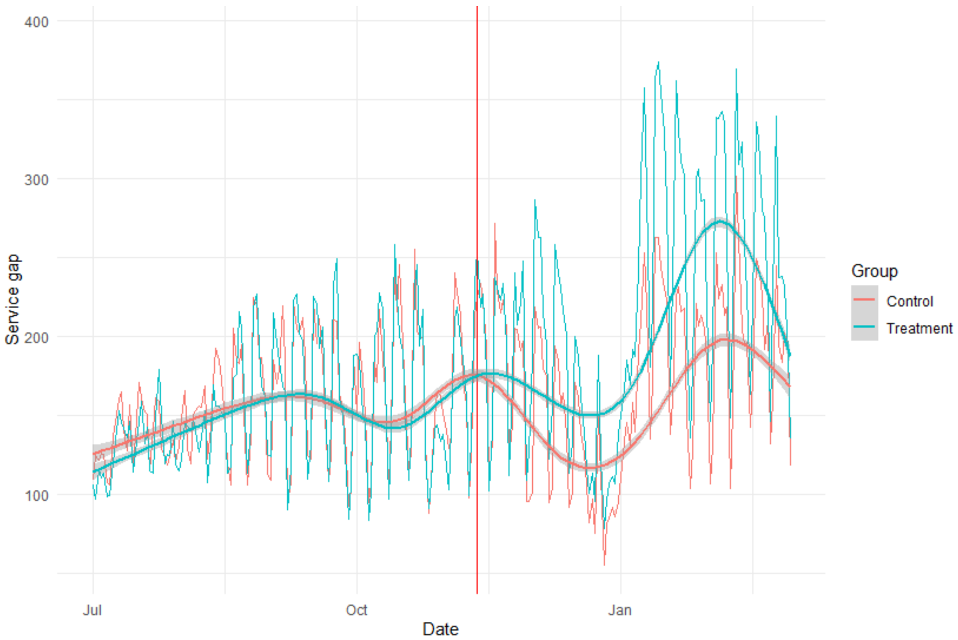


Figure 13: Service gap, distribution over time for treatment group and control group

We can thus see some evidence in support of the identifying assumption c) (parallel trends). The trends of the two variables remain parallel before the introduction of SSTs. After the introduction however, the behaviour of both curves changes, and the *Service gap* of stores in the treatment group increases noticeably faster than in the control group. This would indicate that on average the time between transactions increased and as a result less queues formed.

I will now compute the averages before and after the introduction of SSTs for each store for the whole period. I will also compare the *Service Gap* of manual cashiers and SSTs cashiers within the same store.

Table 16: *Service Gap*, treatment and control group

Average Service Gap		
	Treatment group	
	Pre-treatment	Post-treatment
Store 1	127	176
Store 2	168	215
Store 4	137	190
Store 5	149	183
General	148	192.9

Average Service Gap		
	Control group	
	Pre-treatment	Post-treatment
Store 7	151	145
Store 11	140	137
Store 12	151	161
Store 13	162	162
General	151.1	150

When I split the statistics by store, I notice that *Service Gap* increased in every store in the treatment group and did not vary significantly in the control group: the increase in the treatment group was of an average of 44.9 while the average service gap in the control group fell by 1.1 seconds.

4.2.2 Regression analysis

I estimate 4 initial regression specifications using *Service gap* as my outcome variable. *Service gap* is the number of seconds between transactions in the same cashier and *SST* is an indicator for a transaction occurring in a store which had already introduced self-scanners. Specifications (1) to (3) are as described in the beginning of the section *Empirical Analysis.* In specification (4), I include all controls from models (1) to (3) (time and store fixed effects).

I can assess whether introducing monthly fixed effects and store fixed effects improves my models by employing F-tests, which have as a null hypothesis H0 that the observed and unobserved effects are equal to zero. If the null hypothesis is rejected, the monthly fixed effects or the store fixed effects model is an improvement over the OLS. The p-value of both F-tests is approximately 0, which indicates that the store fixed effects model and the time fixed effects model are an improvement over model (1).

Table 17: Service gap

	OLS (1)	Month FE (2)	Store FE (3)	Month and store (4)
SST	44.083***	-9.512**	44.819***	-5.145
	(0.9831)	(3.18)	(0.983)	(3.187)
Adjusted R^2	0.001552	0.00465	0.002416	0.005516

Time FE	No	Yes	No	Yes
Store FE	No	No	Yes	Yes
N	2,159,010	2,159,010	2,159,010	2,159,010

The standard errors are in parentheses; * p < 0.10, ** p <0.05, *** p <0.01

I now control for seasonal effects as described in the beginning of the section *Empirical Analysis*.

Table 18: Service gap (seasonalities)

	Weekday (1.1)	Day-of-month (1.2)	Day-of-year (1.3)	Final (1.4)
SST	47.42***	48.5493***	15.789520*	35.883***
	(0.9842)	(0.9954)	(6.822021)	(6.896)
Adjusted R^2	0.0048	0.0028	0.0128	0.01399
Store FE	No	No	No	Yes
N	2,159,010	2,159,010	2,159,010	2,159,010

The standard errors are in parentheses; * p < 0.10, ** p <0.05, *** p <0.01

The coefficient of the OLS model (1) is 44.08, while those of models (4) and (1.4) were respectively -5.14 and 35.88. With the exception of model (4), all coefficients are highly significant. The estimates vary somewhat between models, but five specifications out of eight have coefficients between 48.5 and 35.8 and six out of eight show a positive effect. Overall, it seems that model (1.4) is the best and that the introduction of SSTs increased the *Service gap* by approximately 35.8 seconds. The reason behind my selection of this model is that it includes all controls when the coefficient of *SST* was significant in specifications (1) to (3) and (1.1) to (1.3). Moreover, all F-tests suggested to keep all time and store effects. Finally, this model has the highest explanatory power. Given the very high value of N (2,159,010), it is safe to say that all models are not overfitted.

Nonetheless, there remain some problems with these specifications. The Adjusted R squared of every model however is very low, suggesting that my controls have low explanatory power. Moreover, the interpretation of the coefficients with regards to queuing is somewhat difficult: an increase in the service gap shows that queues are less likely to occur, but it does not explain by how much.

4.2.3 Descriptive patterns (Number of Queues)

As they were created from the *Service gap,* these two measures require that all identifying assumptions made for that variable *Service gap* hold.

For the identification assumption d) to hold it is necessary to assume that my indicator *Queue* describes when queues occur. Using two definitions of *Queue* will thus make this more likely. The results of this analysis are similar, although the effects found are of different sizes.

Figure 14 shows the average *Number of queues* over time for the treatment and control groups. The vertical line shows the first date in which the SSTs were introduced in the first store (the time of the first treatment on November 12, 2019):

Figure 14: Number of queues, distribution over time for treatment group and control group

The graph shows evidence in favour of identifying assumption c) (parallel trends). The trends of the control and treatment groups were very similar until after the introduction, when the two curves converged considerably. This would show that the number of queues decreased over time and decreased faster in the treatment than in the control group.

I will now compute the averages before and after the introduction of SSTs for each store for the whole period.

Table 19: *Number of Queues*, treatment and control group

Average Number of queues		
	Treatment group	
	Pre-treatment	Post-treatment
Store 1	754	615
Store 2	905	658
Store 4	381	276
Store 5	487	484
General	629.5	512

Average Number of queues		
	Control group	
	Pre-treatment	Post-treatment
Store 7	368	394
Store 11	481	505
Store 12	255	254
Store 13	408	404
General	378	389.2

I notice that the *Number of queues* decreases in the treatment group and increases in the control group (except for Store 12 and Store 13, where they remain constant). The fall in the treatment group is very noticeable, with the number of queues falling by an average of 117. Generally speaking, the means fluctuate significantly both between stores and before and after the introduction.

4.2.4 Regression analysis

I estimate 4 initial regression specifications using *Number of queues* as my outcome variable. $Number\ of\ queues$ is the number of daily transactions with a queue and *SST* is an indicator for a transaction occurring in a store which had already introduced self-scanners.

Specifications (1) to (3) are as described in the beginning of the section *Empirical Analysis.* In specification (4), I include all controls from models (1) to (3) (time and store fixed effects).

I can assess whether introducing monthly fixed effects and store fixed effects improves my models by employing F-tests, which have as a null hypothesis H0 that the observed and unobserved effects are equal to zero. If the null hypothesis is rejected, the monthly fixed effects or the store fixed effects model is an improvement over the OLS. The p-value of both F-tests is approximately 0, which indicates that the store fixed effects model and the monthly fixed effects model are an improvement over model (1).

Table 20: Number of queues

	OLS (1)	Month FE (2)	Store FE (3)	Final (4)
SST	36.328* (18.365)	136.89*** (21.17)	-123.59*** (17.54)	-100.489*** (21.942)
Adjusted R^2	0.001503	0.7365	0.7892	0.8227
Time FE	No	Yes	No	Yes
Store FE	No	No	Yes	Yes
N	1,936	1,936	1,936	1,936

The standard errors are in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

I now control for seasonal effects as described in the beginning of the section *Empirical Analysis*.

Table 21: Number of queues (seasonalities)

	Weekday (1.1)	Day-of-month (1.2)	Day-of-year (1.3)	All time effects (1.4)
SST	31.75	31.65	115.65***	-145.294***
	(16.52)	(17.91)	(16.60)	(12.421)
Adjusted R^2	0.7412	0.7103	0.8426	0.8457
Store FE	No	No	No	Yes
N	1,936	1,936	1,936	1,936

The standard errors are in parentheses; * p < 0.10, ** p <0.05, *** p <0.01

The F-tests for all these specifications are in the appendix: they suggest to include all time controls.

The coefficients vary considerably between specifications. Nonetheless, once *Store Fixed Effects* are introduced they tend to be between -100 and -140. In the final models (4) and (1.4), the coefficients were respectively -100.48 and -145.29.

The most adequate model is (4): the reason for this is that model (1.4) includes a total of 296 controlling variables, which pose the risk of overfitting in my sample of 1936 observations (296 x 20= 5980 observations would be required). Model (4) instead has sufficient data (19 controls x 20 = 380). Moreover, it is in general not advisable to include a high number of controls in a model unless they change the coefficient of interest to avoid losing degrees of freedom: while the coefficients of (4) and (1.4) are not the same, the additional explanatory power gained not be worth adding approximately 200 controls. As a result, I will adopt the more conservative approach and consider model (4), which shows the smallest effect.

Narrowing the definition of *Queue*

When I repeat this analysis with a narrower definition of *Queue*, I find similar results. I now consider *Queue* as:

$$Queue = \begin{cases} 1, & If\ the\ Service\ Gap\ is\ smaller\ than\ 10\ (there\ is\ a\ queue) \\ 0, & If\ the\ Service\ Gap\ is\ larger\ than\ 10\ (there\ is\ no\ queue) \end{cases}$$

Table 22: Number of queues

	OLS (1)	Month FE (2)	Store FE (3)	Final (4)
SST	-6.743	22.116***	-50.225***	-41.567***
	(3.970)	(4.551)	(3.606)	(4.511)
Adjusted R^2	0.0009728	0.7183	0.7937	0.8265
Time FE	No	Yes	No	Yes
Store FE	No	No	Yes	Yes
N	1,936	1,936	1,936	1,936

The standard errors are in parentheses; * p < 0.10, ** p <0.05, *** p <0.01

Table 23: Number of queues (seasonalities)

	Weekday (1.1)	Day-of-month (1.2)	Day-of-year (1.3)	All time effects (1.4)
SST	-7.640*	-7.467	19.47***	-48.7147***
	(3.718)	(3.896)	(3.89)	(2.9815)
Adjusted R^2	0.7105	0.6827	0.8	0.9275
Store FE	No	No	No	Yes
N	1,936	1,936	1,936	1,936

The standard errors are in parentheses; * p < 0.10, ** p <0.05, *** p <0.01

I notice that the behavior of the coefficients is very similar to previous models. However, the effects found are considerably smaller. The coefficient of the OLS model (1) is -6.74, while those of models (4) and (1.4) were respectively -41.56 and -48.71. The signs of the various models are also more consistent as every specification with the exception of (2) and (1.3) have a negative sign. There is also less difference between the coefficients of models (4) and (1.4).Once again, F-tests suggest to include all time and store effects, and as the sample is unchanged, the considerations made on overfitting do not vary as I vary the definition of *Queue*.

In conclusion, there thus seems to be no uncertainty on whether *SSTs* had an effect on the number of queues. Depending on the definition of *Queue* however, the magnitude of the effect will differ. If *Queue* is defined broadly, this reduction was of 100 daily queues (on average). If *Queue* is defined narrowly, the decrease was of 41 daily queues (on average).

4.2.5 Descriptive patterns (Proportion of Queues)

Figure 15 shows the average proportion of queues in the treatment and control groups over time. The vertical line shows the first date in which the SSTs were introduced in the first store (the time of the first treatment on November 12, 2019):

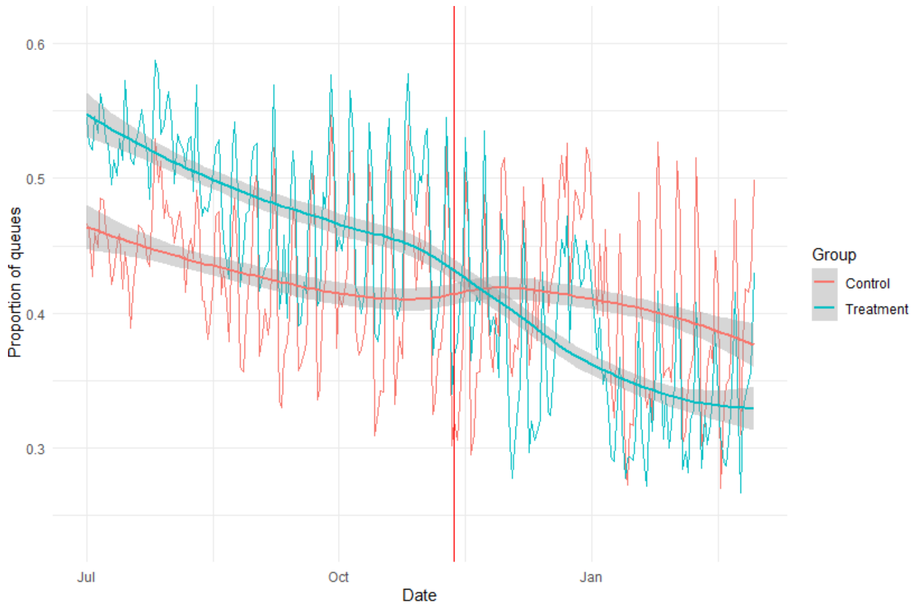


Figure 15: Proportion of queues, distribution over time for treatment group and control group

We can thus see some evidence in support of the identifying assumption c) (parallel trends). The trends of the two variables remain parallel for most of the sample before the introduction of SSTs. Shortly before the introduction though, the two curves begin to converge and the distance between the two had halved on the 12th of December. After the introduction however, the behaviour of both curves changes: the proportion of daily queues increases in the control group but decreases faster than before in the treatment group. These changes are most pronounced immediately after the time of introduction of the technology.

I will now compute the averages before and after the introduction of SSTs for each store for the whole period.

Table 24: ***Proportion of Queues*****, treatment and control group**

Average Proportion of queues		
	Treatment group	
	Pre-treatment	Post-treatment
Store 1	0.516	0.405
Store 2	0.485	0.327
Store 4	0.492	0.295
Store 5	0.433	0.379
General	0.4814	0.3529

Average Proportion of queues		
	Control group	
	Pre-treatment	Post-treatment
Store 7	0.44	0.428
Store 11	0.454	0.434
Store 12	0.402	0.375
Store 13	0.413	0.389
General	0.4275	0.4063

I find that in every store *Proportion of queues* fell after the introduction in both groups. In the treatment group this fall (by an average of 13%) seems more significant than in the control group (by an average of 2%).

4.2.6 Regression analysis

I estimate 4 initial regression specifications using *proportion of queues* as my outcome variable. P$roportion\ of\ queues$ is the number of daily transactions with a queue over the total number of daily transactions and *SST* is an indicator for a transaction occurring in a store which had already introduced self-scanners. The definition for what constitutes a queue is the following:

$$Queue = \begin{cases} 1, & If\ the\ Service\ Gap\ is\ smaller\ than\ 30\ (there\ is\ a\ queue) \\ 0, & If\ the\ Service\ Gap\ is\ larger\ than\ 30\ (there\ is\ no\ queue) \end{cases}$$

Specifications (1) to (3) are as described in the beginning of the section *Empirical Analysis.* In specification (4), I include all controls from models (1) to (3) (time and store fixed effects).

I can assess whether introducing monthly fixed effects and store fixed effects improves my models by employing F-tests, which have as a null hypothesis H0 that the observed and unobserved effects are equal to zero. If the null hypothesis is rejected, the monthly fixed effects or the store fixed effects model is an improvement over the OLS. The p-value of both F-tests is approximately 0, which indicates that the store fixed effects model and the time fixed effects model are an improvement over model (1).

Table 25: Proportion of queues

	OLS (1)	Month FE (2)	Store FE (3)	Final (4)
SST	-0.0922*** (0.00451)	-0.0539*** (0.00508)	-0.131415*** (0.004648)	-0.103963*** (0.00584)
Adjusted R^2	0. 1774	0.32	0.972	0.9762
Time FE	No	Yes	No	Yes

Store FE	No	No	Yes	Yes
N	1,936	1,936	1,936	1,936

The standard errors are in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

I now control for seasonal effects as described in the beginning of the section *Empirical Analysis*. The results are summarized in the table below:

Table 26: Proportion of queues (seasonalities)

	Weekday (1.1)	Day-of-month (1.2)	Day-of-year (1.3)	All time effects (1.4)
SST	-0.0936***	-0.093025***	-0.00574***	-0.111826***
	(0.00390)	(0.00437)	(0.00362)	(0.003494)
Adjusted R^2	0.9739	0.9674	0.9858	0.9919
Store FE	No	No	No	Yes
N	1,936	1,936	1,936	1,936

The standard errors are in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The F-tests for all these specifications are in the appendix: all tests suggest including the seasonal effects.

The coefficient of the OLS model (1) is -0.09, while those of models (4) and (1.4) were respectively -0.10 and -0.11. All specifications have highly significant coefficients which vary very slightly between models: seven out of eight models find an effect between -9.2% and -13.1%. The most adequate model is (4). The main reason for this is that model (1.4) includes a total of 297 controlling variables, which pose the risk of overfitting in my sample of 1936 observations (297 x 20= 6000 observations would be required). Model (4) instead has an acceptable number of controls (19 controls x 20 = 380). It is in general not advisable to include a high number of controls in a model unless their introduction changes the coefficient of interest to avoid losing degrees of freedom. Finally, as both models are highly significant and contain similar coefficients, it also seems reasonable to be conservative and consider the one with the coefficient that displays the smallest effect.

Narrowing the definition of *Queue*

When I repeat this analysis with a narrower definition of *Queue*, I find similar results. I now consider *Queue* as:

$$Queue = \begin{cases} 1, & If\ the\ Service\ Gap\ is\ smaller\ than\ 10\ (there\ is\ a\ queue) \\ 0, & If\ the\ Service\ Gap\ is\ larger\ than\ 10\ (there\ is\ no\ queue) \end{cases}$$

Table 27: Proportion of queues

	OLS (1)	Month FE (2)	Store FE (3)	Final (4)
SST	-0.02771*** (0.001447)	-0.0126*** (0.001616)	-0.044049*** (0.001408)	-0.034038*** (0.001769)
Adjusted R^2	0. 1588	0.3191	0.942	0.9508
Time FE	No	Yes	No	Yes
Store FE	No	No	Yes	Yes
N	1,936	1,936	1,936	1,936

The standard errors are in parentheses; * p < 0.10, ** p <0.05, *** p <0.01

Table 28: Proportion of queues (seasonalities)

	Weekday (1.1)	Day-of-month (1.2)	Day-of-year (1.3)	All time effects (1.4)
SST	-0.028*** (0.001377)	-0.027794*** (0.001429)	-0.012352*** (0.001453)	-0.03460*** (0.001476)
Adjusted R^2	0.9267	0.9213	0.9485	0.9672
Store FE	No	No	No	Yes
N	1,936	1,936	1,936	1,936

The standard errors are in parentheses; * p < 0.10, ** p <0.05, *** p <0.01

The coefficient of the OLS model (1) is -0.02, while those of models (4) and (1.4) were both -0.03. I notice that the behavior of the coefficients is very similar to previous models: this time, all coefficients are rather similar and between -1.23% and -4.4%. However, the effects found are considerably smaller. The coefficients of models (4) and (1.4) are almost identical. Changing the definition of queue seems to have the same effect on *Number of Queues* and *Proportion of Queues*. Once again, the F-tests suggest to include all time and store effects, and as the sample is unchanged, the considerations made on overfitting do not vary as I vary the definition of *Queue*.

In conclusion, there thus seems to be no uncertainty on whether *SSTs* had an effect on the *Proportion of Queues*. Depending on the definition of *Queue* however, the magnitude of the effect will differ. If *Queue* is defined broadly, this reduction will be of 10.3% in the treatment group. If *Queue* is defined narrowly instead, the decrease will be of 3.4% of daily queues (on average).

4.3 Conversion rate

4.3.1 Descriptive patterns

On top of this, it is important to underline that the ratio used to describe *Conversion Rates* is the *Number of Items Sold / Number of Daily Visitors*. This ratio was selected upon suggestion of the company because its alternative *Number of Transactions / Number of Daily Visitors* does not respect the identifying assumption c) (parallel trends). I have nonetheless calculated all coefficients for this second ratio: they can be examined with the relevant graphs in the appendix. These two measures are quite similar and, while the coefficients calculated are different, the conclusions made would have been the same.

Figure 16 shows the average *Conversion rate* over time for the treatment and control groups. The vertical line shows the first date in which the SSTs were introduced in the first store (the time of the first treatment on November 12, 2019):

Figure 16: Conversion rate, distribution over time for treatment group and control group

The graph suggests that while the stores that had introduced SSTs had a comparatively lower conversion rate than their counterparts, the conversion rate level stabilized after the introduction.

The difference between the *Conversion rate* of the control group and the treatment group remains constant until early October, but then decreases to 0 before the first introduction of the technology on November 12. Both conversion rates decrease over this period, but the conversion rate of stores in the control group fall faster than those of the treatment group. The parallel trends assumption c) does not seem to hold.

I will now compute the averages before and after the introduction of SSTs for each store for the whole period.

Table 29: *Conversion Rate*, treatment and control group

Average Conversion rate		
	Treatment group	
	Pre-treatment	Post-treatment
Store 1	0.644	0.590
Store 2	0.618	0.572
Store 4	0.632	0.543
Store 5	0.418	0.403
General	0.5731	0.5279

Average Conversion rate		
	Control group	
	Pre-treatment	Post-treatment
Store 7	0.632	0.568
Store 11	0.605	0.552
Store 12	0.618	0.550
Store 13	0.588	0.497
General	0.6107	0.5415

I find that in every store the *conversion rate* fell after the introduction. In the control group this fall seems more significant in the control group (by an average of 5%) than in the treatment group (by an average of 6%).

Moreover, from late December onwards the conversion rate of the control group increase while those of the treatment group remain constant. This indicates that if the parallel assumption held and SSTs did have an effect, they would negatively affect conversion rates.

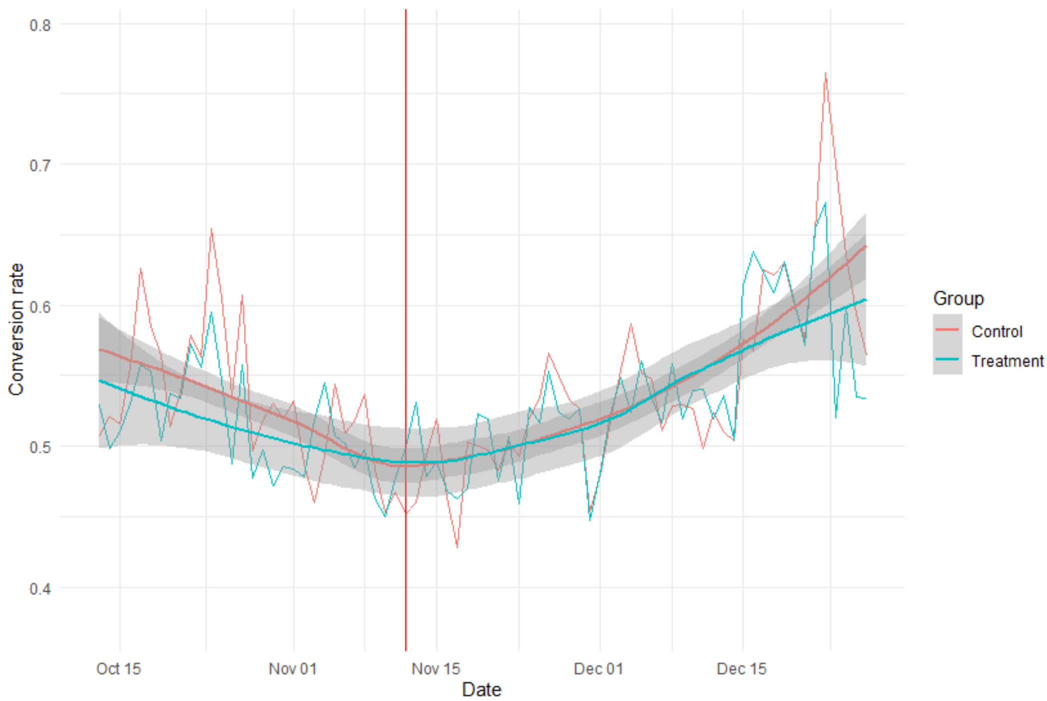


Figure 17: Conversion rate, distribution over time for treatment and control groups,narrow sample

When the sample is restricted to the period from the 12th of October to the 28th of December in figure 17 (maintaining one month of data before the first introduction and after the last one) we obtain the same results. While assumption c) seems more plausible, in early November conversion rates fall inexplicably for the control group.

4.3.2 Regression analysis

I now estimate 4 initial regression specifications using *Conversion rates* as my outcome variable. *Conversion rates* is the number of items sold over the number of daily visitors entering a store and *SST* is an indicator for a transaction occurring in a store which had already introduced self-scanners. Specifications (1) to (3) are as described in the beginning of the section *Empirical Analysis.* In specification (4), I include all controls from models (1) to (3) (time and store fixed effects).

I can assess whether introducing monthly fixed effects and store fixed effects improves my models by employing F-tests, which have as a null hypothesis H0 that the observed and unobserved effects are equal to zero. If the null hypothesis is rejected, the monthly fixed effects or the store fixed effects model is an improvement over the OLS. The p-value of both F-tests is approximately 0, which indicates that the store fixed effects model and the monthly fixed effects model are an improvement over model (1).

Table 30: Conversion rate (whole sample)

	OLS (1)	Month FE (2)	Store FE (3)	Final (4)
SST	-0.04524***	0.044073	-0.04934***	0.013504
	(0.00734)	(0.024816)	(0.004391)	(0.013197)
Adjusted R^2	0.03924	0.9681	0.987	0.9911
Time FE	No	Yes	No	Yes
Store FE	No	No	Yes	Yes
N	1,904	1,904	1,904	1,904

The standard errors are in parentheses; * p < 0.10, ** p <0.05, *** p <0.01

I now control for seasonal effects as described in the beginning of the section *Empirical Analysis*.

Table 31: Conversion rates (seasonalities, whole sample)

	Weekday (1.1)	Day-of-month (1.2)	Day-of-year (1.3)	All time effects (1.4)
SST	-0.04542***	-0.04613***	0.08477	0.01841
	(0.00733)	(0.00738)	(0.04498)	(0.0177)
Adjusted R^2	0.9637	0.9634	0.9637	0.9945
Store FE	No	No	No	Yes
N	1,904	1,904	1,904	1,904

The standard errors are in parentheses; * p < 0.10, ** p <0.05, *** p <0.01

The coefficient of the OLS model (1) is -0.04, while those of models (4) and (1.4) were both 0.01. The final specification indicates that introducing SSTs leads to a 1.3% increase in conversion rates. However, this result is not significant. I therefore study in greater depth seasonal effects. The model is not overfitted as it requires a total of 300 observation (20 observations multiplied by 15 controls).

The F-tests for all these specifications are summarized in the appendix: they suggest to include all seasonal variables with the exception of day-of-month. After I introduce them one by one, the coefficient varies significantly between specification and they are not statistically significant in the specifications (6) and (1.4).

Narrowing down the sample

As a result, I narrow down my sample to the period from the 12th of October to the 28th of December. This narrower sample is more likely to respect the parallel trends assumption and might show a statistically significant effect. My results are below:

Table 32: Conversion rates (narrow sample)

	OLS (1)	Month FE (2)	Store FE (3)	Month and store (4)
SST	0.05439***	0.04407	0.04707***	0.01573
	(0.01303)	(0.02562)	(0.00698)	(0.01294)
Adjusted R^2	0.0593	0.9628	0.9896	0.9908
Time FE	No	Yes	No	Yes
Store FE	No	No	Yes	Yes
N	564	564	564	564

The standard errors are in parentheses; * p < 0.10, ** p <0.05, *** p <0.01

I now control for seasonal effects as described in the beginning of the section *Empirical Analysis*.

Table 33: Conversion rates (narrow sample, seasonalities)

	Weekday (1.1)	Day of month (1.2)	Day of year (1.3)	Final (1.4)
SST	0.05530***	0.06106***	0.08477	0.01358
	(0.01309)	(0.01365)	(0.04744)	(0.01744)
Adjusted R^2	0.9622	0.96804	0.9558	0.9947
Time FE	Yes	Yes	Yes	Yes
Store FE	No	No	No	Yes
N	564	564	564	564

The standard errors are in parentheses; * p < 0.10, ** p <0.05, *** p <0.01

The F-tests for all these specifications are also in the appendix. This time, they suggest to include all variables.

The coefficient of the OLS model (1) is 0.05, while those of models (4) and (1.4) were both 0.01. The results do not vary significantly from those of the larger sample. We find strong, significant effects in specifications (1), (3), (4) and (5), but not (6). Once monthly, day-of-month and day-of-year effects are added all significance is lost, but the F-tests recommend not to include them in the final model. The best model is (1.4). This is also due to the smaller number of controls, which is necessary given the sample size (9 x 20 = 180 observations are required).

4.4 Productivity

4.4.1 Descriptive patterns (Productivity measure (a), Quantity / Employee hours)

Figure 18 shows the average labour productivity based on *Quantity* and *Employee hours,* productivity measure (a), over time for the treatment and control groups. The vertical line shows the first date in which the SSTs were introduced in the first store (the time of the first treatment on November 12, 2019):

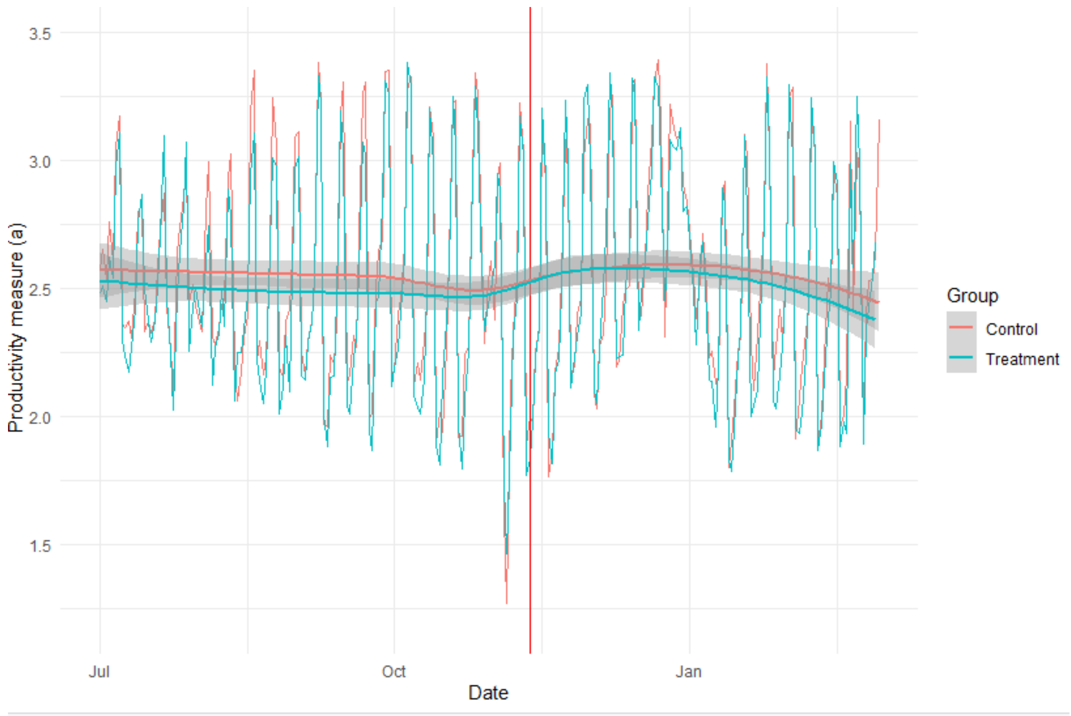


Figure 18: Productivity measure (a), distribution over time for treatment group and control group

The graph shows more evidence that the control group is slightly more productive than the treatment group. What is most striking though is the oscillating behaviour of *quantity / employee hours*: this is produced by different levels of productivity on different days of the week, where the peaks represent weekends. I find the following:

Table 34: Productivity over the days of the week

Weekday	All stores	Treatment Group	Control Group
Monday	2.19	2.15	2.20
Tuesday	2.16	2.12	2.17
Wednesday	2.27	2.21	2.29
Thursday	2.44	2.39	2.45
Friday	2.57	2.57	2.56
Saturday	3.10	3.10	3.10
Sunday	3.06	3.02	3.10

During the week, levels of productivity seem to rise from Mondays / Tuesdays to then peak on the weekend: this behaviour is consistent with the previous graph. In the weekends and especially on Saturdays, stores tend to be much more productive, with a level of productivity even 50% higher than on Mondays and Tuesdays. These results hold for all stores, for the treatment group and the control group. The same table (with similar results) is available for *Productivity measure (b)* in the appendix.

I will now compute the averages before and after the introduction of SSTs for each store for the whole period.

Table 35: *Productivity (a)*, treatment and control group

Average Productivity (a)		
	Treatment group	
	Pre-treatment	Post-treatment
Store 1	2.41	2.48
Store 2	2.51	2.53
Store 4	2.60	2.65
Store 5	2.41	2.54
General	2.482	2.548

Average Productivity (a)		
	Control group	
	Pre-treatment	Post-treatment
Store 7	2.52	2.56
Store 11	2.53	2.57
Store 12	2.61	2.65
Store 13	2.47	2.55
General	2.534	2.582

I find that in every store *Productivity (a)* increased after the introduction. The increase is similar in the treatment group (0.6) and the control group (0.5) and is small in both cases. It is thus hard to see an effect, as is the case with the graph above. The absolute values of the stores are very similar, but the value for Store 5 increases more than the others.

When I examine the trendlines before the introduction of SSTs, I see that the two groups behave very similarly until early October. However, after that moment the two groups converge and diverge in the end of February just as the *conversion rate* did.

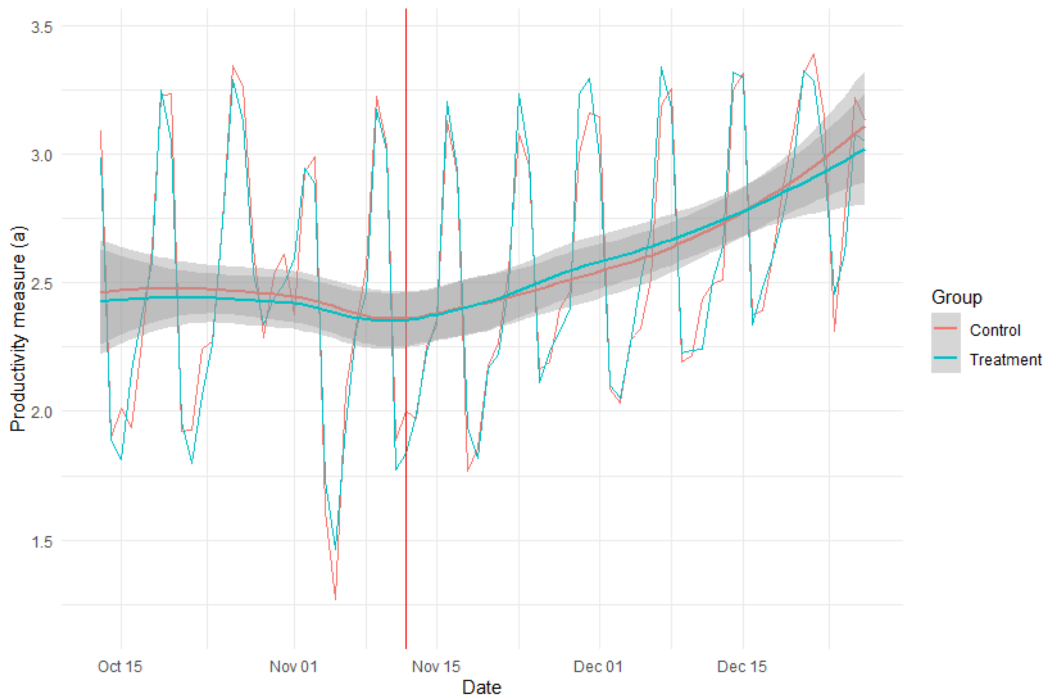


Figure 19: Productivity measure (a), distribution over time for treatment group and control group, narrow sample

When I narrow the sample to the period from 12th of October to the 28th of December in figure 19, the identifying assumption c) (parallel trends) seems more plausible. The trendlines are parallel until just a couple of days before the first introduction. After the introduction, the two groups converge until the end of December.

4.4.2 Regression analysis

I now estimate 4 initial regression specifications using *Productivity measure (a)* (*Quantity / Employee hours*). My outcome variable is the log of *Productivity measure (a)* and *SST* is an indicator for a transaction occurring in a store which had already introduced self-scanners. Specifications (1) to (3) are as described in the beginning of the section *Empirical Analysis.* I initially include *weekday* effects instead of *monthly* effects due to the different levels of productivity exhibited on different days of the week (see section 2.7, *Study of the Database).* In specification (4), I include all controls from models (1) to (3) (time and store fixed effects).

I can assess whether introducing monthly fixed effects and store fixed effects improves my models by employing F-tests, which have as a null hypothesis H0 that the observed and unobserved effects are equal to zero. If the null hypothesis is rejected, the monthly fixed effects or the store fixed effects model is an improvement over the OLS. The p-value of both F-tests is approximately 0 (it is 0.000006 for store fixed effects), which indicates that models (2) and (3) are an improvement over model (1).

Table 36: Productivity (a) (full sample)

	OLS (1)	Weekday FE (2)	Store FE (3)	Weekday and Store (4)
SST	0.06571*	0.05860**	0.06895*	0.06183***
	(0.03083)	(0.01806)	(0.03053)	(0.01750)
Adjusted R^2	0.0036	0.9836	0.9675	0.9893
Time FE	No	Yes	No	Yes
Store FE	No	No	Yes	Yes
N	1,923	1,923	1,923	1,923

The standard errors are in parentheses; * p < 0.10, ** p <0.05, *** p <0.01

I now control for seasonal effects as described in the beginning of the section *Empirical Analysis*.

Table 37: Productivity (a) (full sample, seasonalities)

	Month (1.1)	Day of month (1.2)	Day of year (1.3)	Final* (1.4)
SST	0.28422**	0.06254*	-0.20937**	-0.1057143
	(0.10943)	(0.03040)	(0.07594)	(0.0695239)
Adjusted R^2	0.9683	0.9679	0.9949	0.9958
Time FE	Yes	Yes	Yes	Yes
Store FE	No	No	No	Yes
N	1,923	1,923	1,923	1,923

The standard errors are in parentheses; * p < 0.10, ** p <0.05, *** p <0.01

*The results for specification (1.4) do not vary if we choose to include or not Monthly effects

The F-tests for all these specifications are in the appendix: they suggest to include all seasonal variables in the final specification. The coefficient of the OLS model (1) is 0.06, while those of models (4) and (1.4) were respectively 0.06 and -0.10. When they are introduced in specification (1.4), the coefficients of interest change significantly from (4) and all statistical significance is lost. After they are transformed from the logarithmic form, the coefficients found are 6.37% for model (4) and −11.07% for model (1.4)[4]. When we compare specifications (1.1), (1.2), (1,3) we also see that they significantly change from the coefficient of the baseline OLS model (1) (0.06).

[4] As the dependent variable is logged, it is necessary to transform the coefficient. The effect on Quantity / Employee Hours is: $(\exp(\beta) - 1) * 100$ where β is the coefficient.

Narrowing down the sample

As a result, I narrow down my sample to the period from the 12th of October to the 28th of December. This narrower sample is more likely to respect the parallel trends assumption and might show a statistically significant effect. My results are below:

Table 38: Productivity (a) (narrow sample)

	OLS (1)	Weekday FE (2)	Store FE (3)	Weekday and store (4)	Monthly (1.1)	Day of month (1.2)	Day of year (1.3)	Final* (1.4)
SST	0.28530*** (0.05895)	0.27701*** (0.03114)	0. 30241*** (0.05853)	0.29402*** (0.02951)	0.28422* (0.12096)	0.28500*** (0.05584)	-0.20937** (0.07748)	-0.062864 (0.072182)
Adjusted R^2	0.07064	0.9898	0. 964	0.9909	0.963	0.9682	0.9949	0.996
Time FE	No	Yes	No	Yes	Yes	Yes	Yes	Yes
Store FE	No	No	Yes	Yes	No	No	No	Yes
N	564	564	564	564	564	564	564	564

The standard errors are in parentheses; * p < 0.10, ** p <0.05, *** p <0.01

*The results for specification (1.4) do not vary if we choose to include or not Monthly effects

The F-tests for all these specifications are in the appendix: they suggest to include all variables with the exception of monthly effects.

The coefficient of the OLS model (1) is 0.28, while those of models (4) and (1.4) were respectively 0.29 and -0.06. We find strong, significant effects in specifications (1) to (4) which are larger than those of the full sample. Once monthly, day-of-month and day-of-year effects are added though, all significance is lost. The coefficient becomes significant only if *month* and *day-of-the-year* are removed, while removing only one of the two yields similar but insignificant coefficients of -0.08 (without *day-of-the-year*), and -0.062 (without *month*).

Summary of results

While some of my models show a significant effect of SSTs on productivity of approximately 6% (see models (1)-(4)), these coefficients are not very robust and all significance is lost once d*ay-of-the-year* is added to the final model (1.4). As it was shown that both my productivity measures oscillate over time, controlling for seasonal effects is rather important. On top of this, while the final model may not be appropriate due to overfitting; the coefficients found for models (1)-(4) are also not reliable as they are rather high and do not match the visual inspection which suggests the presence of a small effect or no effect at all. Basker (2012) found that introducing scanners increased productivity of US stores by 4.5%: it would be curious if merely making these scanners automatic had a 6% increase instead. It is also possible that not controlling for capital inputs distorted the measures of productivity as the denominator (inputs) only describes labour inputs: a small denominator could thus result in an abnormally high effect. Unfortunately, data on capital investments is not available and cannot be used to create a Total Factor Productivity measure. Finally, it is not clear whether the parallel trends assumption holds. Narrowing the sample makes this assumption somewhat more believable, but the resulting regressions produce extreme coefficients of approximately 28%. In conclusion, there does not seem to be a clear effect of SSTs on productivity.

4.4.3 Descriptive patterns (Productivity measure (b), Net sales / Employee hours)

Figure 20 shows the average labour productivity based on *Net sales* and *Employee hours,* productivity measure (b), over time for the treatment and control groups. The vertical line shows the first date in which the SSTs were introduced in the first store (the time of the first treatment on November 12, 2019):

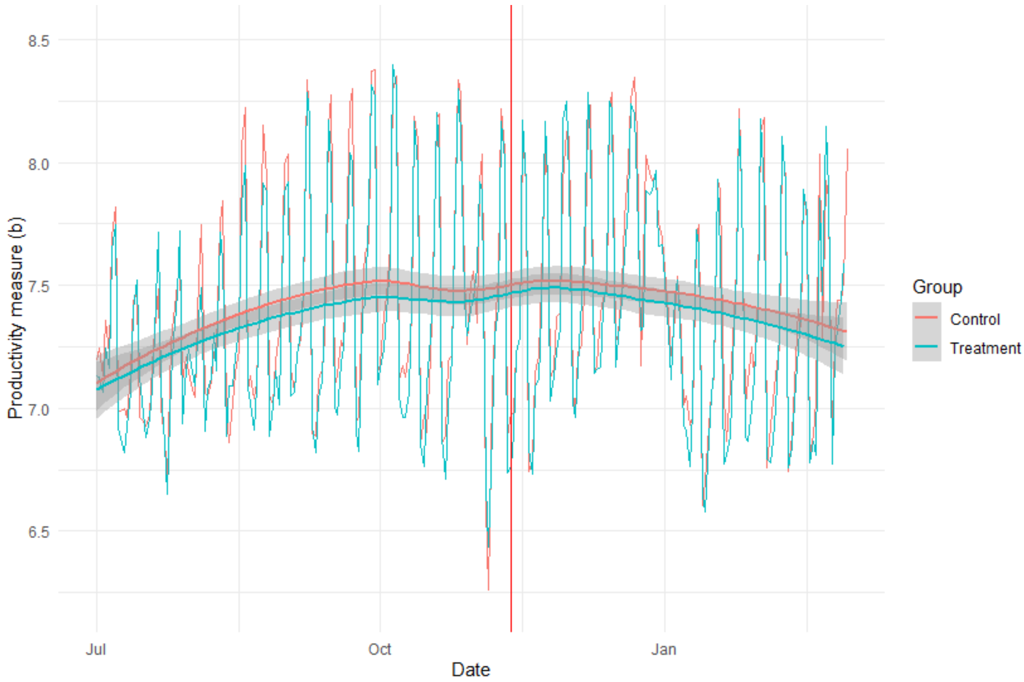


Figure 20: Productivity measure (b), distribution over time for treatment group and control group

The graph shows once again that the control group is slightly more productive and the considerable variation of productivity over different days of the week. When I examine the trendlines before the introduction of SSTs, I see that the two groups behave very similarly: the identifying assumption c) (parallel trends) seems to hold. However, the trendlines seem to remain parallel after the treatment, and it is thus also hard to visualize a treatment effect.

I will now compute the averages before and after the introduction of SSTs for each store for the whole period.

Table 39: *Productivity (b),* treatment and control group

Average Productivity (b)		
	Treatment group	
	Pre-treatment	Post-treatment
Store 1	7.27	7.37
Store 2	7.43	7.46
Store 4	7.45	7.50
Store 5	7.22	7.33
General	7.346	7.413

Average Productivity (b)		
	Control group	
	Pre-treatment	Post-treatment
Store 7	7.43	7.48
Store 11	7.41	7.46
Store 12	7.43	7.51
Store 13	7.31	7.41
General	7.396	7.465

When I split the statistics by store, I find that in every store *Productivity (b)* increased after the introduction. These increases are very similar in the treatment group (0.7) and the control group (0.7). It is thus hard to see an effect, as is the case with the graph above. The absolute values of the stores are also very similar.

When I restrict the sample to the period from the 12th of October to the 28th of December in figure 21, I notice that the gap between the treatment and control group slightly narrows after the treatment until mid-December, from when it widens more significantly. The previous graph though shows that in the larger sample this distance remains more or less constant. It is possible that the effect of SSTs on productivity is small or insignificant.

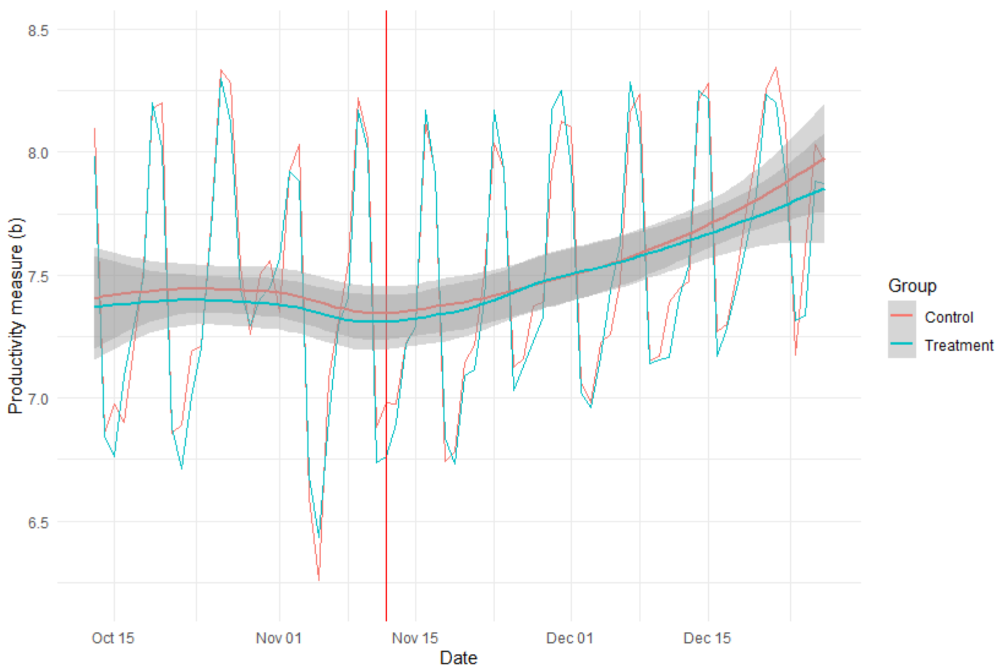


Figure 21: Productivity measure (b), distribution over time for treatment group and control group

4.4.4 Regression analysis (Productivity measure (b))

I now estimate 4 initial regression specifications using *Productivity measure (b)* (*Net sales / Employee hours*). My outcome variable is the log of *Productivity measure (b)* and *SST* is an indicator for a transaction occurring in a store which had already introduced self-scanners. Specifications (1) to (3) are as described in the beginning of the section *Empirical Analysis.* I initially include *weekday* effects instead of *monthly* effects due to the different levels of productivity exhibited on different days of the week (see section 2.7, *Study of the Database).* In model (4), I add *Price* as a control to my regression as indicated by the literature on productivity (eg Syverson, 2011). The reason why this is the standard practice with revenue measures of outcome is to ensure that price variations due to unobserved causes (such as market power) do not affect *Net sales* and therefore my outcome variable. In specification (5), I include all controls from models (1) to (3) (time and store fixed effects).

I can assess whether introducing weekday fixed effects and store fixed effects improves my models by employing F-tests, which have as a null hypothesis H0 that the observed and unobserved effects are equal to zero. If the null hypothesis is rejected, the weekday fixed effects or the store fixed effects model is an improvement over the OLS. The p-value of both F-tests is approximately 0 (it is 0.000006 for store fixed effects), which indicates that models (2) and (3) are an improvement over model (1).

Table 40: Productivity (b)

	OLS (1)	Weekday FE (2)	Store FE (3)	Price OLS (4)	Final (5)
SST	0.06736*	0.06006**	0.07043*	0.074305*	0.06836***
	(0.03237)	(0.01952)	(0.0319)	(0.030678)	(0.01745)
Price	No	No	No	0.010573***	0.00745***
				(0.001011)	(0.00063)
Adjusted R^2	0.003473	0.6375	0.033	0.1051	0.9987
Time FE	No	Yes	No	No	Yes
Store FE	No	No	Yes	No	Yes
N	1,923	1,923	1,923	1,923	1,923

The standard errors are in parentheses; * p < 0.10, ** p <0.05, *** p <0.01

I now control for seasonal effects as described in the beginning of the section *Empirical Analysis*.

Table 41: Productivity (b) (seasonalities)

	Month (1.1)	Day-of-month (1.2)	Day-of-year (1.3)	All time effects (1.4)
SST	0.27579*	0.06398*	-0.17939*	-0.094365
	(0.11258)	(0.03198)	(0.07996)	(0.069152)
Price	No	No	No	0.002859
				(0.00147)
Adjusted R^2	0.996	0.9958	0.9993	0.9995
Store FE	No	No	No	Yes
N	1,923	1,923	1,923	1,923

The standard errors are in parentheses; * p < 0.10, ** p <0.05, *** p <0.01

The coefficient of the OLS model (1) is 0.06, while those of models (5) and (1.4) were respectively 0.06 and -0.09. The final specification indicates that introducing SSTs leads to a 6.8% average percentage change in labour productivity. This result is highly significant. The coefficients above do not vary considerably between specifications: once they are transformed from the logarithmic form, they generally remain between 6.9% and 8%[5].

The F-tests for all these specifications are summarized in the appendix: they suggest to include all seasonal controls in the final specification. When I do this in specification (1.4) my coefficient of interest varies from specification (4) and is insignificant. Moreover, (1.1) and (1,3) yield very different coefficients from the rest of the models: their behaviour is similar to the coefficients of productivity (a). Additional regressions (not shown) indicate that significance is regained only if *day-of-the-year* is removed, but the resulting coefficient of 16% (p-value of 0.005) seems extremely high given the visual inspection in section 4.4.3 which suggests a small effect or no effect. As for *Productivity measure(a),* including *day-of-the-year* in the final model risks causing overfitting (289 controls x 20 = 5780) and losing an excessive number of degrees of freedom. Unlike productivity measure (b), identifying assumption c) (parallel trends) seems to hold.

Summary of results: Productivity measures (a) and (b)

In conclusion, I do not find a clear effect for both productivity measures. While models (1)-(4) have consistent estimates for both measures, these estimates are not robust and considerably high. Moreover, for one measure the parallel trends assumption might not hold. As a result, I conclude that there is no clear effect of the introduction of SSTs on productivity.

[5] As the dependent variable is logged, it is necessary to transform the coefficient. The effect on Quantity / Employee Hours is: (exp(β) – 1) * 100 where β is the coefficient.

5 Summary of results

In my thesis, I have analysed the effect of the introduction of SSTs on 8 dependent variables: the *Speed* to checkout, *Speed to Scan an Item,* the *Service Gap*, the *Number of Queues per day,* the *Proportion of Queues per day*, the *Conversion Rate, Quantity / Employee Hours* (*Productivity measure (a)*)*, Net Sales / Employee Hours* (*Productivity measure (b)*).

Speed

I find that stores which introduced SSTs increased their time to checkout (*Speed)* by 8.6 seconds on average, which represents a 16.9% increase. This result is highly significant and robust. The actual difference between the time employed to checkout for manual and automatic checkouts is much higher: mean comparison suggested that transactions from automated checkouts take approximately twice the time to complete than their manual counterparts (100 seconds vs 50 seconds). The size of the effect found is explained by the relatively small proportion of transactions that are carried out through automated checkouts.

Speed per item

I find that stores which introduced SSTs increased their average time to scan an Item (*Speed per item)* by 5 seconds on average, which represents a 20% increase. This result is highly significant and robust. This result is in line with the result of *Speed* and definitively shows that SSTs negatively affected the speed to check out, both measuring it as the time to complete a transaction and the time to scan an item.

Queuing

All queuing measures show a fall in the presence of queues after SSTs were introduced. The service gap increased by 35 seconds on average, while depending on the definition of queue in each store the *Number of Daily Queues* falls by either 45 or 100 on average and t*he Proportion of Daily queues* falls by either 3.4% or 10%. All coefficients are significant for both measures, but the estimates of *number of daily queues* are not very stable. However, all effects found point in the same direction (a reduction of queues) and are consistent with the visual inspection: while there may be uncertainty on the size of the effect which derives from how a queue is defined, there is no uncertainty on the presence of an effect. Therefore, I consider these results to be sufficiently robust. I will now take a conservative approach and select the smallest set of coefficients as my final results: I thus conclude that SSTs have a statistically significant effect in increasing the service gap by 35 seconds on average, reducing the *Number of Daily Queues* by 45 seconds and reducing the *Proportion of Daily Queues* by 3.4%. The first estimate is slightly less robust than the second, but both are sufficiently robust.

Conversion rates

The estimates of *conversion rates* show no significant effect. The value of the estimates vary wildly through specification and the final estimates are not statistically significant. The reason for this seems to be the violation of the parallel trends assumption which invalidates the difference-in-difference model used. Changing the sample size and examining a smaller time period over which the parallel trends assumption is more likely to hold does not affect the results. I thus conclude that no statistically significant effect of SSTs on conversion rates could be found.

Productivity

Both productivity measures are more or less inconclusive (at the moment). Most coefficients for both productivity measures defined using quantity (a) or sales (b) show an increase in productivity by 6% on average. These coefficients however are not robust as they oscillate between models and the significance is lost depending on which controls are included. Moreover, visual inspection shows that, while the parallel trends assumption holds, effects due to the introduction of SSTs are small. This is consistent with the findings on queuing and *speed:* the positive (decreased queuing times) and negative (*speed* of transaction) effects of the introduction at least in part cancel each other out. As a result, I conclude that there is not enough evidence to conclude that SSTs have a statistically significant effect on productivity.

6. Conclusion and Discussion

The degree of success of new technologies is very hard to predict. Sometimes innovations like Information Technology are disruptive and lead to great productivity improvements. In some other cases though, they are less successful and are soon forgotten: while Facebook.com and Microsoft's laptops are valued and appreciated, few remember Facebook Phone or Microsoft Zoon. Making new products or technologies that are useful for businesses and consumers is extremely difficult. Nonetheless, technological development is one of the key drivers of productivity and growth, and even unsuccessful innovations serve a purpose by becoming a stepping stone for better tools. Very few inventions have the drastic effect that the wheel or the printing press had, but incremental progress can turn primitive computers slower than humans in basic algebra into the powerful machines that they are today.

Self Service Technologies falls somewhere in between the two extremes. In this study, I found that introducing SSTs in retail stores increased the time to checkout, decreased the time to queue and had no effect on conversion rates or productivity. Since the effects on *Speed* and queuing measures are opposite in sign and it is not possible to quantify how many seconds of queuing time were saved, the outcome of the technology on idle time is uncertain. Based on my results, it thus seems that the introduction of SSTs did not have a significant effect on store performance in the short run. Nonetheless, the machines are comparable to manual cashiers in most respects and are capable of performing the same operations without losses in productivity. Moreover, introducing automated checkouts may prove to be useful in certain situations. For example, if a business were in great need of reducing queuing times, it might be worth introducing the checkouts even if the time to checkout would increase.

Nonetheless, more research is needed to fully understand how this technology interacts with business operations. Learning by doing for example has been shown to be a key determinant of productivity (Benkard 2000), and greater experience over time with the checkouts both by customers and the store staff may make SSTs more productive and the stores more profitable in the future. It would also be useful to understand if learning by doing could affect *Speed* as customers become more acquainted with the technology. In addition, it would be interesting to analyse the reallocation of employees in detail. These results would have both economic importance by adding to the relevant literatures on productivity and service quality and be of business relevance to companies attempting to introduce automated checkouts.

Another issue that needs further exploration is the role of costs and capital inputs. My analysis was predominantly made from an "output perspective": with the exception of productivity measures (which took into account the number of employee hours used in the stores), the costs of implementing the technology were not examined. In this study I focused on labour productivity instead of Total Factor Productivity (TFP) measures. However, introducing the capital inputs required for the machines in the analysis by recovering a TFP measure is an interesting area for future research.

Appendix

Literature Review

Basker & Klimek (2017)

Figure 22 shows the behaviour of two measures of productivity in a subset of their sample. *Employee productivity* includes both the productivity of customers and workers, while *worker productivity* only considers that of employees. The gap between these two measures widens between 1977 and 1992 by 12.5% as self service pumps are introduced, showing that employee productivity rises faster than worker productivity. The true productivity of the stations lies in between the two lines.

Figure 22: Log Physical Productivity, of Employee and Worker productivity[6]

[6] Source: Basker & Klimek (2017), pp. 66, figure 5

Summary Statistics

Dependent variables (unfiltered)

	Speed	Speed per item	Service gap	Number of Queues	Proportion of Queues	Conversion rate	Productivity (a)	Productivity (b)
Mean	52	27.34	177	482.7	0.4248	Inf	Inf	Inf
Median	36	20	37	401	0.4272	0.61868	2.463	7.329
Min	0	0	-557	53	0.1667	-0.03846	1.11	6.148
Max	80762	80762	1542807	2199	0.6633	Inf	Inf	Inf
Range	80762	80762	1543364	2146	0.4966	NA	NA	NA
Standard deviation	171.60	136.99	1978.70	320.203	0.0850451	NA	NA	NA
Skewness	256.55	335.51	492.979	1.76138	-0.185262	NA	NA	NA
Kurtosis	86559	144225	319652	3.967491	-0.270952	NA	NA	NA
5th percentile	11	7.666	8	142	0.2800978	0.2473791	1.866931	6.752524
25th percentile	22	13	17	263.75	0.3683241	0.490144	2.196022	7.057274
75th percentile	61	31	96	595	0.4861268	0.7594047	2.890143	7.756516
95th percentile	141	65.5	589	1160	0.5600945	1.0772453	3.396061	8.335381
Missing values	0	208108	1	0	0	0	0	0
N	3,656,477	3,656,477	3,656,477	29,691	29,691	29,691	29,691	29,691

Averages before and after for the dependent variables

When I compare the average values of my dependent variables for the treatment and control group before the introduction of SSTs, I find that they are similar. These values are all calculated *after* they have been properly filtered as described in the next section (*filtering)*.

	Speed	Speed per item	Service Gap	Number of queues	Proportion of queues	Conversion rate	Productivity (a)	Productivity (b)
Store 1	52.5	24.4	132	757	0.51	0.64	2.41	7.27
Store 2	49.5	23.6	182	920	0.48	0.62	2.51	7.43
Store 4	48.4	23.7	147	390	0.49	0.80	2.6	7.45
Store 5	46.5	25.4	164	492	0.43	0.42	2.41	7.22
Store 7	51.1	25.8	169	368	0.44	0.63	2.52	7.43
Store 11	53.3	24.9	162	481	0.45	0.60	2.53	7.41
Store 12	51.6	25.8	152	255	0.40	0.61	2.61	7.43
Store 13	55.9	27.5	179	408	0.41	0.58	2.47	7.31

When I examine the average values of my dependent variables before SSTs were introduced, they seem rather similar (with perhaps the exception of *Number of queues*). This table thus shows further evidence that the stores I selected are comparable.

Study of the Dataset

Dependent variables:

Speed

Figure 23, Boxplot of *Speed* (all stores, unfiltered):

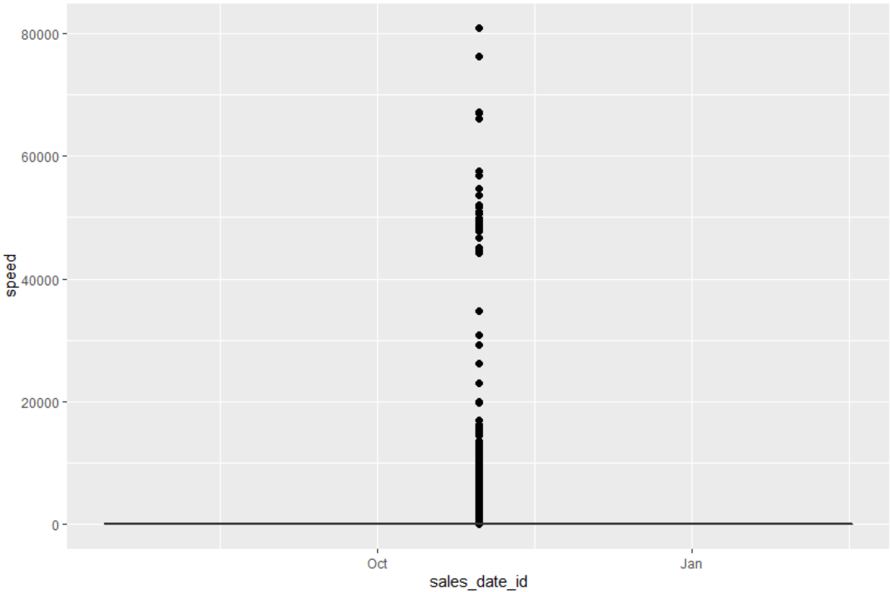
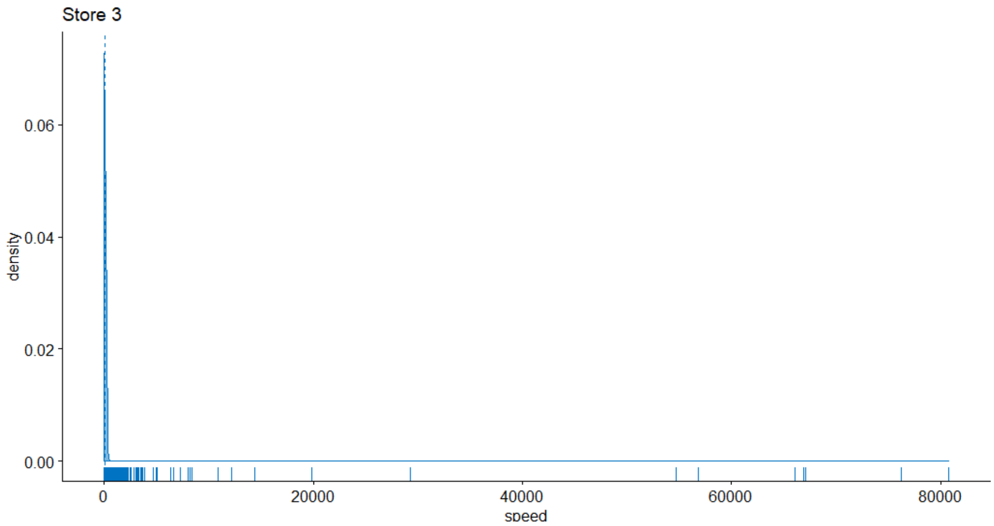
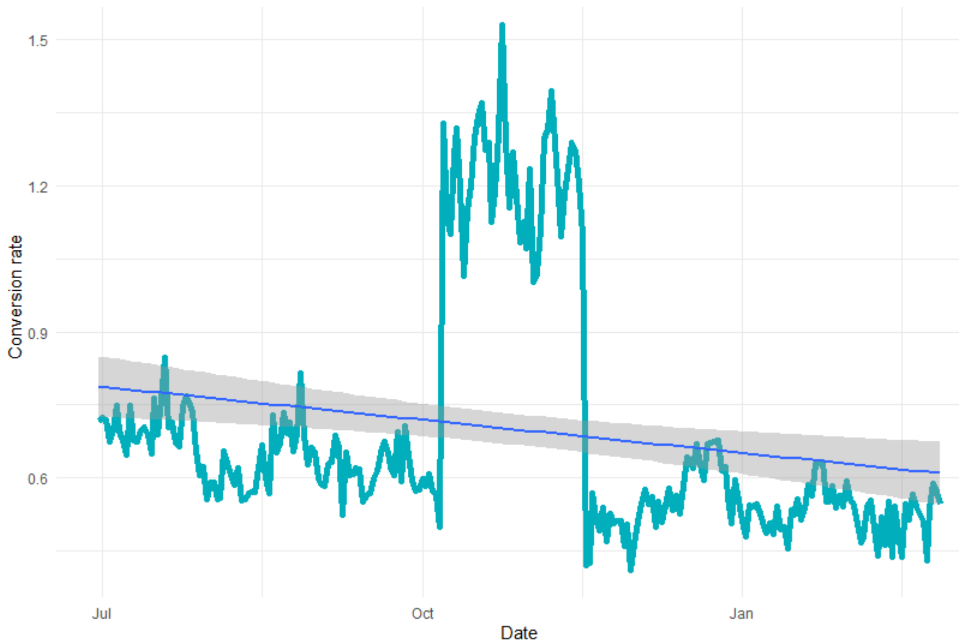


Figure 24: density of *Speed* (Store 3 only, unfiltered):



Conversion rate

Figure 25: conversion rate of store 4 over time (unfiltered sample):



Productivity measure (b)

Weekday	All stores	Treatment Group	Control Group

Monday	7.06	7.00	7.06
Tuesday	7.03	6.97	7.03
Wednesday	7.13	7.06	7.15
Thursday	7.31	7.25	7.31
Friday	7.43	7.43	7.43
Saturday	7.99	7.98	7.99
Sunday	7.95	7.91	7.99

Other variables:

The first set of variables shown are those which change at the time of introduction, while the remaining ones follow afterwards.

Number of Transactions (daily)

I examine the distribution over time of the *Number of Transactions (daily)*.

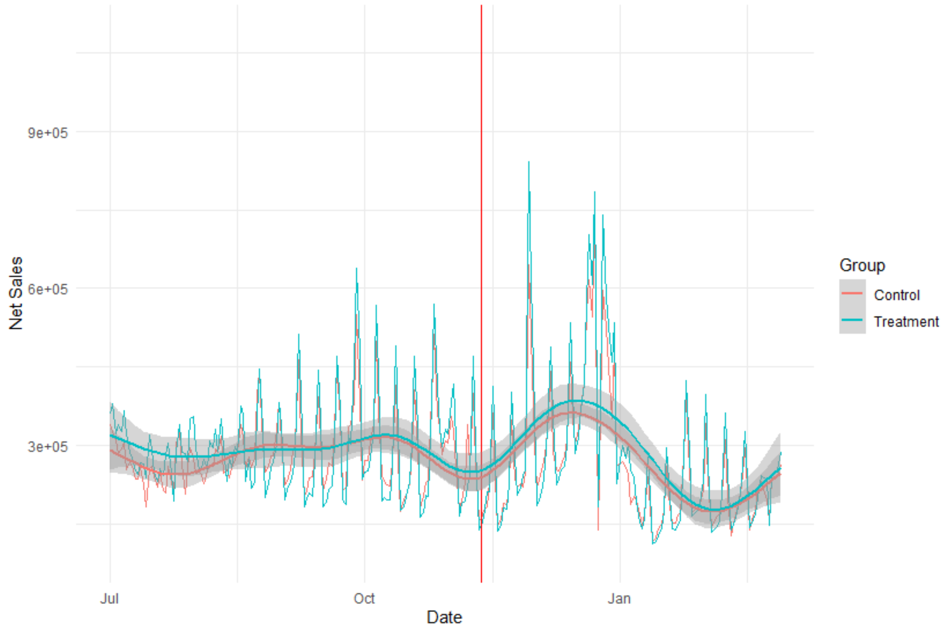
Day	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Percentage of clients	13.03%	12.14%	11.96%	13.86%	16.37%	18.62%	14.03%

Month	July	August	September	October	November	December	January	February
Percentage of clients	13.70%	12.12%	11.73%	12.45%	11.49%	17.80%	10.69%	10.03%

The tables above show the percentage of total clients every day of the week and every month. We can see that over the week, clients are slightly more concentrated on Friday, Saturday and Sunday. Moreover, there are more clients in December than in the other months of the year. This result suggests that there may be weekly and monthly seasonal effects which may influence the models: I shall thus control for them appropriately when I run the various regressions. I also expect the higher number of clients on weekends to explain the higher levels of productivity on those same days.

Net sales (daily)

When I examine the distribution over time of *Net sales* for stores in the treatment and control group in figure 26, I find that *Net Sales* were slightly higher in the treatment group at the time of treatment (the sample is unfiltered):



(some values were removed by R as they were non-finite)

Figure 26: Net sales, distribution over time for treatment group and control group

The density and distribution over time of the daily net sales (unfiltered sample) are shown below in figure 27:

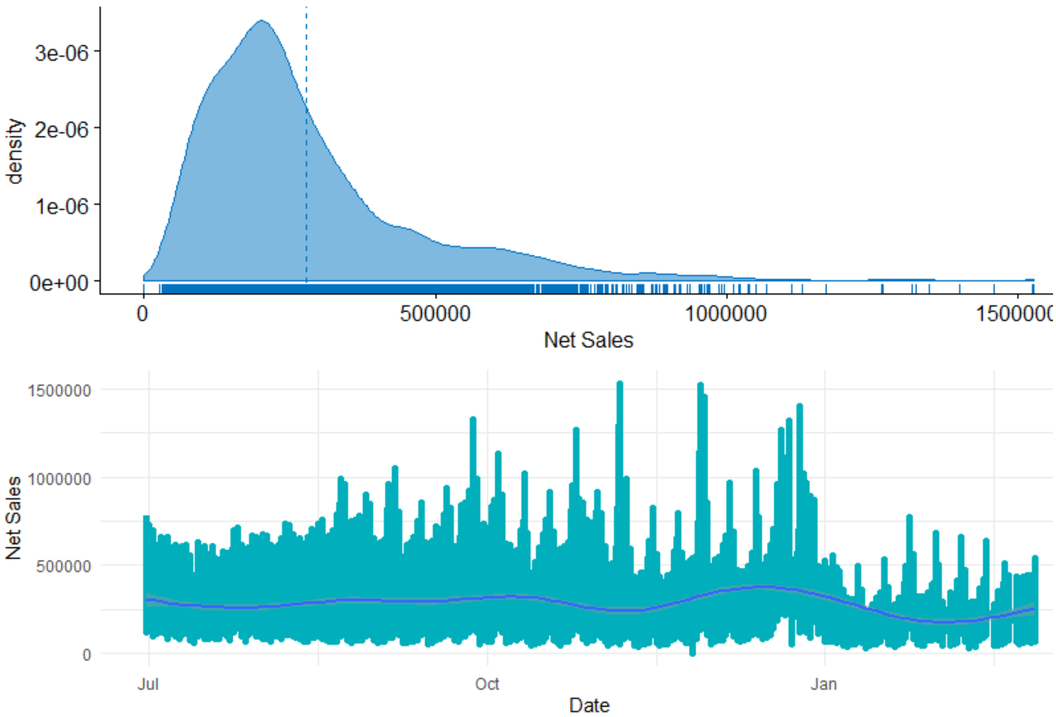
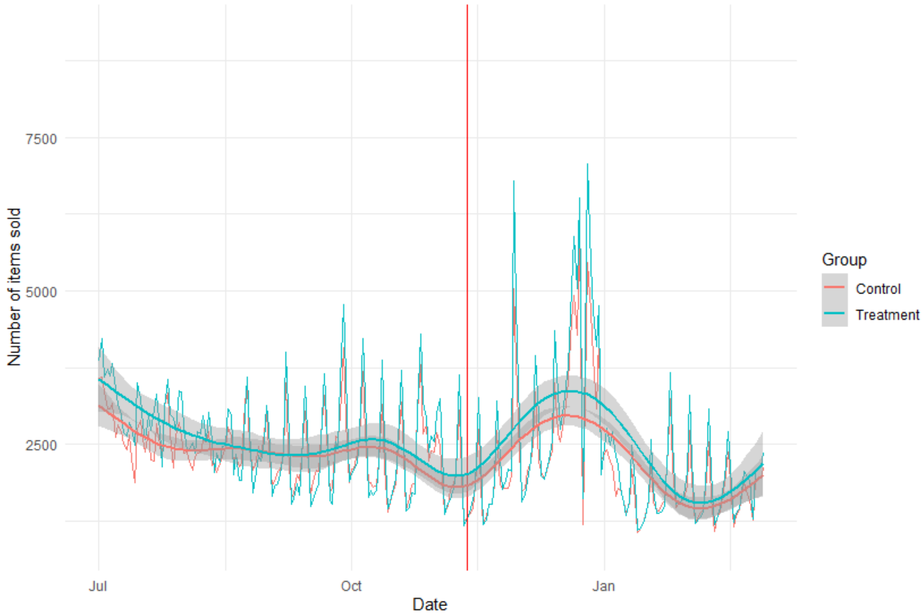


Figure 27: Net sales, density and distribution over time

Number of items sold (daily)

When I examine the distribution over time for stores in the treatment and control group in figure 28, I find that in the end of December the *Net Sales* increased faster in the former than in the latter, shortly after the time of treatment (the sample is unfiltered):



(some values were removed by R as they were non-finite)

Figure 28: Number of items sold, distribution over time for treatment group and control group

The density and distribution over time of the daily number of items sold (unfiltered sample) is shown below in figure 29:

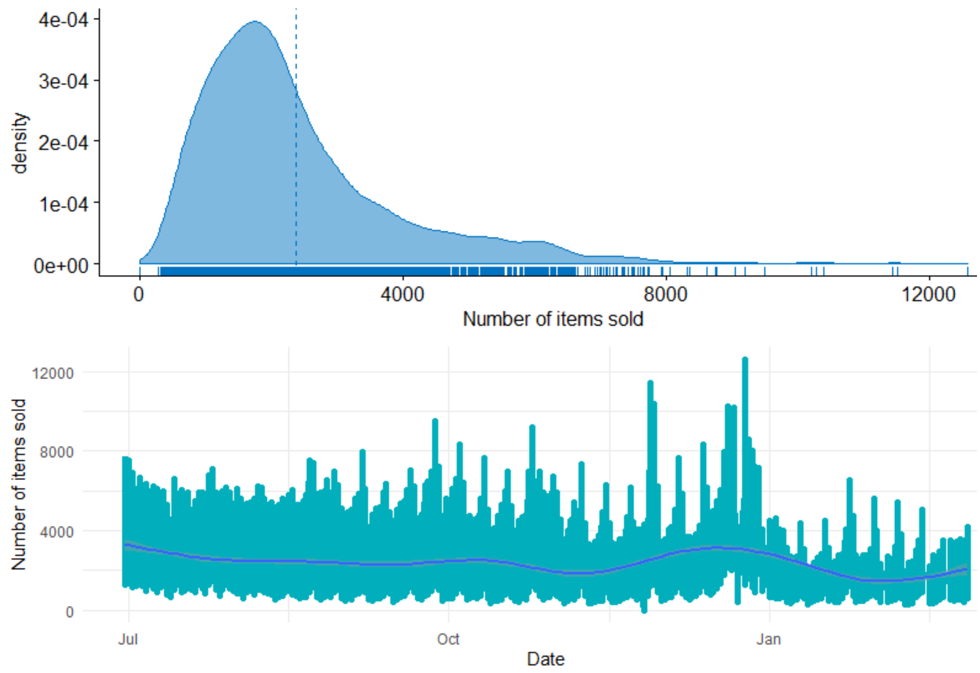
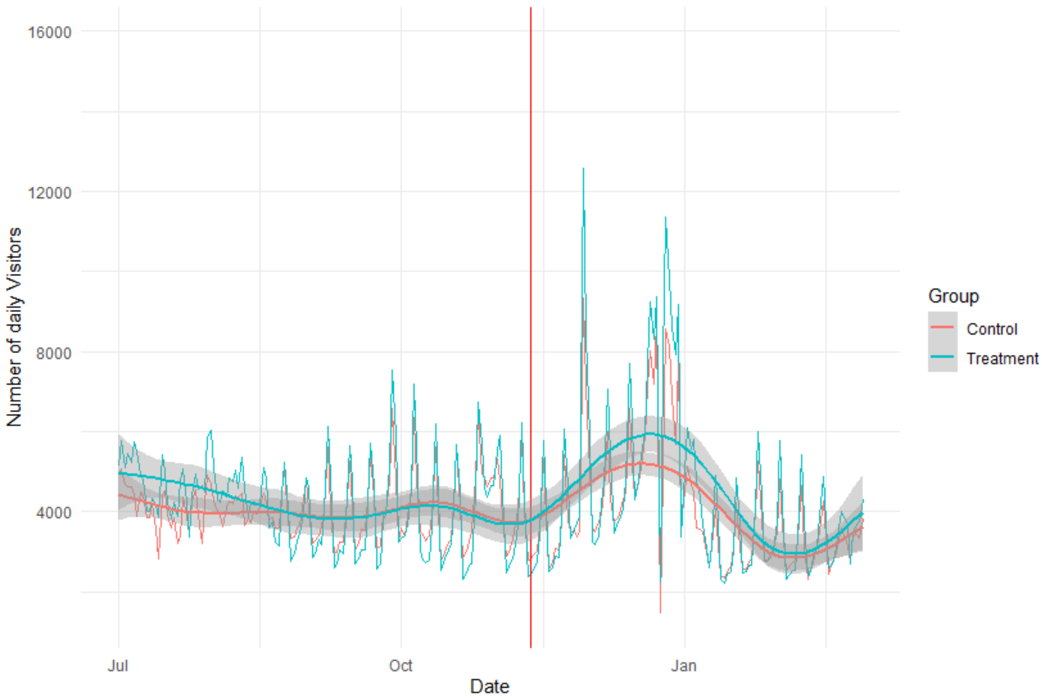


Figure 29: Number of items sold, density and distribution over time

Number of Visitors (daily)

When I examine the distribution over time for stores in the treatment and control group in figure 30, I find that in the end of December the *Number of daily Visitors* increased faster in the former than in the latter, shortly after the time of treatment (the sample is unfiltered):



(some values were removed by R as they were non-finite)

Figure 30: Number of daily visitors, distribution over time for treatment group and control group

The density and distribution over time of the daily number of customers (unfiltered sample) is shown below in figure 31:

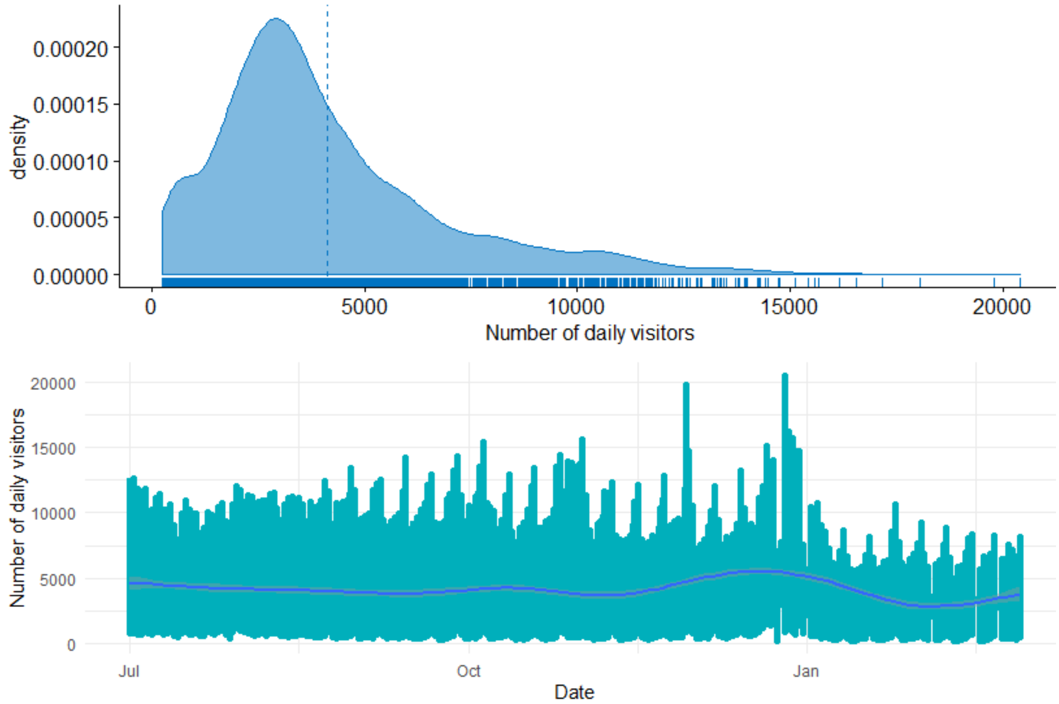


Figure 31: Number of daily visitors, distribution over time for treatment group and control group

I notice that the variables takes many values of 0, which would cause the *Conversion Rate* to be infinite. However, when I restrict the sample to the Treatment and the Control group all these values disappear making no filtering necessary.

Employee hours (daily)

The density and distribution over time of employee hours is below in figure 32 (unfiltered sample):

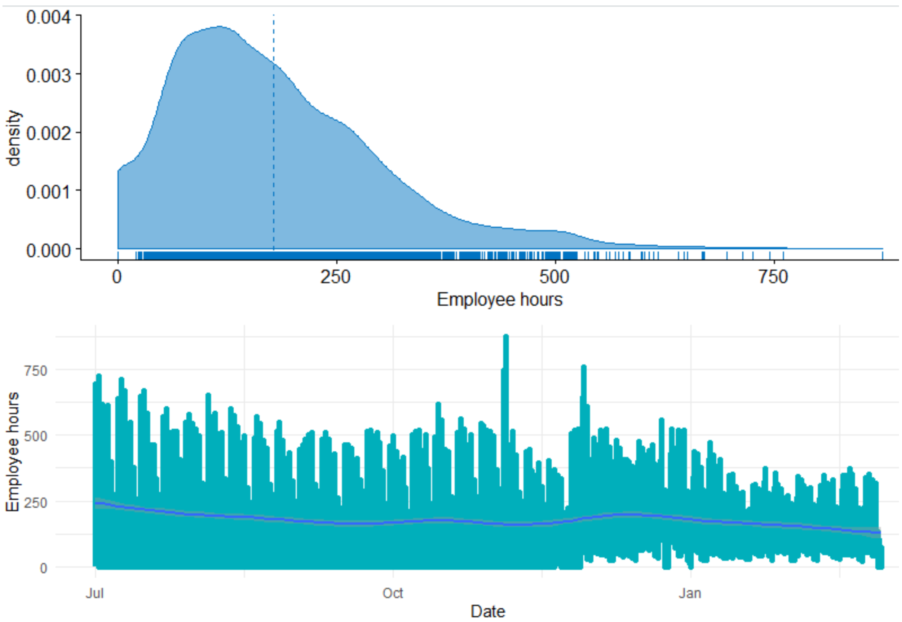


Figure 32: Employee hours, density and distribution over time, unfiltered

I can see that the average number of opening hours in the 14 stores remained more or less constant over time, falling slightly over the sample and becoming entirely greater than zero from December. The oscillating behaviour of the variable over the sample is due to weekday variations in the number of opening hours. From the density graph I notice that there are many cases in which the number of employee hours is equal to zero. Almost all these values come from Store 8. Once Store 8 (only) has been removed from the sample, we obtain the following (figure 33):

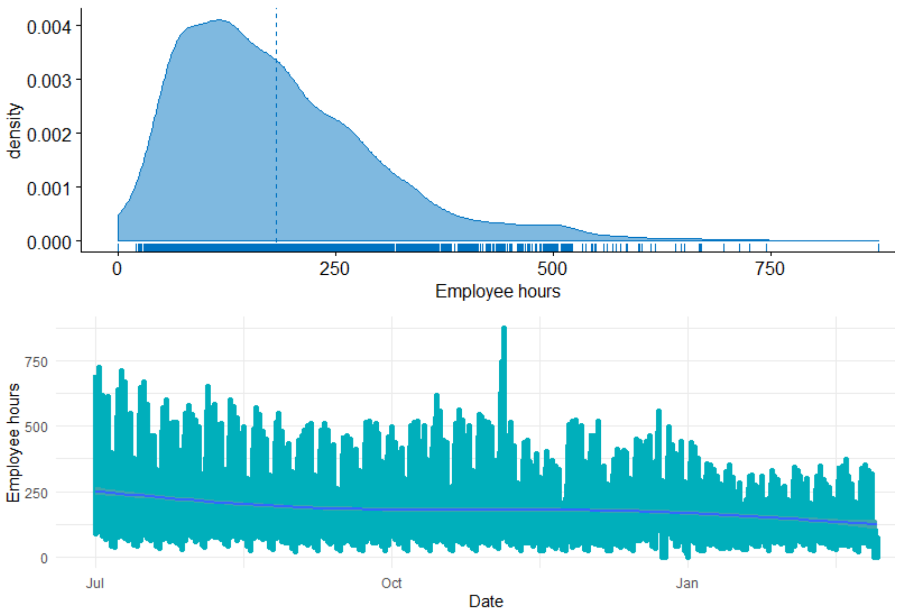


Figure 33: Employee hours, density and distribution over time, filtered

This decision validates my choice to exclude Store 8 from the Control group.

When I split the distribution over time for stores in the treatment and control group, I find that *Employee hours* changed after the time of treatment in figure 34 (the sample is unfiltered):

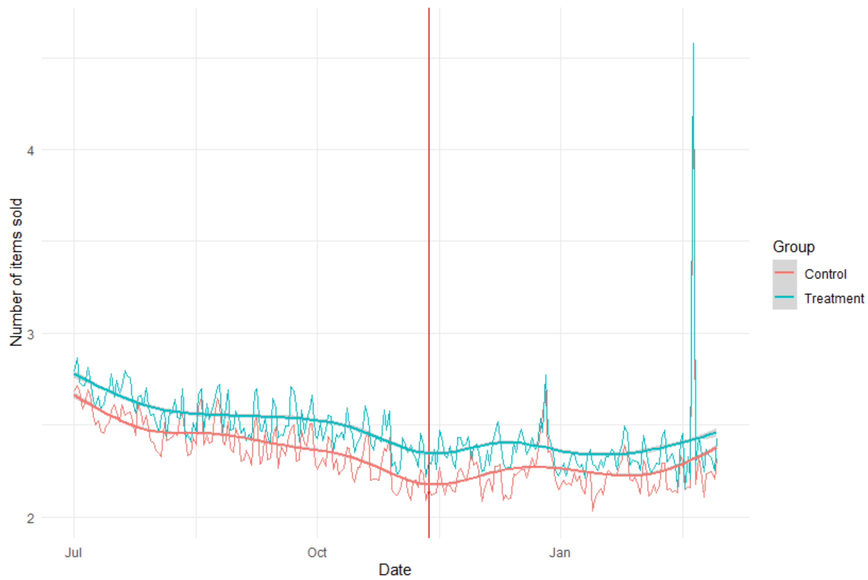
(some values were removed by R as they were non-finite)

Figure 34: Employee hours, distribution over time for treatment group and control group

The different behaviour of the two groups could be explained by the introduction of SSTs that lead to labour savings in the stores. I would expect this to positively productivity levels in treatment stores which, all else equal, would be using less *Employee hours.*

Number of items sold (per transaction)

This variable from the *receipts* database shows the number of items sold in every transaction. When I split the distribution over time for stores in the treatment and control group in figure 35, I find that no shocks occurred to *Number of items sold* at the time of treatment (the sample is unfiltered):

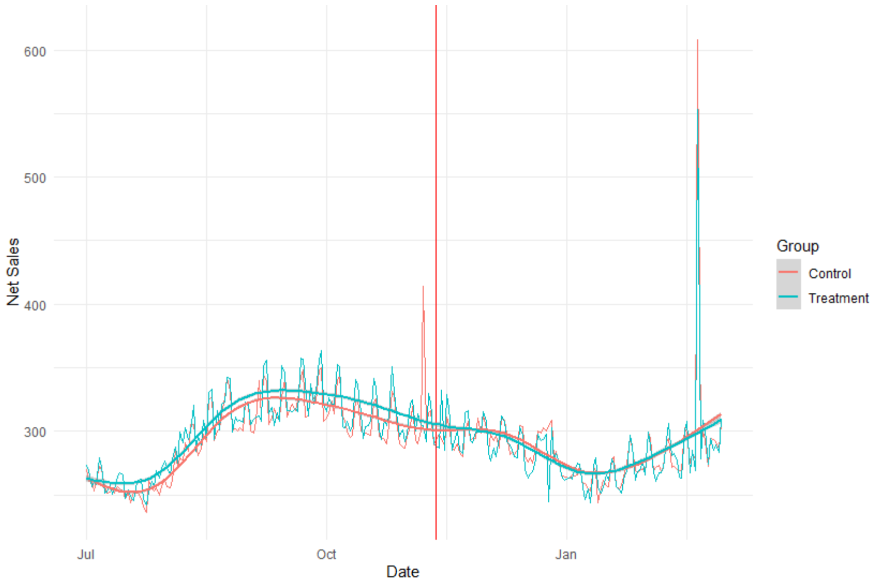


(some values were removed by R as they were non-finite)

Figure 35: Number of items sold, distribution over time for treatment group and control group

Net Sales (per transaction)

When I split the distribution over time for stores in the treatment and control group in figure 36, I find that no shocks occurred to *Net Sales per transaction* (the revenue from every transaction) at the time of treatment (the sample is unfiltered):



(some values were removed by R as they were non-finite)

Figure 36: Net sales, distribution over time for treatment group and control group

Number of Transactions (daily)

When I split the distribution over time for stores in the treatment and control group in figure 37, I find that no shocks occurred to *Number of Transactions* (the revenue from every transaction) at the time of treatment (the sample is unfiltered):

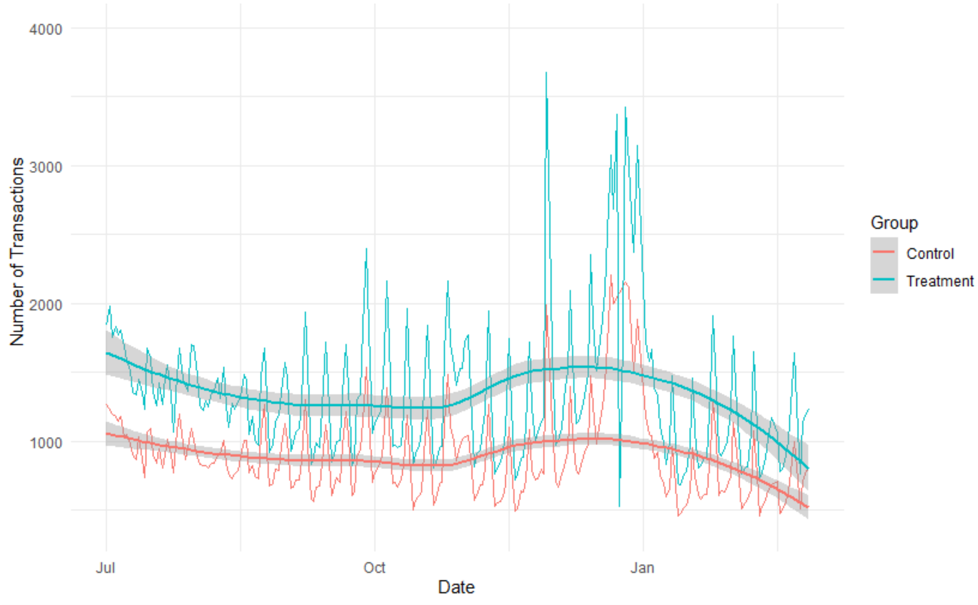


Figure 37: Number of transactions, distribution over time for treatment group and control group

Average Price (daily)

The density and distribution over time of this variable is displayed below in figure 38. The graphs on the left are unfiltered, while in those on the right I removed outliers where *price* = 0:

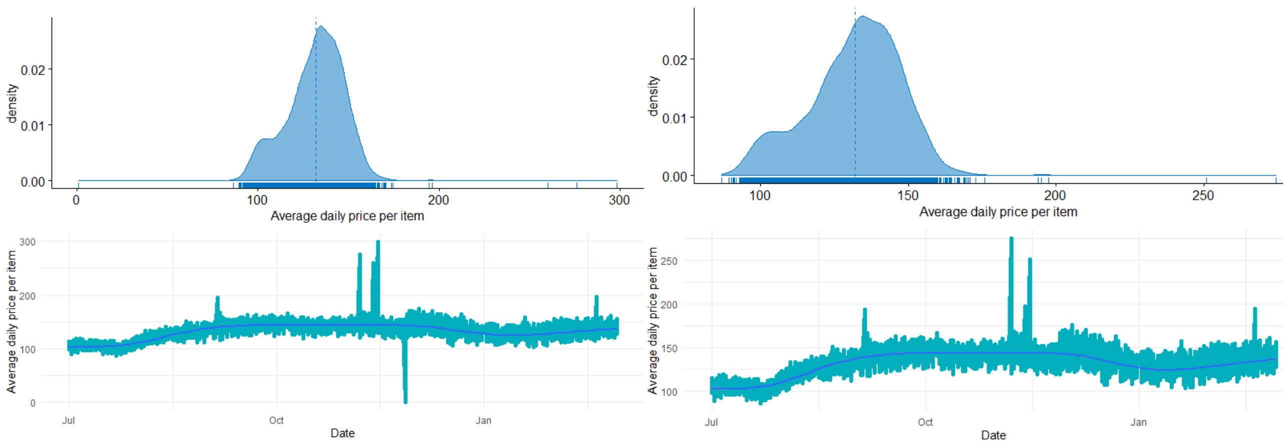


Figure 38: Average price, density and distribution over time, filtered and unfiltered

The distribution over time in the treatment and control groups is displayed below in figure 39:

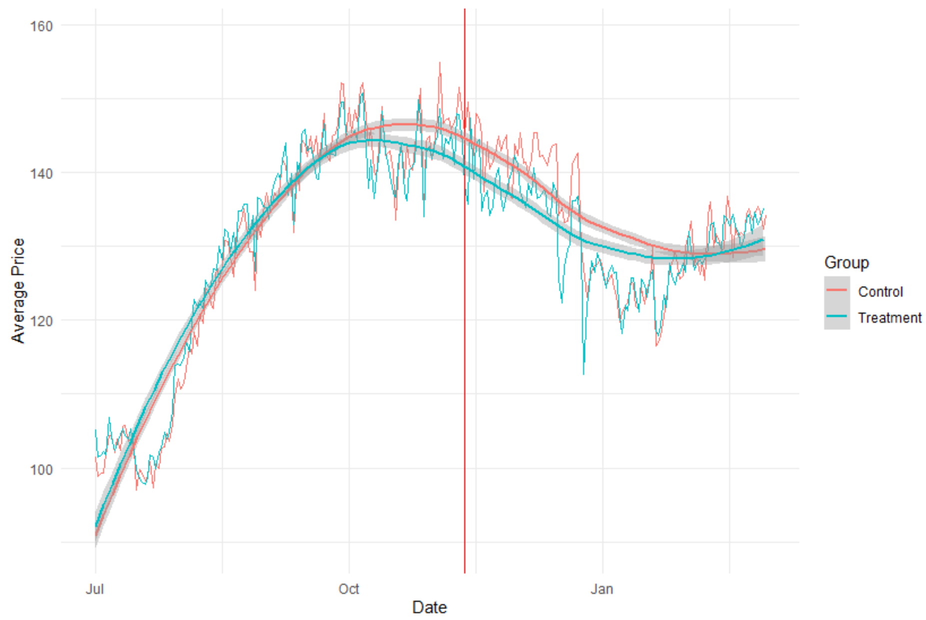


Figure 39: Average price, distribution over time for treatment group and control group

While the two groups separate, this occurs months before the date of first introduction and the trendlines remain parallel afterwards, until the beginning of February.

Regression analysis

The following section contains more information on the coefficients of every specification.

Speed of checkout

Figure 40 shows the behaviour of *speed* for the control group:

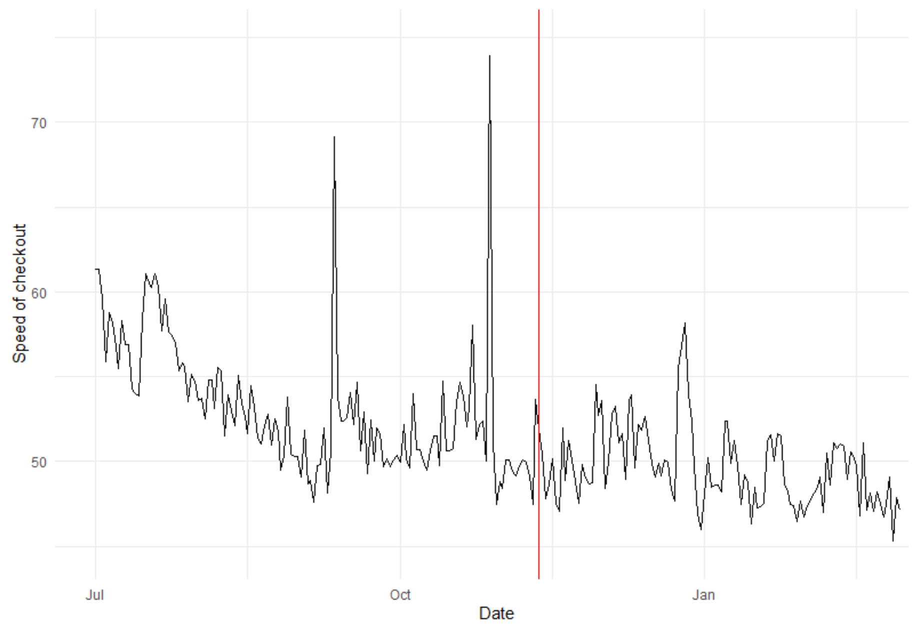
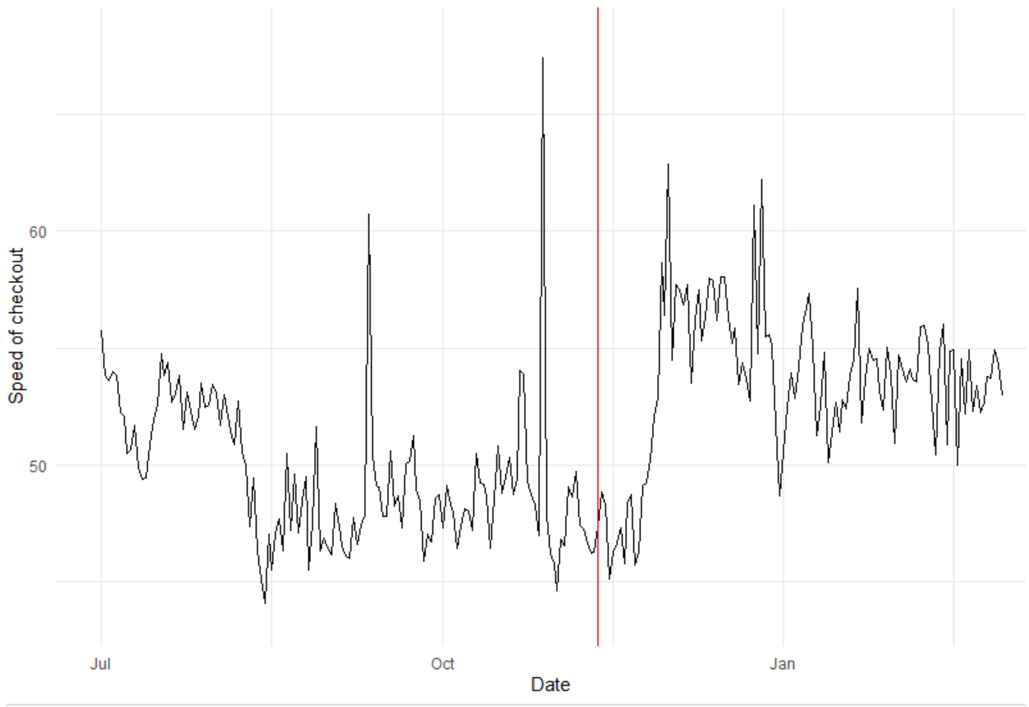


Figure 41 instead shows the same data for the treatment group:



Specifications

$$(1)\ Speed = \beta\ SST + \varepsilon$$

$$(2)\ Speed = \delta_t + \beta\ SST + \varepsilon$$

$$(3)\ Speed = \alpha_i + \beta\ SST + \varepsilon$$

$$(4)\ Speed = \alpha_i + \delta_t + \beta\ SST + \varepsilon$$

Robustness of results (specifications)

$$(1.1)\qquad Speed = \delta_{weekday} + \beta\ SST + \varepsilon$$

$$(1.2)\qquad Speed = \delta_{dayofmonth} + \beta\ SST + \varepsilon$$

$$(1.3)\qquad Speed = \delta_{dayofyear} + \beta\ SST + \varepsilon$$

$$(1.4)\qquad Speed = \alpha_i + \delta_t + \delta_{weekday} + \delta_{dayofmonth} + \delta_{dayofyear} + \beta\ SST + \varepsilon$$

F-tests

(2), monthly effects:

```
                F test for individual effects

data:  speed ~ SST
F = 992.63, df1 = 7, df2 = 2179249, p-value < 2.2e-16
alternative hypothesis: significant effects
```

(3), store effects:

```
                F test for individual effects

data:  speed ~ SST
F = 401.88, df1 = 7, df2 = 2179249, p-value < 2.2e-16
alternative hypothesis: significant effects
```

(1.1), weekday effects:

```
                F test for individual effects

data:  speed ~ SST
F = 29.955, df1 = 6, df2 = 2179250, p-value < 2.2e-16
alternative hypothesis: significant effects
```

(1.2), day-of-month effects:

```
                F test for individual effects

data:  speed ~ SST
F = 15.165, df1 = 30, df2 = 2179226, p-value < 2.2e-16
alternative hypothesis: significant effects
```

(1.3), day-of-year effects:

```
                F test for individual effects

data:  speed ~ SST
F = 32.419, df1 = 242, df2 = 2179014, p-value < 2.2e-16
alternative hypothesis: significant effects
```

Speed per item

Figure 42 shows the behaviour of the *Speed per item* for the control group:

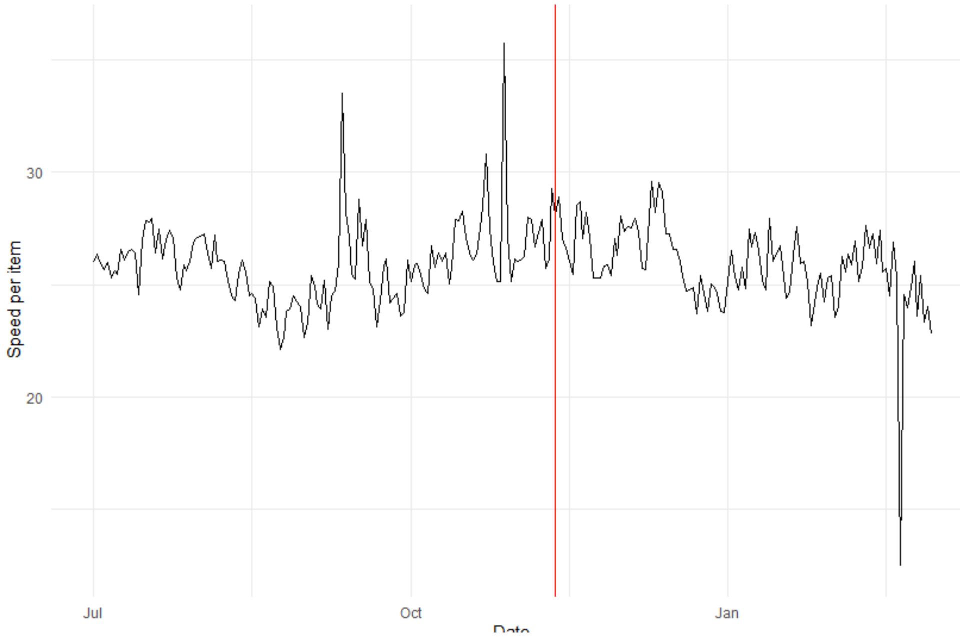
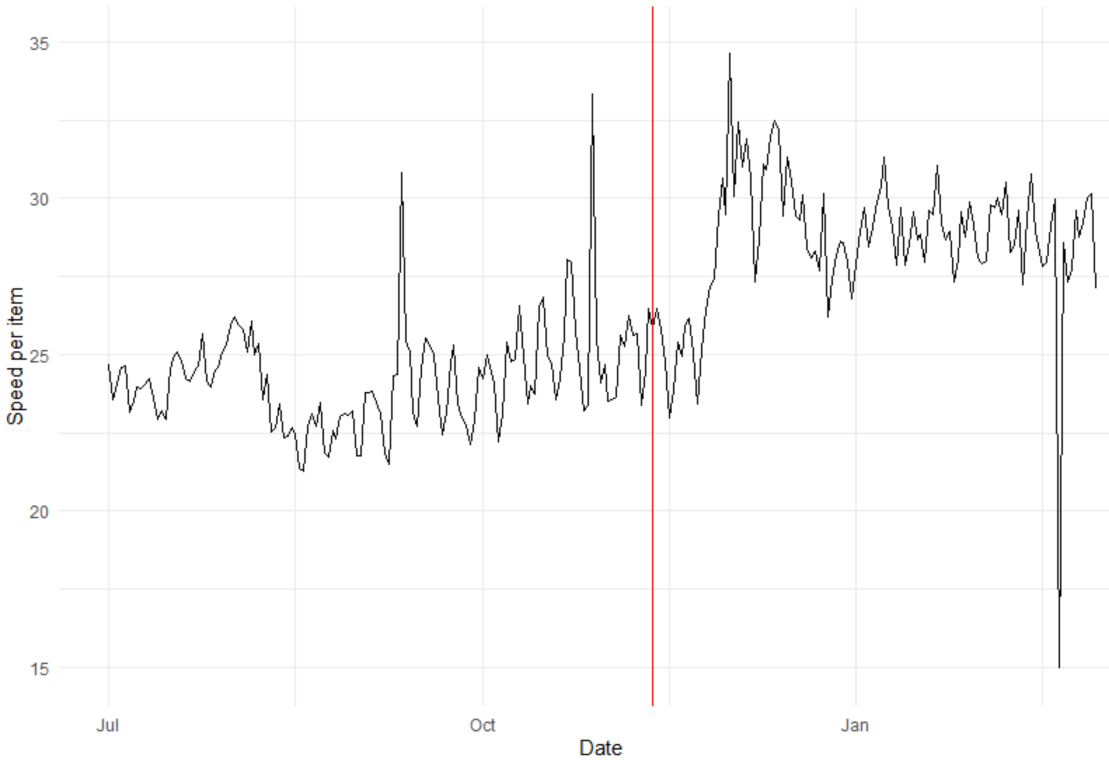


Figure 43 instead shows the same data for the treatment group:



Specifications

$$(1)\ Speed\ by\ item = \beta\ SST + \varepsilon$$

$$(2)\ Speed\ by\ item = \delta_t + \beta\ SST + \varepsilon$$

$$(3)\ Speed\ by\ item = \alpha_i + \beta\ SST + \varepsilon$$

$$(4)\ Speed\ by\ item = \alpha_i + \delta_t + \beta\ SST + \varepsilon$$

Robustness of results (Specifications)

$$(1.1)\qquad Speed\ by\ item = \delta_{weekday} + \beta\ SST + \varepsilon$$

$$(1.2)\qquad Speed\ by\ item = \delta_{dayofmonth} + \beta\ SST + \varepsilon$$

$$(1.3)\qquad Speed\ by\ item = \delta_{dayofyear} + \beta\ SST + \varepsilon$$

$$(1.4)\qquad Speed\ by\ item = \alpha_i + \delta_t + \delta_{weekday} + \delta_{dayofmonth} + \delta_{dayofyear} + \beta\ SST + \varepsilon$$

F-tests

(2), monthly effects:

```
        F test for individual effects

data:  speeditems ~ SST
F = 200.22, df1 = 7, df2 = 2062579, p-value < 2.2e-16
alternative hypothesis: significant effects
```

(3), store effects:

```
        F test for individual effects

data:  speeditems ~ SST
F = 482.74, df1 = 7, df2 = 2062579, p-value < 2.2e-16
alternative hypothesis: significant effects
```

(1.1), weekday effects:

```
        F test for individual effects

data:  speeditems ~ SST
F = 219.24, df1 = 6, df2 = 2062580, p-value < 2.2e-16
alternative hypothesis: significant effects
```

(1.2), day-of-month effects:

```
          F test for individual effects

data:  speeditems ~ SST
F = 33.293, df1 = 30, df2 = 2062556, p-value < 2.2e-16
alternative hypothesis: significant effects
```

(1.3), day-of-year effects:

```
          F test for individual effects

data:  speeditems ~ SST
F = 42.845, df1 = 242, df2 = 2062344, p-value < 2.2e-16
alternative hypothesis: significant effects
```

Service gap

Figure 44 shows the behaviour of the *Service gap* for the control group:

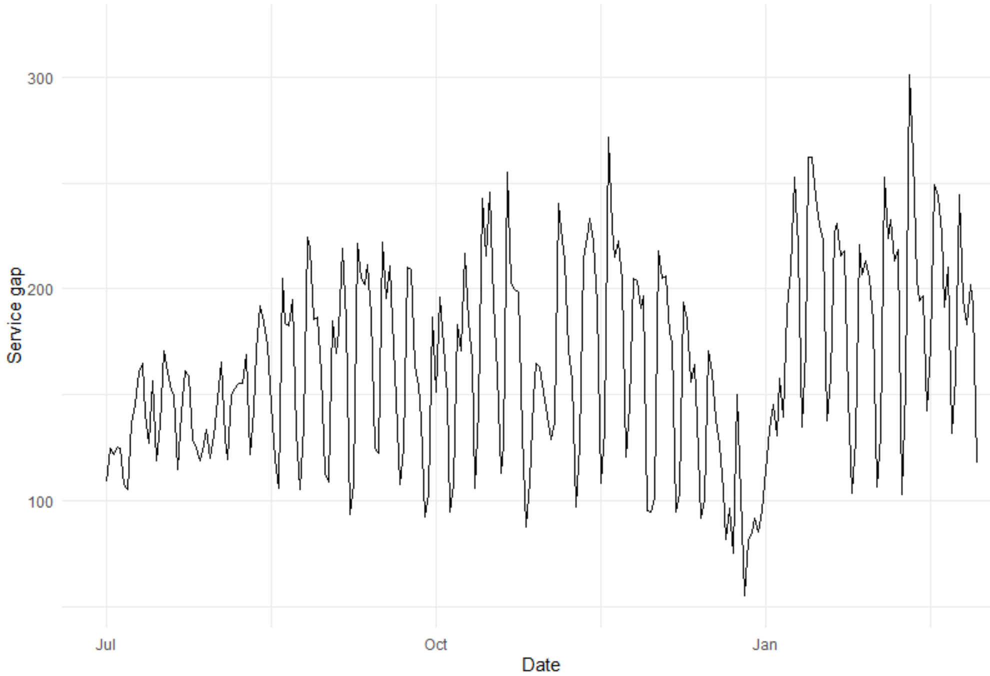
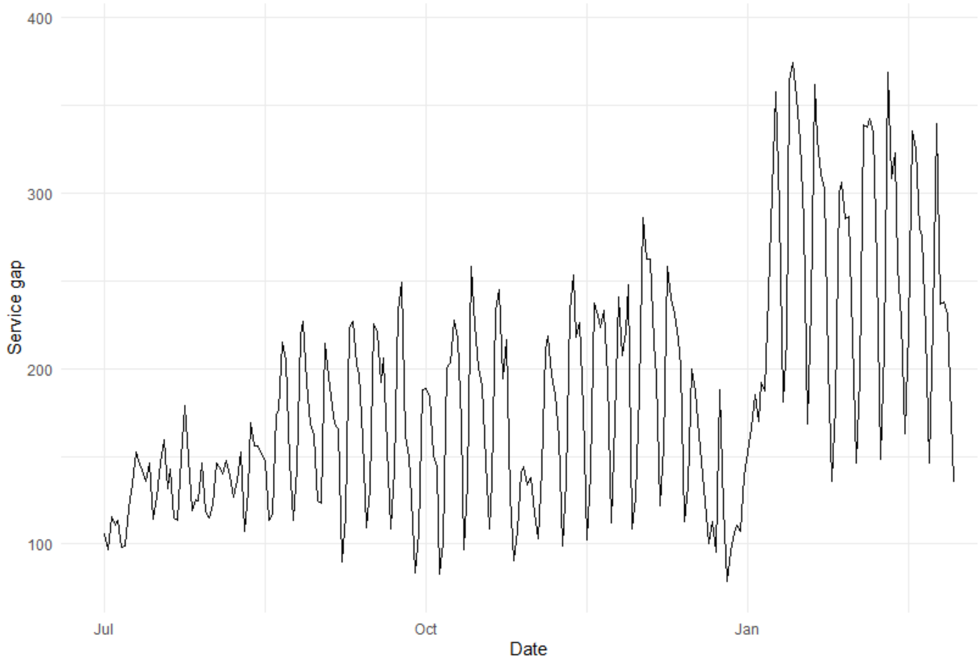


Figure 45 instead shows the same data for the treatment group:



Specifications

$$(1)\ Service\ gap = \beta\ SST + \varepsilon$$

$$(2)\ Service\ gap = \delta_t + \beta\ SST + \varepsilon$$

$$(3)\ Service\ gap = \alpha_i + \beta\ SST + \varepsilon$$

$$(4)\ Service\ gap = \alpha_i + \delta_t + \beta\ SST + \varepsilon$$

Robustness of results (Specifications)

$$(1.1)\qquad Service\ gap = \delta_{weekday} + \beta\ SST + \varepsilon$$

$$(1.2)\qquad Service\ gaps = \delta_{dayofmonth} + \beta\ SST + \varepsilon$$

$$(1.3)\qquad Service\ gap = \delta_{dayofyear} + \beta\ SST + \varepsilon$$

$$(1.4)\qquad Service\ gap = \alpha_i + \delta_t + \delta_{weekday} + \delta_{dayofmonth} + \delta_{dayofyear} + \beta\ SST + \varepsilon$$

Queuing

Service gap

Figure 46 shows the behaviour of the *Service gap* for the control group:

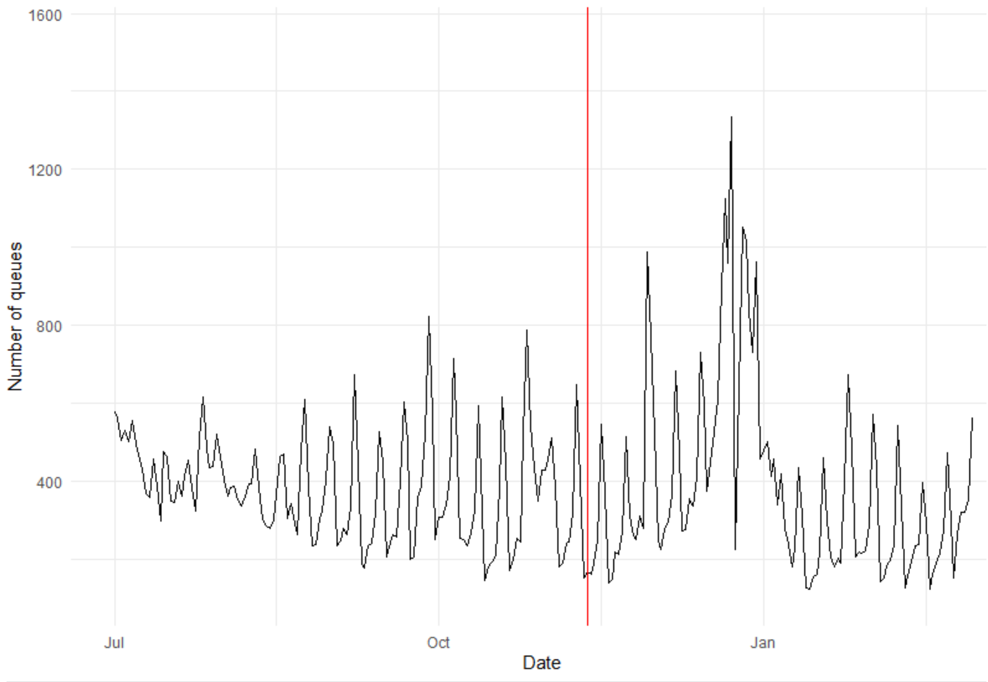
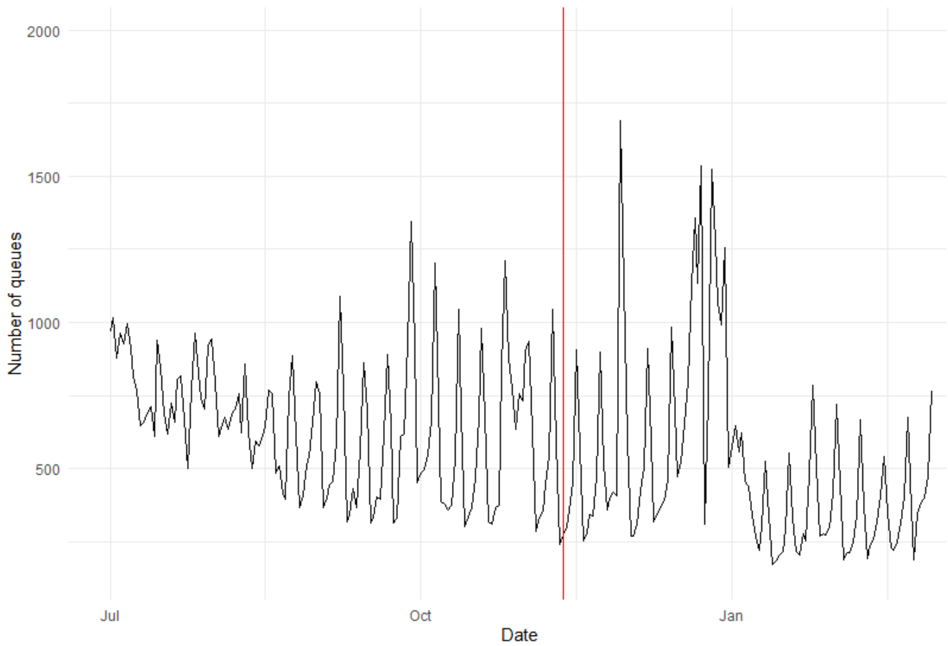


Figure 47 instead shows the same data for the treatment group:



Specifications

$$(1)\ Service\ Gap = \beta\ SST + \varepsilon$$

$$(2)\ Service\ Gap = \delta_t + \beta\ SST + \varepsilon$$

$$(3)\ Service\ Gap = \alpha_i + \beta\ SST + \varepsilon$$

$$(4)\ Service\ Gap = \alpha_i + \delta_t + \beta\ SST + \varepsilon$$

Robustness of results (Specifications)

$$(1.1) \qquad Service\ Gap = \delta_{weekday} + \beta\ SST + \varepsilon$$

$$(1.2) \qquad Service\ Gap = \delta_{dayofmonth} + \beta\ SST + \varepsilon$$

$$(1.3) \qquad Service\ Gap = \delta_{dayofyear} + \beta\ SST + \varepsilon$$

$$(1.4) \qquad Service\ Gap = \alpha_i + \delta_t + \delta_{weekday} + \delta_{dayofmonth} + \delta_{dayofyear} + \beta\ SST + \varepsilon$$

F-tests

(2), month effects:

```
        F test for individual effects

data:  sg ~ SST
F = 575.6, df1 = 7, df2 = 1292543, p-value < 2.2e-16
alternative hypothesis: significant effects
```

(3), store effects:

```
        F test for individual effects

data:  sg ~ SST
F = 374.01, df1 = 3, df2 = 1292547, p-value < 2.2e-16
alternative hypothesis: significant effects
```

(1.1), weekday effects:

```
        F test for individual effects

data:  sg ~ SST
F = 725.06, df1 = 6, df2 = 1292544, p-value < 2.2e-16
alternative hypothesis: significant effects
```

(1.2), day-of-month effects:

```
        F test for individual effects

data:  sg ~ SST
F = 56.553, df1 = 30, df2 = 1292520, p-value < 2.2e-16
alternative hypothesis: significant effects
```

(1.3), day-of-year effects:

```
        F test for individual effects

data:  sg ~ SST
F = 62.051, df1 = 242, df2 = 1292308, p-value < 2.2e-16
alternative hypothesis: significant effects
```

Number of queues

Figure 48 shows the behaviour of *number of queues* for the control group:

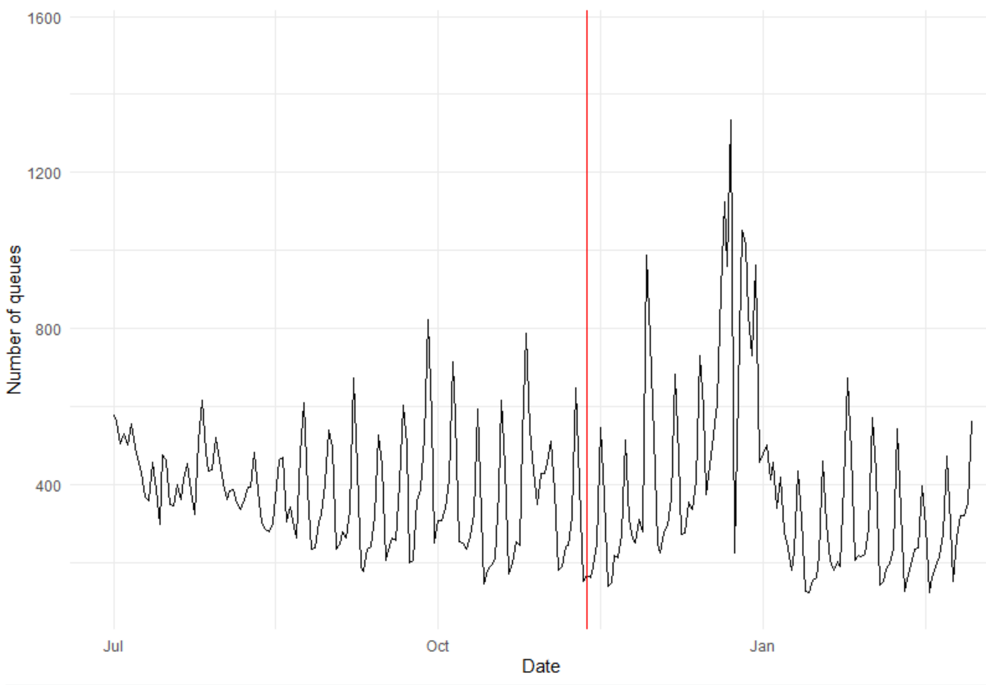
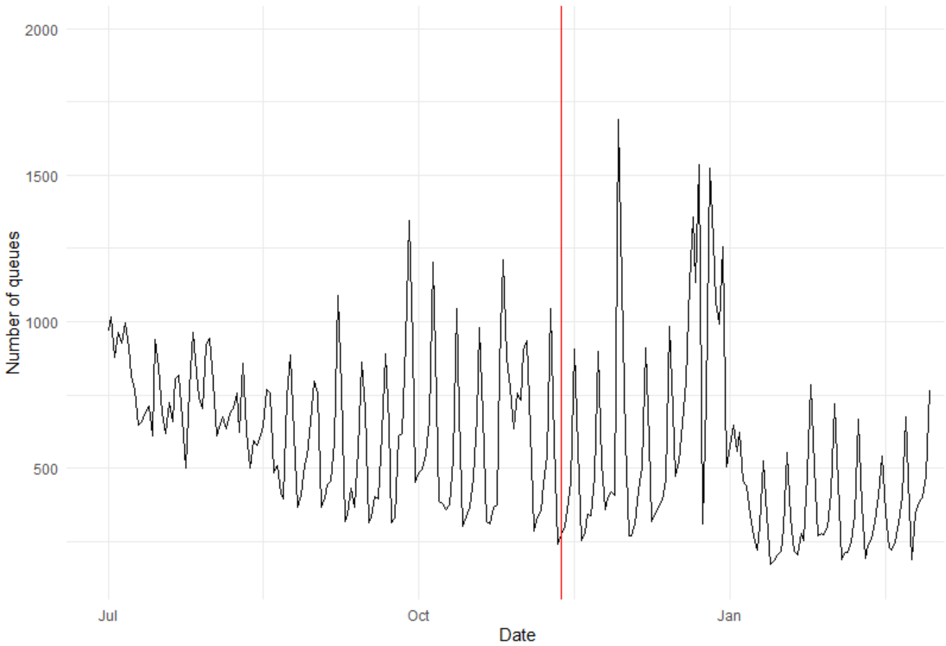


Figure 49 instead shows the same data for the treatment group:



Specifications

$$(1)\ Number\ of\ Queues = \beta\ SST + \varepsilon$$

$$(2)\ Number\ of\ Queues = \delta_t + \beta\ SST + \varepsilon$$

$$(3)\ Number\ of\ Queues = \alpha_i + \beta\ SST + \varepsilon$$

$$(4)\ Number\ of\ Queues = \alpha_i + \delta_t + \beta\ SST + \varepsilon$$

Robustness of results (Specifications)

$$(1.1) \qquad Number\ of\ Queues = \delta_{weekday} + \beta\ SST + \varepsilon$$

$$(1.2) \qquad Number\ of\ Queues = \delta_{dayofmonth} + \beta\ SST + \varepsilon$$

$$(1.3) \qquad Number\ of\ Queues = \delta_{dayofyear} + \beta\ SST + \varepsilon$$

$$(1.4) \qquad Number\ of\ Queues = \alpha_i + \delta_t + \delta_{weekday} + \delta_{dayofmonth} + \delta_{dayofyear} + \beta\ SST + \varepsilon$$

F-tests (broad definition of *Queue*)

(2), month effects:

```
              F test for individual effects

data:  count ~ SST
F = 44.489, df1 = 7, df2 = 1927, p-value < 2.2e-16
alternative hypothesis: significant effects
```

(3), store effects:

```
              F test for individual effects

data:  count ~ SST
F = 127.64, df1 = 7, df2 = 1927, p-value < 2.2e-16
alternative hypothesis: significant effects
```

(1.1), weekday effects:

```
              F test for individual effects

data:  count ~ SST
F = 58.247, df1 = 6, df2 = 1928, p-value < 2.2e-16
alternative hypothesis: significant effects
```

(1.2), day-of-month effects:

```
              F test for individual effects

data:  count ~ SST
F = 4.4231, df1 = 30, df2 = 1904, p-value = 2.814e-14
alternative hypothesis: significant effects
```

(1.3), day-of-year effects:

```
              F test for individual effects

data:  count ~ SST
F = 8.3835, df1 = 242, df2 = 1692, p-value < 2.2e-16
alternative hypothesis: significant effects
```

F-tests (narrow definition of *Queue*)

(2), month effects:

```
              F test for individual effects

data:  count ~ SST
F = 48.509, df1 = 7, df2 = 1927, p-value < 2.2e-16
alternative hypothesis: significant effects
```

(3), store effects:

```
        F test for individual effects

data:  count ~ SST
F = 166.76, df1 = 7, df2 = 1927, p-value < 2.2e-16
alternative hypothesis: significant effects
```

(1.1), weekday effects:

```
        F test for individual effects

data:  count ~ SST
F = 46.283, df1 = 6, df2 = 1928, p-value < 2.2e-16
alternative hypothesis: significant effects
```

(1.2), day-of-month effects:

```
        F test for individual effects

data:  count ~ SST
F = 3.6203, df1 = 30, df2 = 1904, p-value = 1.768e-10
alternative hypothesis: significant effects
```

(1.3), day-of-year effects:

```
        F test for individual effects

data:  count ~ SST
F = 6.2011, df1 = 242, df2 = 1692, p-value < 2.2e-16
alternative hypothesis: significant effects
```

Proportion of queues

Figure 50 shows the behaviour of *Proportion of Queues* for the control group:

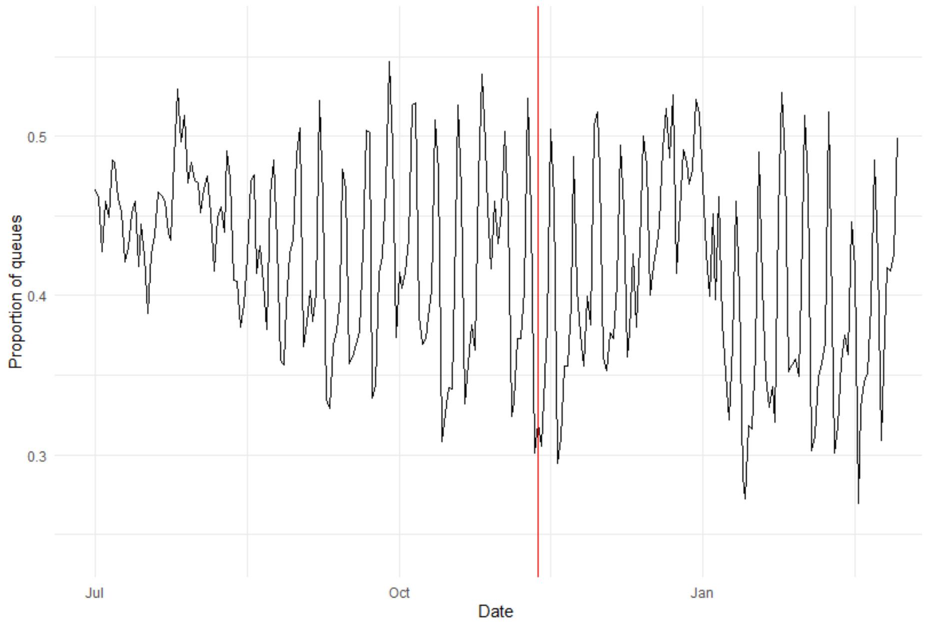
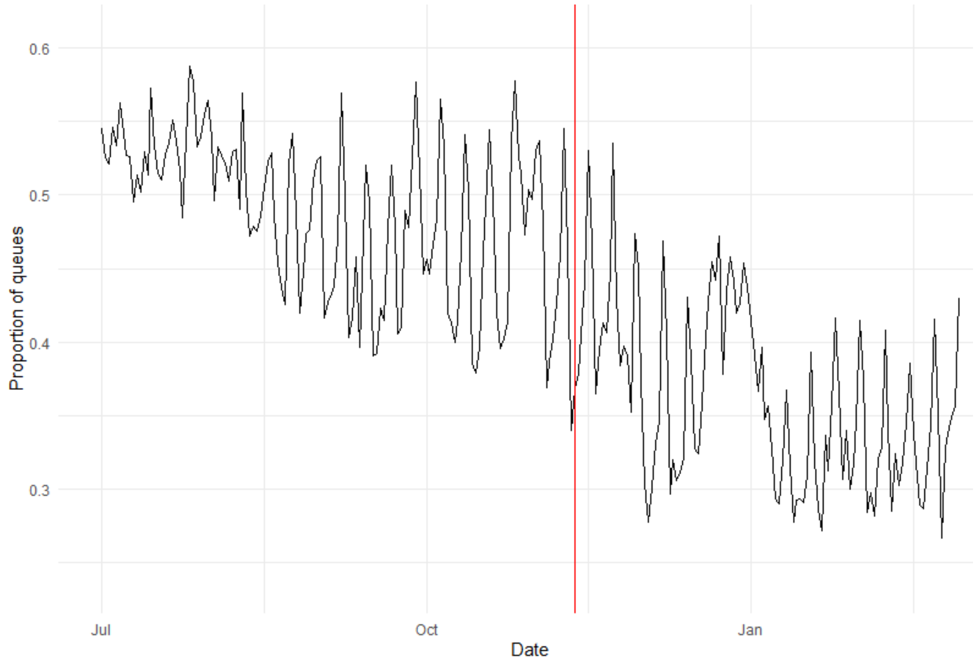


Figure 51 instead shows the same data for the treatment group:



Specifications

$$(1)\ Proportion\ of\ Queues = \beta\ SST + \varepsilon$$

$$(2)\ Proportion\ of\ Queues = \delta_t + \beta\ SST + \varepsilon$$

$$(3)\ Proportion\ of\ Queues = \alpha_i + \beta\ SST + \varepsilon$$

$$(4)\ Proportion\ of\ Queues = \alpha_i + \delta_t + \beta\ SST + \varepsilon$$

Robustness of results (Specifications)

$$(1.1)\qquad Proportion\ of\ Queues = \delta_{weekday} + \beta\ SST + \varepsilon$$

$$(1.2)\qquad Proportion\ of\ Queues = \delta_{dayofmonth} + \beta\ SST + \varepsilon$$

$$(1.3)\qquad Proportion\ of\ Queues = \delta_{dayofyear} + \beta\ SST + \varepsilon$$

$$(1.4)\qquad Proportion\ of\ Queues = \alpha_i + \delta_t + \delta_{weekday} + \delta_{dayofmonth} + \delta_{dayofyear} + \beta\ SST + \varepsilon$$

F-tests (Broad definition of *Queue*)

(2), month fixed effects:

```
        F test for individual effects

data:  prop ~ SST
F = 57.954, df1 = 7, df2 = 1927, p-value < 2.2e-16
alternative hypothesis: significant effects
```

(3), store fixed effects:

```
        F test for individual effects

data:  prop ~ SST
F = 68.49, df1 = 7, df2 = 1927, p-value < 2.2e-16
alternative hypothesis: significant effects
```

(1.1), weekday effects:

```
        F test for individual effects

data:  prop ~ SST
F = 109.17, df1 = 6, df2 = 1928, p-value < 2.2e-16
alternative hypothesis: significant effects
```

(1.2), day-of-month effects:

```
        F test for individual effects

data:  prop ~ SST
F = 5.3219, df1 = 30, df2 = 1904, p-value < 2.2e-16
alternative hypothesis: significant effects
```

(1.3), day-of-year effects:

```
        F test for individual effects

data:  conversion_rate ~ SST
F = 59.154, df1 = 288, df2 = 667, p-value < 2.2e-16
alternative hypothesis: significant effects
```

F-tests (Narrow definition of *Queue*)

(2), month fixed effects:

```
        F test for individual effects

data:  prop ~ SST
F = 66.031, df1 = 7, df2 = 1927, p-value < 2.2e-16
alternative hypothesis: significant effects
```

(3), store fixed effects:

```
            F test for individual effects

data:  prop ~ SST
F = 110.2, df1 = 7, df2 = 1927, p-value < 2.2e-16
alternative hypothesis: significant effects
```

(1.1), weekday effects:

```
            F test for individual effects

data:  prop ~ SST
F = 35.002, df1 = 6, df2 = 1928, p-value < 2.2e-16
alternative hypothesis: significant effects
```

(1.2), day-of-month effects:

```
            F test for individual effects

data:  prop ~ SST
F = 2.9061, df1 = 30, df2 = 1904, p-value = 2.737e-07
alternative hypothesis: significant effects
```

(1.3), day-of-year effects:

```
            F test for individual effects

data:  prop ~ SST
F = 5.5775, df1 = 242, df2 = 1692, p-value < 2.2e-16
alternative hypothesis: significant effects
```

Conversion rates

Figure 52 shows the behaviour of *conversion rates* for the control group:

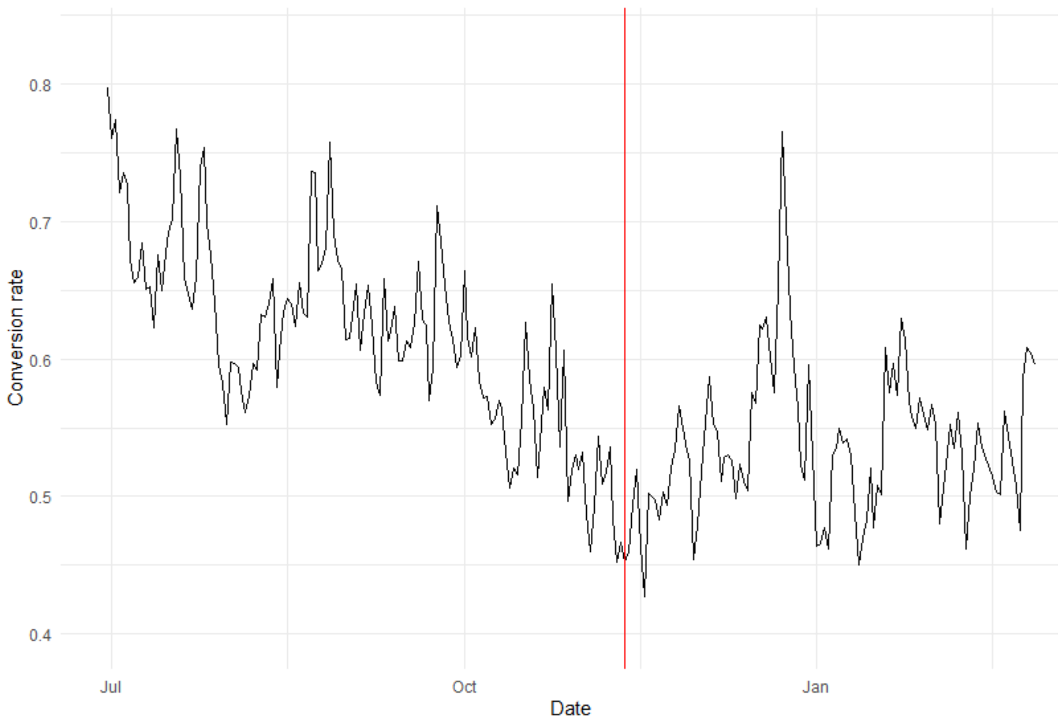
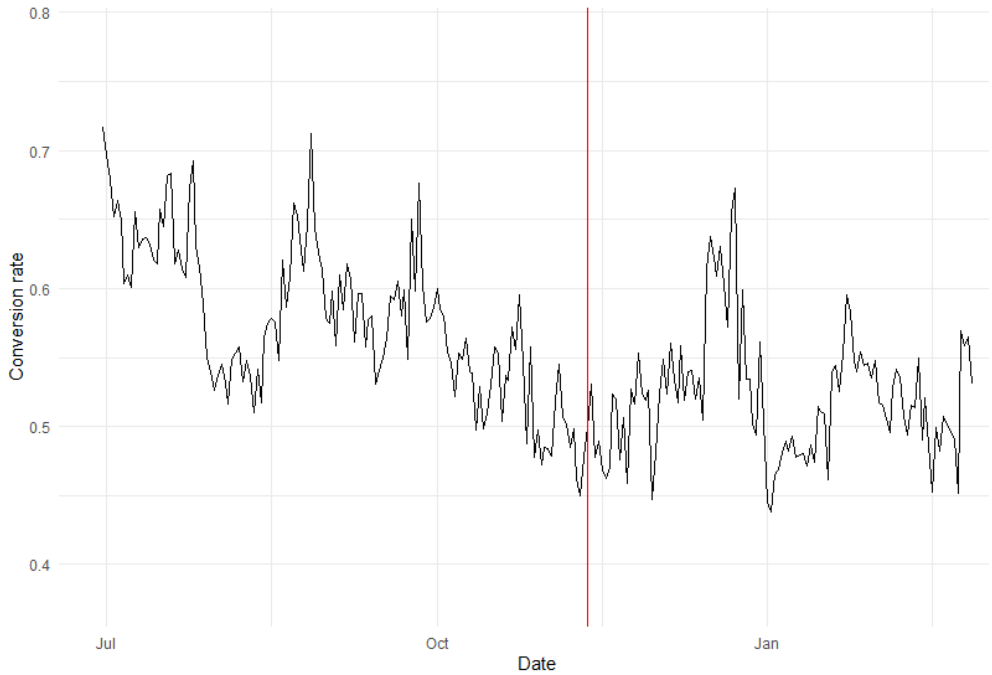


Figure 53 instead shows the same data for the treatment group:



Specifications

$$(1)\ \ Conversion\ Rate = \beta\ SST + \varepsilon$$

$$(2)\ \ Conversion\ Rate = \delta_t + \beta\ SST + \varepsilon$$

$$(3)\ \ Conversion\ Rate = \alpha_i + \beta\ SST + \varepsilon$$

$$(4)\ \ Conversion\ Rate = \alpha_i + \delta_t + \beta\ SST + \varepsilon$$

Robustness of results (Specifications)

$$(1.1)\qquad Conversion\ Rate = \delta_{weekday} + \beta\ SST + \varepsilon$$

$$(1.2)\qquad Conversion\ Rate = \delta_{dayofmonth} + \beta\ SST + \varepsilon$$

$$(1.3)\qquad Conversion\ Rate = \delta_{dayofyear} + \beta\ SST + \varepsilon$$

$$(1.4)\qquad Conversion\ Rate = \alpha_i + \delta_t + \delta_{weekday} + \delta_{dayofmonth} + \delta_{dayofyear} + \beta\ SST + \varepsilon$$

F tests (full sample)

(2), monthly effects:

```
        F test for individual effects

data:  conversion_rate ~ SST
F = 21.878, df1 = 5, df2 = 680, p-value < 2.2e-16
alternative hypothesis: significant effects
```

(3), store effects:

```
        F test for individual effects

data:  conversion_rate ~ SST
F = 432.6, df1 = 3, df2 = 682, p-value < 2.2e-16
alternative hypothesis: significant effects
```

(1.1), weekday effects:

```
        F test for individual effects

data:  conversion_rate ~ SST
F = 332.11, df1 = 58, df2 = 897, p-value < 2.2e-16
alternative hypothesis: significant effects
```

(1.2), day-of-month effects:

```
        F test for individual effects

data:  conversion_rate ~ SST
F = 0.84027, df1 = 30, df2 = 875, p-value = 0.7129
alternative hypothesis: significant effects
```

The F-test suggests not to include day-of-month effects.

(1.3), day-of-year effects:

```
        F test for individual effects

data:  conversion_rate ~ SST
F = 1.0092, df1 = 238, df2 = 667, p-value = 0.4592
alternative hypothesis: significant effects
```

F tests (narrow sample)

(2), monthly effects:

```
        F test for individual effects

data:  conversion_rate ~ SST
F = 3.1306, df1 = 2, df2 = 257, p-value = 0.04536
alternative hypothesis: significant effects
```

(3), store effects:

```
F test for individual effects

data:  conversion_rate ~ SST
F = 229.13, df1 = 3, df2 = 256, p-value < 2.2e-16
alternative hypothesis: significant effects
```

(1.1), weekday effects:

```
F test for individual effects

data:  conversion_rate ~ SST
F = 1.0078, df1 = 6, df2 = 253, p-value = 0.4206
alternative hypothesis: significant effects
```

(1.2), day of month effects:

```
F test for individual effects

data:  conversion_rate ~ SST
F = 0.61503, df1 = 30, df2 = 229, p-value = 0.9439
alternative hypothesis: significant effects
```

(1.3), day-of-year effects:

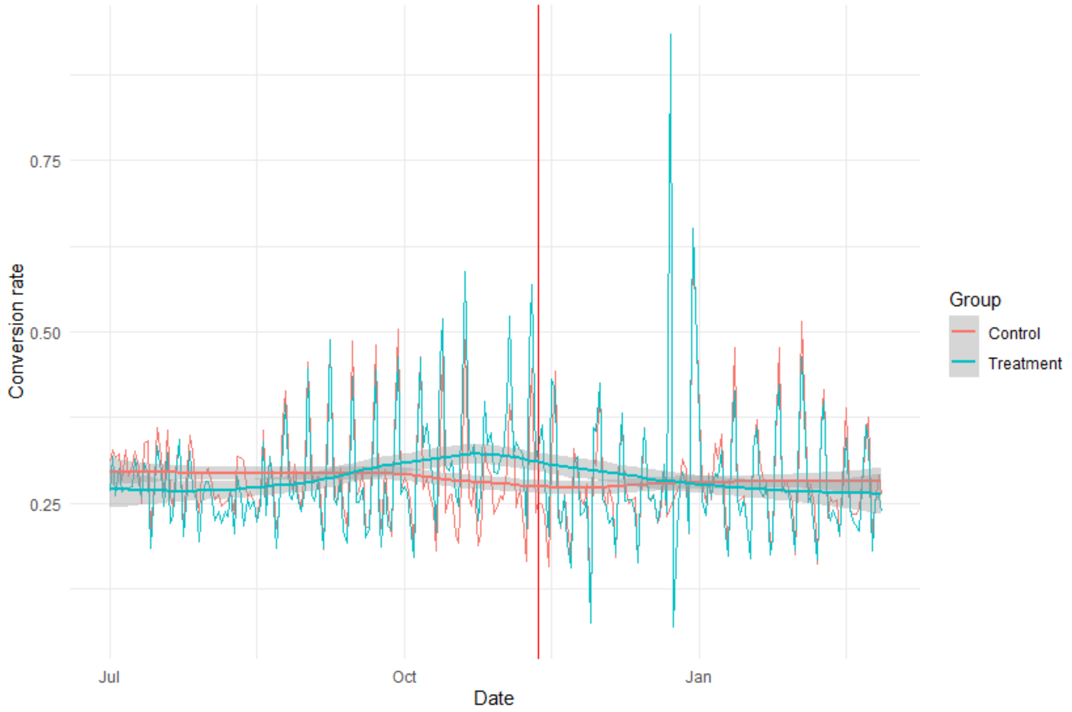
```
F test for individual effects

data:  conversion_rate ~ SST
F = 0.50016, df1 = 75, df2 = 184, p-value = 0.9996
alternative hypothesis: significant effects
```

***Conversion Rate*, alternative definition:**

$$Conversion\ rate = \frac{Number\ of\ transactions}{Number\ of\ Visitors}$$

The distribution over time of this variable for the treatment and control groups is shown in figure 54:



The parallel trends assumptions does not seem to hold.

I now execute all the models used for the main measure for the *Conversion Rate*:

Table 1: Conversion rate, alternative definition

	OLS (1)	Month FE (2)	Store FE (3)	Final (4)
SST	-0.018694* (0.008)	-0.0488 (0.0284)	-0.0181* (0.007)	-0.04838 (0.02667)
N	3128	3128	3128	3128
Adjusted R^2	0.004641	0. 8541	0.6918	0.973
Time FE	No	Yes	No	Yes
Store FE	No	No	Yes	Yes

The standard errors are in parentheses; * p < 0.10, ** p <0.05, *** p <0.01

Table 2: Conversion rate, alternative definition (seasonalities)

	Weekday (1.1)	Day-of-month (1.2)	Day-of-year (1.3)	All time effects (1.4)	Final (4)
SST	-0.01906* (0.00742)	-0.01859* (0.00809)	0.03405 (0.0422)	0.038188 (0.03752)	-0.04838 (0.02667)
N	3128	3128	3128	3128	3128
Adjusted R^2	0.874	0.8521	0.8924	0.9174	0.973

Store FE	No	No	No	Yes	Yes

The standard errors are in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The coefficients do not vary greatly between models but are only barely significant in in models (1), (3), (1.1) and (1.2), while all other coefficients are not significant. No effect of SSTs on conversion rates was thus found.

Productivity measure (a)

Figure 55 shows the behaviour of *productivity measure (a)* for the control group:

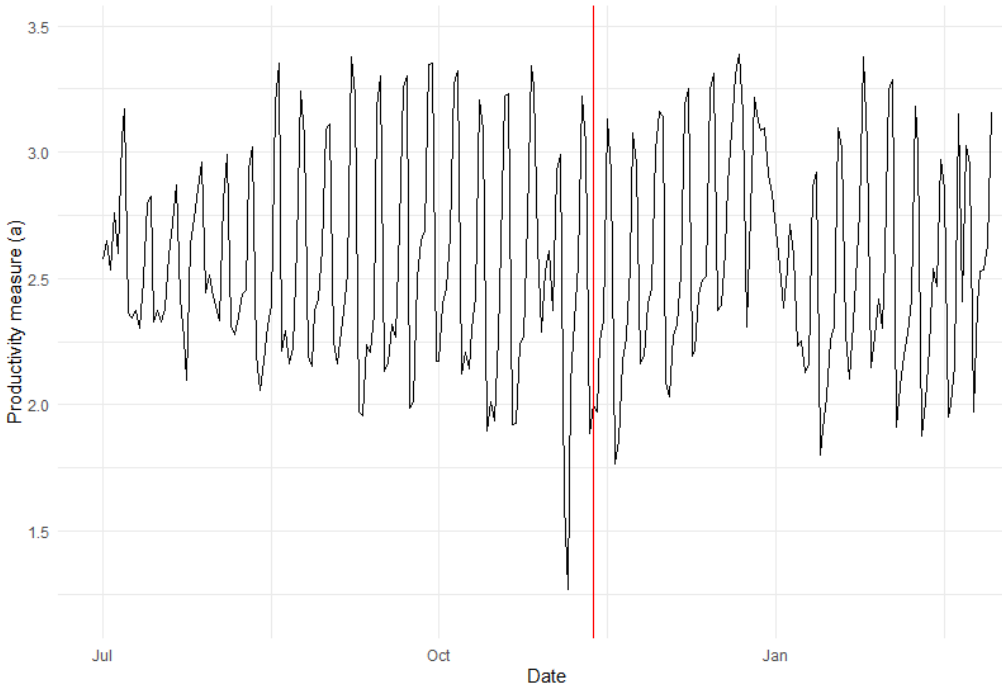
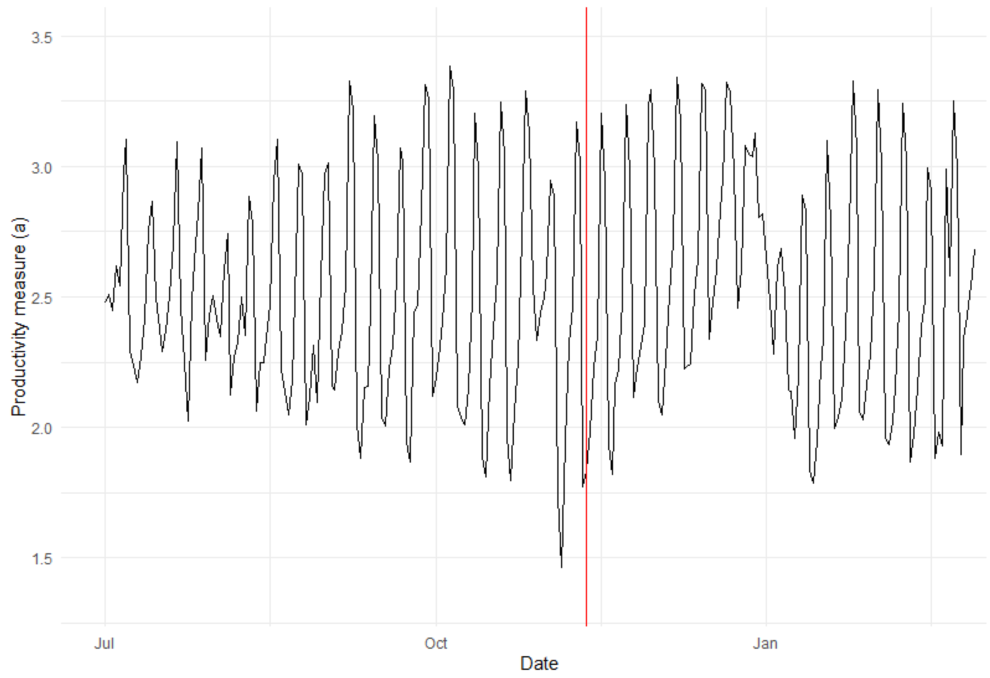


Figure 56 instead shows the same data for the treatment group:



Specifications

$$(1)\ \log\left(\frac{Quantity}{employee\ hours}\right) = \beta\ SST + \varepsilon$$

$$(2)\ \log\left(\frac{Quantity}{employee\ hours}\right) = \delta_t + \beta\ SST + \varepsilon$$

$$(3)\ \log\left(\frac{Quantity}{employee\ hours}\right) = \alpha_i + \beta\ SST + \varepsilon$$

$$(4)\ \log\left(\frac{Quantity}{employee\ hours}\right) = \alpha_i + \delta_t + \beta\ SST + \varepsilon$$

Robustness of results (Specifications)

$$(1.1)\qquad \log\left(\frac{Quantity}{employee\ hours}\right) = \delta_{month} + \beta\ SST + \varepsilon$$

$$(1.2)\qquad \log\left(\frac{Quantity}{employee\ hours}\right) = \delta_{dayofmonth} + \beta\ SST + \varepsilon$$

$$(1.3)\qquad \log\left(\frac{Quantity}{employee\ hours}\right) = \delta_{dayofyear} + \beta\ SST + \varepsilon$$

$$(1.4)\qquad \log\left(\frac{Quantity}{employee\ hours}\right) = \alpha_i + \delta_t + \delta_{month} + \delta_{dayofmonth} + \delta_{dayofyear} + \beta\ SST + \varepsilon$$

F-tests (full sample)

(2), weekday effects:

```
                F test for individual effects

data:  prod2 ~ SST
F = 305.55, df1 = 6, df2 = 949, p-value < 2.2e-16
alternative hypothesis: significant effects
```

(3), store effects:

```
                F test for individual effects

data:  prod2 ~ SST
F = 7.393, df1 = 3, df2 = 952, p-value = 6.726e-05
alternative hypothesis: significant effects
```

(1.1), monthly effects:

```
                F test for individual effects

data:  prod2 ~ SST
F = 7.435, df1 = 7, df2 = 948, p-value = 1.006e-08
alternative hypothesis: significant effects
```

(1.2), day-of-month effects:

```
                F test for individual effects

data:  prod2 ~ SST
F = 2.0315, df1 = 30, df2 = 925, p-value = 0.0009446
alternative hypothesis: significant effects
```

(1.3), day-of-year effects:

```
                F test for individual effects

data:  prod2 ~ SST
F = 22.81, df1 = 241, df2 = 714, p-value < 2.2e-16
alternative hypothesis: significant effects
```

F-tests (narrow sample)

(2), weekday effects:

```
        F test for individual effects

data:  prod2 ~ SST
F = 128.45, df1 = 6, df2 = 288, p-value < 2.2e-16
alternative hypothesis: significant effects
```

(3), store effects:

```
        F test for individual effects

data:  prod2 ~ SST
F = 3.3559, df1 = 3, df2 = 291, p-value = 0.01931
alternative hypothesis: significant effects
```

(1.1), monthly effects:

```
        F test for individual effects

data:  prod2 ~ SST
F = 0.60047, df1 = 2, df2 = 292, p-value = 0.5492
alternative hypothesis: significant effects
```

The F-tests suggests not to include monthly effects.

(1.2), day of month effects:

```
        F test for individual effects

data:  prod2 ~ SST
F = 2.5557, df1 = 30, df2 = 264, p-value = 3.823e-05
alternative hypothesis: significant effects
```

(1.3), day-of-year effects:

```
        F test for individual effects

data:  prod2 ~ SST
F = 25.58, df1 = 75, df2 = 219, p-value < 2.2e-16
alternative hypothesis: significant effects
```

Productivity measure (b)

Figure 57 shows the behaviour of *Productivity Measure (b)* for the control group:

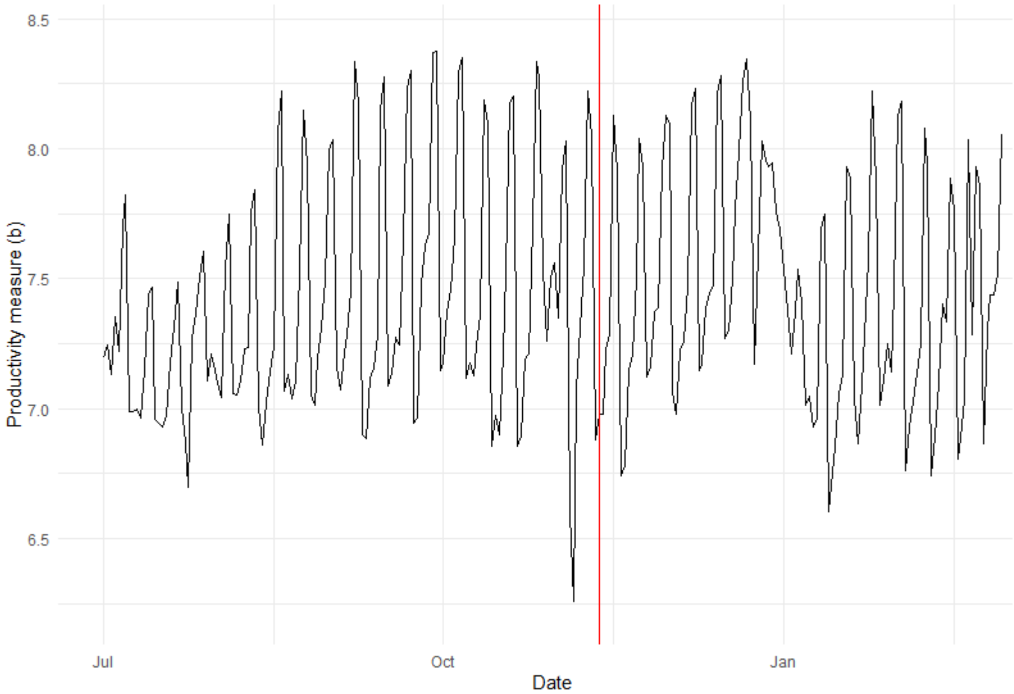
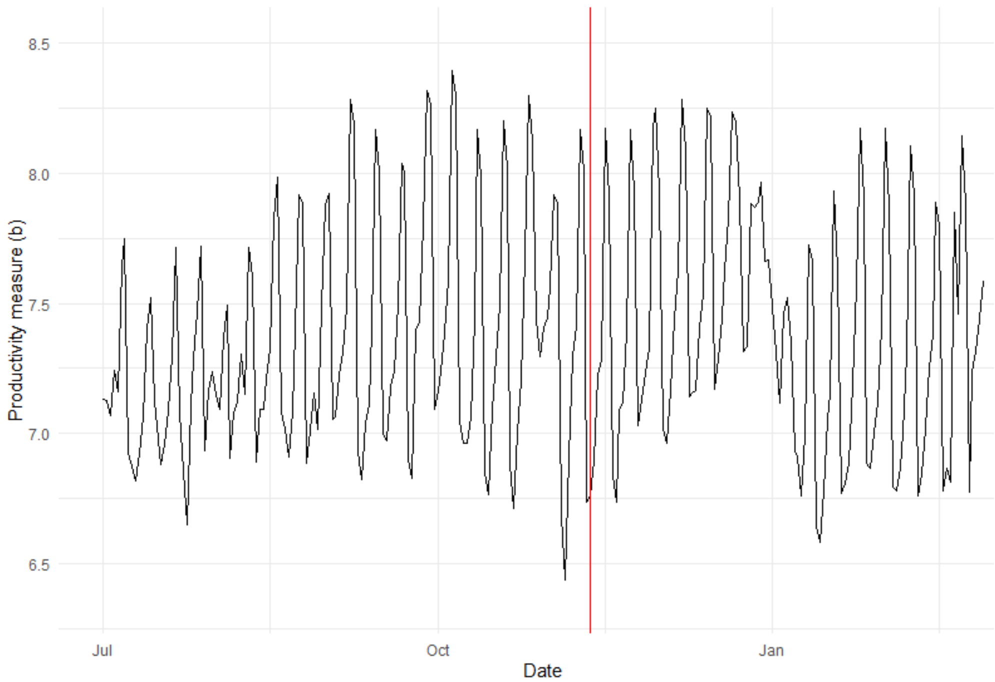


Figure 58 instead shows the same data for the treatment group:



Specifications

$$(1)\ \log\left(\frac{Net\ Sales}{employee\ hours}\right) = \beta\ SST + \ \varepsilon$$

$$(2)\ \log\left(\frac{Net\ Sales}{employee\ hours}\right) = \beta\ SST + price + \ \varepsilon$$

$$(3)\ \log\left(\frac{Net\ Sales}{employee\ hours}\right) = \delta_t + \beta\ SST + \varepsilon$$

$$(4)\ \log\left(\frac{Net\ Sales}{employee\ hours}\right) = \alpha_i + \beta\ SST + \ \varepsilon$$

$$(5)\ \log\left(\frac{Net\ Sales}{employee\ hours}\right) = \alpha_i + \delta_t + \beta\ SST\ + price + \ \varepsilon$$

Robustness of results (Specifications)

$$(1.1)\qquad log\left(\frac{Net\ Sales}{employee\ hours}\right) = \delta_{month} + \beta\ SST\ + \ \varepsilon$$

$$(1.2)\qquad log\left(\frac{Net\ Sales}{employee\ hours}\right) = \delta_{dayofmonth} + \beta\ SST + \ \varepsilon$$

$$(1.3)\qquad log\left(\frac{Net\ Sales}{employee\ hours}\right) = \delta_{dayofyear} + \beta\ SST\ + \ \varepsilon$$

$$(1.4)\qquad log\left(\frac{Net\ Sales}{employee\ hours}\right) = price + \alpha_i + \delta_t + \delta_{month} + \delta_{dayofmonth} + \delta_{dayofyear} + \beta\ SST + \ \varepsilon$$

F-tests (full sample)

(2), weekday effects:

```
        F test for individual effects

data:  prod ~ SST + price
F = 302.98, df1 = 6, df2 = 948, p-value < 2.2e-16
alternative hypothesis: significant effects
```

(3), store effects:

```
        F test for individual effects

data:  prod ~ SST + price
F = 7.463, df1 = 3, df2 = 951, p-value = 6.096e-05
alternative hypothesis: significant effects
```

(1.1), monthly effects:

```
        F test for individual effects

data:  prod ~ SST
F = 13.361, df1 = 7, df2 = 948, p-value < 2.2e-16
alternative hypothesis: significant effects
```

(1.2), day-of-month effects:

```
        F test for individual effects

data:  prod ~ SST
F = 1.9035, df1 = 30, df2 = 925, p-value = 0.002559
alternative hypothesis: significant effects
```

(1.3), day-of-year effects:

```
        F test for individual effects

data:  prod ~ SST
F = 22.664, df1 = 241, df2 = 714, p-value < 2.2e-16
alternative hypothesis: significant effects
```

References

Econometrics

Angrist, J. D. & Pischke, J. S. 2009, *Mostly Harmless Econometrics: An Empiricist's Companion*, 1st edn, Princeton University Press.

Wooldridge, M. J. 2001, *Introductory Econometrics: A Modern Approach,* 2nd edn, South-Western College Pub, Cincinnati.

Queuing (and Productivity)

Schmenner, R. W. & Swink, M. L. 1998, 'On Theory in Operations Management', *Journal of Operations Management*, vol. 17, no. 1, pp. 97-113.

Service Quality

Anselmsson, J. 2001, 'Customer-Perceived Service-Quality and Technology-Based Self Service', PhD. thesis, Lund University, Lund.

Bateson, J. E. 1985, 'Self service consumer: An exploratory study'. *Journal of Retailing*, vol. 61, no. 3, pp. 49–76.

Dabholkar, P. A., Thorpe, D. I., & Rentz, J. O. 1996, 'A Measure of Service Quality for Retail Stores: Scale Development and Validation', *Journal of the Academy of Marketing Science*, vol 24, no. 1, pp. 3-16.

Dabholkar, P., Bobbitt, M. L. & Lee, E. 2003, ʾUnderstanding consumer motivation and behavior related to self-scanning in retailing: Implications for strategy and research on technology-based self service', *International Journal of Service Industry Management*, vol. 14 no. 1, pp. 59-95.

Davis, F. 1989, 'Perceived usefulness, perceived ease of use and user acceptance information technology'. *MIS Quarterly*, vol. 13, no. 3, pp. 319–340.

Marzocchi, G.L. & Zammit, A. 2006, 'Self-scanning technologies in retail: Determinants of adoption', *The Service Industries Journal*, vol. 26, no. 6, pp. 651-669.

Fernandes, T. & Pedroso, R. 2017, 'The effect of self-checkout quality on customer satisfaction and repatronage in a retail context', *Service Business*, vol. 11, no. 1, pp. 69-92.

Productivity

Bartelsman, E. J. & Doms, M. 2000, 'Understanding Productivity: Lessons from Longitudinal Microdata', *Journal of Economic Literature,* vol. 38, no. 3, pp. 569-594.

Basker, E. 2012, 'Raising the Barcode Scanner: Technology and Productivity in the Retail Sector', *American Economic Journal: Applied Economics*, vol. 4, n. 3, pp. 1-27.

Basker, E. 2015, 'Change at the Checkout: Tracing the Impact of a Process Innovation', *Journal of Industrial Economics*, vol. 63, no. 2, pp. 339-370.

Basker, E., Foster, L. & Klimek, S. D. 2017, 'Customer-Employee Substitution: Evidence from Gasoline Stations', *Journal of Economics & Management Strategy*, vol. 26, no. 4, pp. 876-896.

Benkard, C. L. 2000, 'Learning and Forgetting: The Dynamics of Aircraft Production', *American Economic Review*, vol. 90, no. 4, pp. 1034-1054.

Bernard, A. B. Redding, S. J. & Schott, P. K. 2010, 'Multiple-Product Firms and Product Switching.' *American Economic Review*, vol. 100, no. 1, pp. 70-97.

Bloom, N. & Van Reenen, J. 2007, 'Measuring and Explaining Management Practices Across Firms and Countries', *The Quarterly Journal of Economics*, vol. 122, no. 4, pp. 1351-1408.

Brynjolfsson, E. McAfee, A. Sorell, M. & Zhu, F. 2008, Scale Without Mass: Business Process Replication and Industry Dynamics, *Harvard Business School Technology & Operations Mgt. Unit Research Paper,* No. 07-016, Cambridge (MA), Unites States.

Foster, L. S. Haltiwanger, J. & Krizan, C. J. 2006, 'Market Selection, Reallocation, and Restructuring in the U.S. Retail Trade Sector in the 1990s', *The Review of Economics and Statistics*, vol. 88, no. 4, pp. 748-758.

Hamilton, B. H. Nickerson, J. A. & Owan, H. 2003, 'Team Incentives and Worker Heterogeneity: An Empirical Analysis of the Impact of Teams on Productivity and Participation', *Journal of Political Economy,* vol. 111, no. 3, pp. 465-497.

Ichniowski, C. Shaw, K. L. and Prennushi, G. 1997, 'The Effects of Human Resource Management Practices on Productivity: A Study of Steel Finishing Lines', *The American Economic Review,* vol. 87, no. 3, pp. 291-313.

Ilmakunnas, P. Maliranta, M. & Vainiomäki, J. 2004, 'The Roles of Employer and Employee Characteristics for Plant Productivity', *Journal of Productivity Analysis,* vol. 21, no. 3, pp. 249–276.

Johnston, A. Porter, D. Cobbold, T. & Dolamore, R. 2000, 'Productivity in Australia's Wholesale and Retail Trade', *Productivity Commission Working Paper,* no. 1641, University Library of Munich, Germany.

Moretti, E. 2004, 'Workers' Education, Spillovers, and Productivity: Evidence from Plant-Level Production Functions.' *American Economic Review*, vol. 94, no. 3, pp. 656-690.

Plutarchos, S. & Daniel, J. W. 2004, 'Quantifying Embodied Technological Change', *Review of Economic Dynamics*, vol. 7, no. 1, pp. 1-26.

Syverson, C. 2011, 'What Determines Productivity?', *Journal of Economic Literature,* vol. 49, no. 2, pp. 326-365.

Van Biesebroeck, J. 2003, 'Productivity Dynamics with Technology Choice: An Application to Automobile Assembly', *The Review of Economic Studies*, vol. 70, no. 1, pp. 167-198

Johnston, J. 2003, 'Service Productivity: Towards Understanding the Relationship Between Operational and Customer Productivity', *International Journal of Productivity and Performance Management*, vol. 53, no. 3, pp. 201-213

Service Industry

Karmarkar, U. 2004, 'Will You Survive the Services Revolution?', *Harvard Business Review,* vol 82, no 6, pp. 100-107

Lee, S. M. Ribeiro, D. Ohlson, D. Roid & S. 2007, 'The importance of the activities of service business in the economy: welcome to the Service Business, an International Journal', *Service Business,* vol. 1, no. 1, pp. 1-5

Other

The World Bank, World Development Indicators (2012). Percentage of employment in service industry worldwide. Retrieved from: https://data.worldbank.org/indicator/SL.SRV.EMPL.ZS

The World Bank, World Development Indicators (2012). Service industry as a percentage of GDP worldwide. Retrieved from: https://data.worldbank.org/indicator/NV.SRV.TOTL.ZS