Evaluating the Significance of COVID-19 and Contactless Payments on Utilisation of Oyster Cards in London

Grace Anderton
38295571

January, 2025

Contents

1 Introduction **1**
- 1.1 Motivation . . . 1
- 1.2 Literature Review . . . 1

2 Data **3**
- 2.1 Preliminary Analysis . . . 4
- 2.2 Variables . . . 7

3 Methodology **9**
- 3.1 Multiple Linear Regression Model . . . 9
- 3.2 Testing Multiple Restrictions . . . 11
  - 3.2.1 Restricted Model One Test . . . 11
  - 3.2.2 Restricted Model Two Test . . . 12

4 Results **14**

5 Conclusion **19**

Appendix A **21**
- A.1 Data Inputs . . . 21
- A.2 Outliers . . . 24

Appendix B **25**
- B.1 Regression Results . . . 25

References **27**

List of Figures

2.1 Timeline of policy changes, with data inputs between 2012-2024 and the introduction of the oyster card in 2003 3
2.2 Line chart illustrating daily oyster card usage from 2012 to 2024 4
2.3 Line chart illustrating oyster card usage by day of the week 5
2.4 Bar chart illustrating monthly oyster card usage across all years 6
2.5 Histogram illustrating the distribution of daily oyster card uses 6

4.1 Line chart comparing daily oyster card uses to the fitted values from Model 3.2 17
4.2 Box plot illustrating point estimates and confidence intervals for Model 3.2 coefficients 17
4.3 Histograms illustrating the distribution of daily oyster card uses before and after the start of the first COVID-19 Lockdown 18
4.4 Histograms illustrating the distribution of daily oyster card uses before and after the introduction of contactless payments on TfL services 18

List of Tables

2.1 Summary Statistics . . . 8

3.1 Regression statistics for the Unrestricted Model and the Restricted Model One 12
3.2 Regression statistics for the Unrestricted Model and the Restricted Model Two 13

4.1 Exact percentage changes for Model 3.2's dummy explanatory variables . . . 14

A.1 Oyster card usage data . . . 21
A.2 Raw oyster card usage data with duplicates February-March 2017 . . . 22
A.3 Cleaned oyster card usage data February-March 2017 . . . 22
A.4 Raw oyster card usage data with duplicates June 2017 . . . 23
A.5 Cleaned oyster card usage data June 2017 . . . 23
A.6 Card usage on December 25th, Christmas Day, for all available years . . . 24

B.1 Regression results for Model 3.2 with Newey-West std. err. . . . 25
B.2 Regression results for Restricted Model One . . . 26
B.3 Regression results for Restricted Model Two . . . 26

Chapter 1

Introduction

1.1 Motivation

Transport for London (TfL) is the largest public transport provider in the capital city of the United Kingdom; London. Whilst working in London during 2023-2024, I heavily relied on TfL services for transport to and from work and leisure activities. Owning an oyster card was an obvious choice for me, but I noticed a large proportion of TfL users used contactless payments instead. Through searching for a way to refund my Oyster Card balance this December, I found the following dataset profoundly interesting. Whilst the number of oyster cards have been steadily increasing since their inception, the usage of oyster cards on TfL services have been declining in real terms for 10 years. In this paper, we will examine the effects of COVID-19 and contactless payments on the usage of oyster cards in London, and evaluate which had a greater impact.

1.2 Literature Review

There have been several external shocks to the TfL system causing the decline in utilisation of oyster cards. Despite their initial convenience, many customers have switched to contactless payments since 2014. I believe this is also due to mobile phone payments, such as Apple Pay and Google Wallet, being widely adopted across TfL routes. Another easily identified external shock would be the COVID-19 pandemic which caused several 'lockdown' periods between 2020-2022. Literature agrees the oyster card scheme was met with great enthusiasm upon it's launch in 2003, continuing right up until 2014 (Low, 2021). Whilst contactless payments on bus services were introduced in 2012, few consumers made the switch since the oyster fare was almost half the price (Glickenstein, 2013). With the city-wide adoption of contactless payments on TfL services completed in 2014 (TfL, 2014), the utilisation of oyster cards has been in steady decline, with contactless payments overtaking oyster cards in late 2018 (Low, 2021). Boris Johnson announced a full lockdown in the UK on the 23rd March 2020 (UK, 2020), whereby non-essential travel and business was halted and people were ordered by law to stay at home. With only essential travel allowed, the TfL saw a 95% drop in tube and rail journeys in April 2020 (TfL, 2020), leaving the TfL in a dire situation. It is with this backdrop that we begin our examination of the significance of COVID-19 and contactless payments on the utilisation of oyster cards in London.

Chapter 2

Data

Data used in this study has been sourced from the official TFL Publications and Reports page for oyster cards, under the section *Daily breakdown of Oyster card usage*. This PDF was converted into a text file using a basic Python package, then imported into Excel. A sample of March 2024 is shown in Table A.1. This dataset includes daily counts of oyster cards used, from 1st November 2012 until 31st March 2024. We will be using 16th September 2014 as the policy date for contactless payments as stated by (TfL, 2014). We will be using 23rd March 2020 as the policy date for COVID-19 as stated by (UK, 2020). A timeline is included below in Figure 2.1 for the readers' reference.

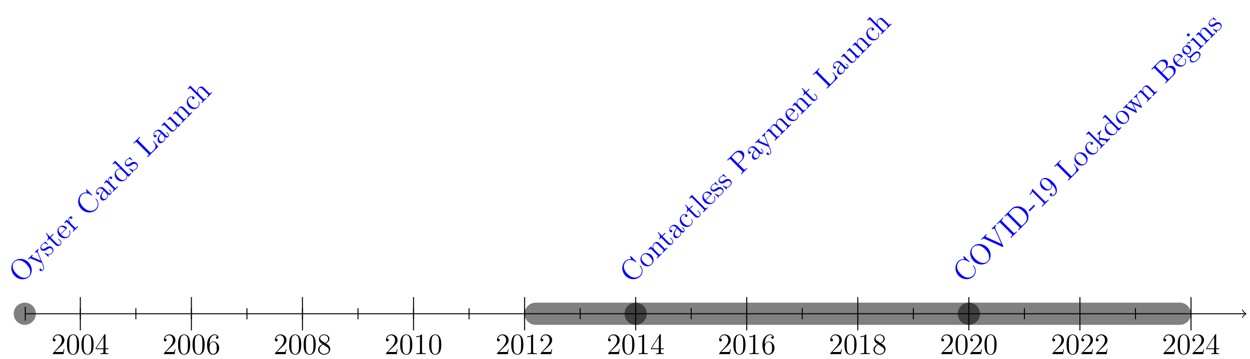
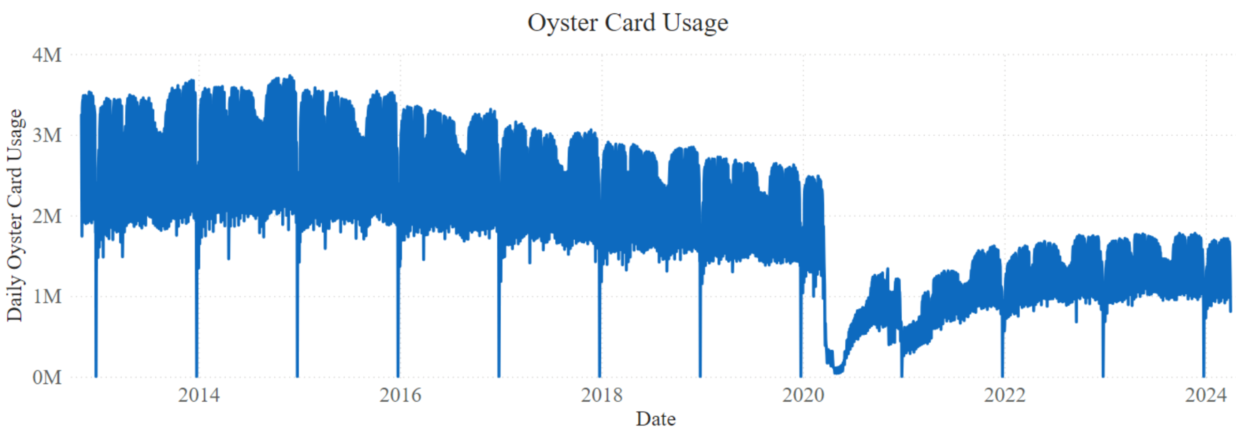


Figure 2.1: Timeline of policy changes, with data inputs between 2012-2024 and the introduction of the oyster card in 2003

Through some preliminary data cleaning, we found duplicated data entries for multiple dates from February to March 2017 and also in June 2017. A record of these duplicates is in Tables A.2 and A.4 alongside the cleaned versions in Tables A.3 and A.5. To clean, we used majority voting to predict the card usage for each unique date. After this cleaning, we progressed onto preliminary analysis of the dataset.

2.1 Preliminary Analysis

All figures in this section were produced in Power BI using the cleaned dataset. In Figure 2.2, one trend we observe is that daily oyster card usage has been steadily declining since the peak in 2015 of almost 4 million uses per day. Another trend is the decline in daily uses since the huge shock in 2020. Despite daily uses slowly recovering since the shock, daily uses have not returned to pre-shock levels. Christmas Day, 25th December, creates a strong trough for daily uses every year, with significantly lower uses than any other surrounding days. To add to that, August reports lower daily uses than the surrounding months each year. To look deeper into these trends, we produced the following figures.



Produced using Power BI

Figure 2.2: Line chart illustrating daily oyster card usage from 2012 to 2024

Figure 2.3 highlights the day-specific trends in oyster card usage. We can see the deep trough on 25th December each year more clearly, which is supported by the average usage on 25th December being 728. Figures for all available 25th Decembers are included in A.6 for the reader's reference. Our most interesting observation is the effect that the day-of-the-week has on usage. Tuesday through to Friday are the most consistent days, with low volatility. Monday is more volatile in comparison, which we hypothesise to be due to Bank Holidays in the UK. By comparing with the UK Bank Holidays dataset from The United Kingdom Debt Management Office we conducted a hypothesis test as follows;

$$H_0 : \mu_m = \mu_h, H_1 : \mu_m \neq \mu_h \tag{2.1}$$

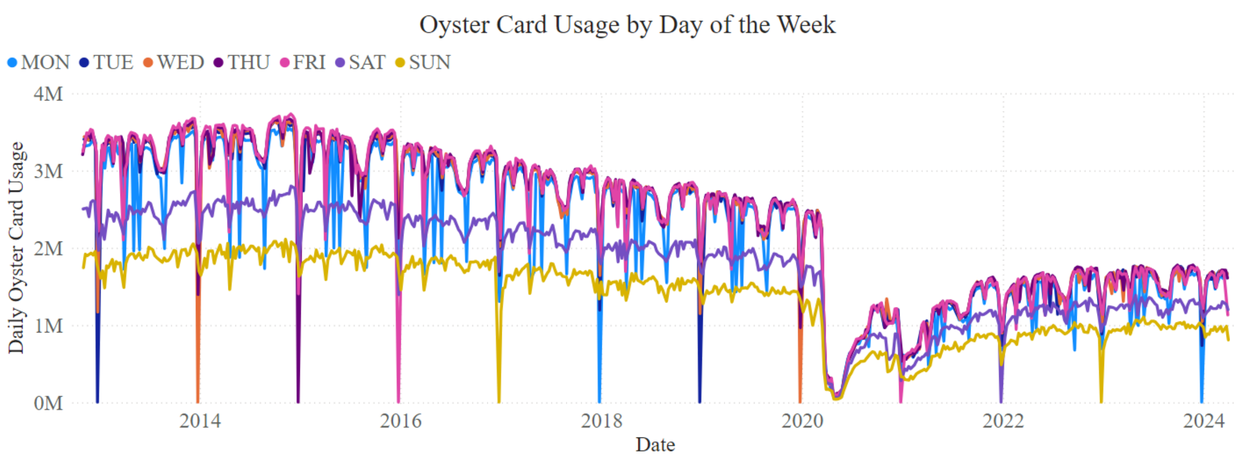
Where μ_m represents the average usage on Mondays that are not bank holidays, and μ_h represents the average usage on Mondays that are bank holidays. Using Excel, we found $\mu_m = 2365251$ and $\mu_h = 1283569$, as well as the sample standard deviations $s_m = 885545.1635$

and $s_h = 566716.5$. The sample sizes are $n_m = 540$ and $n_h = 55$. Assuming both are normally distributed and the samples are independent, we can test the hypothesis by calculating a t-score:

$$t \equiv \frac{\mu_m - \mu_h}{\sqrt{\frac{s_m^2}{n_m} + \frac{s_h^2}{n_h}}} = \frac{2365251 - 1283569}{\sqrt{\frac{885545.1635^2}{540} + \frac{566716.5^2}{55}}} = 12.67 \tag{2.2}$$

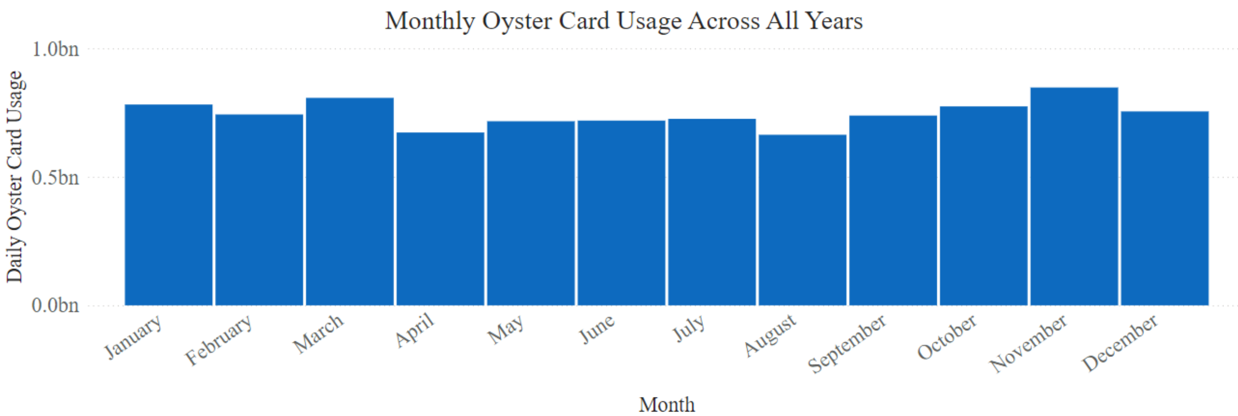
For a two tailed test at the 5% significance level with 593 degrees of freedom, we find the critical value to be approximately 1.96. Given that $12.67 > 1.96$, we can say there is sufficient evidence to reject H_0 in favour of H_1. We conclude that there is most likely a significant difference between average usage on Bank holiday Mondays and normal Mondays, supporting our hypothesis that usage on Bank holidays explain the high volatility in Monday data. Bank holidays tend to bring a localised dip in the usage over the surrounding weekdays, which can be explained by people taking leave around the bank holiday to maximise time off.

We can also observe that the usage line for weekends are much lower than weekdays. Saturday tends to be almost a million lower than weekdays, with Sunday tending to be almost two million lower than weekdays. This trend remains post-COVID, but with gaps of 500,000 and one million respectively. This implies that daily trends are weakened post-COVID, which could be explained by more Work From Home (WFH) schemes available for jobs in the city.



Produced using Power BI

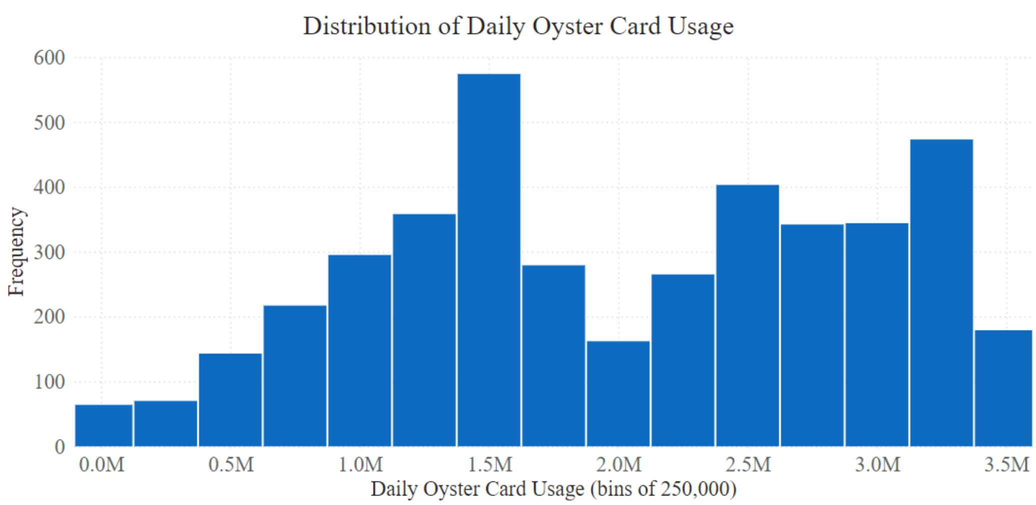
Figure 2.3: Line chart illustrating oyster card usage by day of the week



Produced using Power BI

Figure 2.4: Bar chart illustrating monthly oyster card usage across all years

Figure 2.4 highlights the monthly differences in oyster card usage, with each month's bar being the aggregate uses over all years and days in this month. There is not a large amount of volatility between months, but some months are particularly low. This supports our earlier trend prediction of August being a low usage month, but also brings April into consideration, as they have the lowest total use of $663,826,199$ and $672,636,220$ respectively. These results could be explained by school holiday periods; with August being the Summer Holiday and April being the Easter Holiday. It is clear that there is some seasonality in the data and we recommend monthly dummy explanatory variables.



Produced using Power BI

Figure 2.5: Histogram illustrating the distribution of daily oyster card uses

When plotting a histogram of daily oyster card uses, as shown in Figure 2.5, no clear distribution arises. It appears that there are two groups in the data, with a peak on the left and a peak on the right. This could be explained by the weekend and weekday trends, alongside bank holidays and Christmas Day. It could also be explained by pre and post a policy; this relationship will be discussed further in Chapter 4.

2.2 Variables

Daily Oyster Card Use (dailyoysteruse): We introduce the outcome (dependent) variable for this model as daily use of oyster cards. We are interested in the patterns of oyster card usage in London, so this is what we will be modelling with the following variables.

Contactless Payments Introduction (contpayintro): We introduce a policy dummy explanatory variable as a post contactless payments indicator. This will be 1 for days that are on or after the introduction of contactless payments, and 0 for before.

COVID-19 Lockdown (covidlock): We introduce a policy dummy explanatory variable as a post-COVID-19 indicator. This will be 1 for days that are on or after the start of the first COVID-19 lockdown, and 0 for before.

Christmas Day (xmasday): We introduce a dummy explanatory variable that will help control for Christmas Day drops in usage. This will be 1 for when it is 25th December, and 0 on all other days.

Bank Holidays (bankholiday): We introduce a dummy explanatory variable that will help control the decline in usage around bank holidays each year. This will be 1 for a day that is included in the bank holiday dataset, and 0 if not. Despite overlap between *xmasday* and *bankholiday* on 25th December each year, but we believe that since that day is such an outlier it is fair to assume independence between these variables.

Weekend Use (weekend): We introduce a dummy explanatory variable weekends to account for differences between weekdays and weekends. The trend of Saturday usage is consistently lower then weekdays, with Sundays being even lower. Using *weekend* it will be 1 on a Saturday or Sunday, indicated by the oyster card usage dataset, and 0 when the day is not Saturday or Sunday.

Monthly Use (monthly): We introduce a dummy explanatory variable for monthly usage. We created a *monthly* dataset by assigning a unique integer to each month.

Time Trends (time): We introduce one more dummy explanatory variable for time trends. We created a *time* dataset by assigning a unique integer to day.

Variable	**Obs**	**Mean**	**Std. dev.**	**Min**	**Max**
dailyoysteruse	4,168	2,145,992	924,446.8	183	3,733,516
log(dailyoysteruse)	4,168	14.42737	0.7318505	5.209486	15.13286
contpayintro	4,168	0.8358925	0.3704175	0	1
covidlock	4,168	0.3526871	0.4778637	0	1
xmasday	4,168	0.0028791	0.0535862	0	1
bankholiday	4,168	0.0220729	0.1469384	0	1
weekend	4,168	0.2857486	0.4518244	0	1
monthly	4,168	6.496881	3.505351	1	12
time	4,168	2084.5	1203.342	1	4168

Table 2.1: Summary Statistics

Chapter 3

Methodology

In this study we are examining the magnitude of effect that COVID-19 and contactless payments had on oyster card utilisation. We will carry out a multiple regression with the variables defined in Chapter 2.2. The use of Difference in Differences (DiD) was considered, but ruled out due to collinearity since we don't have a dataset where we can separate the treatment group from the control group (Schwerdt and Woessmann, 2020). This means the joint impact can't be found between the oyster card usage and *contpayintro* or *covidlock*. Another way we can examine the magnitudes of effect could be through restricted models. If we conduct hypothesis tests on the restricted model vs the unrestricted model, we will be able to conclude on which policy had a greater effect. This will be the second half of our methodology and will consist of two joint hypothesis tests. One hypothesis result we predict is that the start of the first COVID-19 lockdown will have a more significant impact on daily oyster card uses than the introduction of contactless payments, based off the large shock that immediately plummeted daily uses illustrated in Figure 2.2. Additionally, the second hypothesis result we predict is that the introduction of contactless payments and the start of the first COVID-19 lockdown *both* have a significant effect on daily oyster card uses, and that their presence in the model is beneficial.

3.1 Multiple Linear Regression Model

Multiple Linear Regression (MLR) is generally used when there are multiple factors affecting the outcome variable. In this case, we have that daily oyster card usage was affected by COVID-19 and contactless payments, as well as several time dummy variables. Since we are using daily data over 12 years, it may be affected by autocorrelation, seasonality or time trends. Adding time-series methods into our MLR model should reduce their effects. We add in monthly and weekday dummy explanatory variables to account for seasonality, $monthly_t$ and $weekend_t$ respectively. Seasonality has been identified in Chapter 2.1 so adding these in will account for repeating patterns. We also add in a time dummy explanatory variable, $time_t$, to capture long term trends in the time series data. We considered adding lagged variables into the model, but since the R^2 value was already high we didn't want to overcomplicate, so we will use the Newey-West standard error to correct autocorrelation. We propose the use of *log(dailyoysteruse)* as a log-linearised dependent variable. This enables us to compute percentage change coefficients, making model evaluation easier. To add to that, since there are some outlier values, such as Christmas Day, using a log-linearised dependent variable helps to stabilise the regression, as well as reducing potential heteroscedasticity (despite heteroscedasticity being ruled out through our preliminary analysis). Since there are also no zero values in *dailyoysteruse*, we propose the following model 3.1;

$$\begin{aligned} log(dailyoysteruse_t) &= \beta_0 + \beta_1 contpayintro_t + \beta_2 covidlock_t + \beta_3 xmasday_t \\ &+ \beta_4 bankholiday_t + \beta_5 weekend_t + \beta_6 monthly_t \\ &+ \beta_7 time_t + \epsilon_t \end{aligned} \tag{3.1}$$

We can't say it is an independent and random sample since daily oyster card uses are a time series so past results tend to influence future results, and they are not random since events such as COVID-19 may influence results. As we are working with a large sample size of 4,168 observations, we can relax the assumptions on our MLR model. We can require OLS to be consistent not unbiased, making exogeneity less strict. We can use the central limit theorem to relax normality, in our case this is not needed but we will apply it to meet asymptotic normality requirements and use a Newey-West standard error. Despite assuming the above, there is high likelihood of omitted variable bias where there are other variables impacting daily oyster card usage, such as ULEZ expansion and fare changes. This makes it less likely that the model is consistent. Regardless, we will continue and discuss the effects in our results.

Since we are employing the Newey-West standard error, we will use the `newey` function in Stata with `lag(8)`, calculated by $4168^{0.25}$ (Greene, 2003). Running `newey` on Model 3.1 in Stata gives the results in Table B.1 for the reader's reference. These results give the following model:

$$\begin{aligned} log(dailyoysteruse_t) &= 14.91 - 0.278 contpayintro_t - 1.040 covidlock_t - 7.65 xmasday_t \\ &- 0.620 bankholiday_t - 0.427 weekend_t + 0.0107 monthly_t \\ &+ 0.0000973 time_t + \epsilon_t \end{aligned} \tag{3.2}$$

3.2 Testing Multiple Restrictions

Testing multiple restrictions on a MLR model can help us identify if a simpler model might fit the data better. A restricted model is where we set several coefficients to be equal to zero, e.g. $\beta_3 = 0, \beta_4 = 0, \beta_5 = 0$. By comparing fits of both models simultaneously we eliminate Type I errors in our test and a false conclusion. Using joint hypothesis tests, we can test whether COVID-19 had a more significant impact on daily oyster card utilisation than the introduction of contactless payments. We can also test if both policies had no effect on daily oyster card usage. Continuing with these joint tests, we propose the unrestricted model and the following restricted models in the following sections, in addition to conducting the appropriate hypothesis tests.

3.2.1 Restricted Model One Test

Unrestricted Model:

$$\begin{aligned} log(dailyoysteruse_t) = {} & \beta_0 + \beta_1 contpayintro_t + \beta_2 covidlock_t + \beta_3 xmasday_t \\ & + \beta_4 bankholiday_t + \beta_5 weekend_t + \beta_6 monthly_t \\ & + \beta_7 time_t + \epsilon_t \end{aligned} \tag{3.1}$$

Restricted Model One:

$$\begin{aligned} log(dailyoysteruse_t) = {} & \beta_0 + \gamma(contpayintro_t + covidlock_t) + \beta_3 xmasday_t \\ & + \beta_4 bankholiday_t + \beta_5 weekend_t + \beta_6 monthly_t \\ & + \beta_7 time_t + \epsilon_t \end{aligned} \tag{3.3}$$

where $\gamma = \beta_1 = \beta_2$. Our proposed hypothesis based on Restricted Model One 3.3 is:

$$H_0 : \beta_1 = \beta_2, H_1 : \beta_1 \neq \beta_2 \tag{3.4}$$

where we are testing whether COVID-19 lockdown or the introduction of contactless pay had a statistically significant difference on daily oyster card uses. We will generate a new variable labelled *contpayintropluscovidlock* which is equal to *contpayinto* + *covidlock*, allowing for γ to be calculated in Stata. We assume that there is no omitted variable bias affecting the coefficients, as well as the proper measurement of the policy dummy explanatory variables. From using `regress` on Models 3.1 and 3.3, we get the statistics in Table 3.1. Full regression results are shown in Appendix B.1 and B.2.

Statistic	Unrestricted Model	Restricted Model One
SSR	479.639513	615.231523
R-squared	0.7851	0.7243

Table 3.1: Regression statistics for the Unrestricted Model and the Restricted Model One

With $(4168 - 7 - 1)$ degrees of freedom for the denominator and 1 restriction, we can calculate a normal F test statistic due to assumed normality of the error term with the large sample size.

$$F \equiv \frac{SSR_{r1} - SSR_{ur}/q}{SSR_{ur}/(N-k-1)} = \frac{615.23 - 479.64/1}{479.64/(4168-7-1)} = 1175.995 \tag{3.5}$$

Choosing a 5% significance level, we find the two-tailed critical value for $F \sim F_{0.95,(1,4160)} = 3.844$. Given that $1175.995 > 3.844$, we can say there is sufficient evidence to reject H_0 in favour of accepting H_1. We conclude that the differences between the impact of the start of the first COVID-19 lockdown and the introduction of contactless payments are statistically significant. Using the coefficients from Model 3.2 we determined earlier, we know $\beta_1 = -0.278$ and $\beta_2 = -1.040$. So we can say $|\beta_1| < |\beta_2|$, supporting the argument that the COVID-19 lockdown had a statistically significantly greater impact on daily oyster card utilisation than the introduction of contactless payments.

3.2.2 Restricted Model Two Test

Unrestricted Model:

$$\begin{aligned} log(dailyoysteruse_t) = \beta_0 &+ \beta_1 contpayintro_t + \beta_2 covidlock_t + \beta_3 xmasday_t \\ &+ \beta_4 bankholiday_t + \beta_5 weekend_t + \beta_6 monthly_t \\ &+ \beta_7 time_t + \epsilon_t \end{aligned} \tag{3.1}$$

Restricted Model Two:

$$\begin{aligned} log(dailyoysteruse_t) = \beta_0 &+ \beta_3 xmasday_t + \beta_4 bankholiday_t + \beta_5 weekend_t \\ &+ \beta_6 monthly_t + \beta_7 time_t + \epsilon_t \end{aligned} \tag{3.6}$$

where $\beta_1 = 0$ and $\beta_2 = 0$. Our proposed hypothesis based on Restricted Model Two 3.6 is:

$$H_0 : \beta_1 = 0 \text{ and } \beta_2 = 0, H_1 : \beta_1 \neq 0 \text{ and/or } \beta_2 \neq 0 \tag{3.7}$$

where we are testing whether at least one of COVID-19 lockdown or the introduction of contactless payments had a significant impact on daily oyster card utilisation. We assume that there is no omitted variable bias affecting the coefficients, as well as the proper measurement of the policy dummy explanatory variables. We will calculate a F-statistic to test the hypothesis. From using `regress` on Models 3.6 and 3.1, we get the statistics in Table 3.2. Full regression results are shown in Appendices B.1 and B.3.

Statistic	Unrestricted Model	Restricted Model Two
SSR	479.639513	743.036449
R-squared	0.7851	0.6671

Table 3.2: Regression statistics for the Unrestricted Model and the Restricted Model Two

With (4160) degrees of freedom for the denominator and 2 restrictions, we can calculate a normal F test statistic due to assumed normality of the error term with the large sample size.

$$F \equiv \frac{SSR_{r2} - SSR_{ur}/q}{SSR_{ur}/(N-k-1)} = \frac{743.04 - 479.64/2}{479.64/(4168-7-1)} = 1142.26 \tag{3.8}$$

Choosing a 5% significance level, we find the two-tailed critical value for $F \sim F_{0.95,(2,4160)} = 2.998$. Given that $1142.26 > 2.998$, we can say there is sufficient evidence to reject H_0 in favour of accepting H_1. We conclude that there is strong evidence at the 5% significance level of at least one of the policy dummy explanatory variables being highly impactful towards daily oyster card utilisation. This supports the argument that most of the model can be explained by at least one policy dummy explanatory variable. Time series adjustments cannot fit the data with high accuracy without the addition of policy dummy explanatory variables in this model.

Chapter 4

Results

To properly interpret the coefficients of dummy explanatory variables when the dependent variable is log-linearised, we must calculate the exact percentage differences by exponentiating and subtracting one (Wooldridge, 2020 - 2020). Using the *covidlock* variable as an example, we have $dailyoysteruse_1$ when $covidlock_t = 1$ and $dailyoysteruse_0$ when $covidlock_t = 0$, and keeping all other explanatory variables zero:

$$\frac{dailyoysteruse_1 - dailyoysteruse_0}{dailyoysteruse_0} = exp(-1.040) - 1 \approx -0.646545 \tag{4.1}$$

Implementing this for each of Model 3.2's coefficients leaves us with the exact percentage changes, which are shown in Table 4.1, and make much more sense than the original coefficients.

Dummy Explanatory Variable	**Model 3.2 Coefficient**	**Exact Percentage Change**
contpayintro	-0.278	-24.27%
covidlock	-1.040	-64.65%
xmasday	-7.65	-99.95%
bankholiday	-0.0620	-6.01%
weekend	-0.427	-34.75%
monthly	0.0107	1.07%
time	0.0000973	0.00973%

Table 4.1: Exact percentage changes for Model 3.2's dummy explanatory variables

From our results in Chapter 3.1 of Model 3.2 and the exact percentages calculated in Table 4.1, we can make several observations. Firstly, the introduction of contactless pay had a negative effect on *log(dailyoysteruse)* with a -24.27% drop in *log(dailyoysteruse)* after the introduction of contactless payments, holding others constant. Whereas, *covidlock* had a relatively larger effect of -64.65%, echoing the conclusions made in Chapter 3.2.1. Having an intercept of 14.91% is expected as this is the level of *log(dailyoysteruse)* when all dummy explanatory variables are zero, showing the initial level of daily use to be circa 3,001,465. The variable *monthly* had a slightly positive impact of 1.07% on daily oyster card uses, with months later in the year contributing more. The *time* dummy had a positive but small impact of less than 0.01%, which contrasts with our view of terminally falling usage of oyster cards. Weekends and bank holidays had negative effects of -34.75% and -6.01% respectively, supporting the predicted lower rider-ship on public holidays and weekends. Christmas Day had an impact of -99.95% on daily oyster card uses, which is very large. This illustrates the high importance of the religious holiday to the population of London.

With a R^2 value of 0.7851, we consider our model well fitted to the time-series dataset. As illustrated in Figure 4.1, the fitted values tend to share the same trends as the real data, even down to the Christmas Day outliers. Limitations of the model were found around the external shocks of the policies. After the introduction of contactless payments, the model saw a larger than true drop in daily uses, and then proceeded to slowly increase daily uses up until the start of the first COVID-19 lockdown. This does not entirely reflect the sentiments of the true data. To add to that, the severity of the COVID-19 pandemic was not captured fully by the model, and sees a drop to $\sim$ 13.5 log(dailyoysteruse), which is $\sim$ 25% higher than the true drop in the data, which perhaps non-linear models like Probit/Logit could correctly model. Despite this, our model fits well. When examining the Confidence Intervals (CI) for each coefficient, we find that the CI for the time dummy explanatory variables, *weekend*, *monthly* and *time*, are so small that their upper and lower boundaries do not show. This implies that their coefficients are highly accurate for this model. To add to that, we find that *xmasday* has the widest CI, which could be explained by a small sample of 12 days.

The reader may recall from Chapter 2.1 we discussed the distribution of daily oyster card uses. We found a distribution with two groups and we suggested weekends and bank holidays as an explanation for the abnormal shape of the histogram in Figure 2.5. However, the conclusion of our hypothesis test in Chapter 3.2.1 forced us to reconsider our original suggestion. To test an idea, we plotted a histogram filtered by dates before 23rd March 2020, and a histogram filtered by dates on or after 23rd March 2020, creating a pre and post COVID-19 analysis of the distribution of daily oyster card uses. The resulting Figure 4.3 illustrates our idea perfectly. The two groups could more likely be pre and post the start of the first COVID-19 lockdown, with each group exhibiting almost normal distribution. Whilst is notable that both groups have a right skew, both groups display a similar skew, making it a yet more likely explanation. Observing that post 23rd March 2020 the average daily use was 1.17 million, which is just 43% of pre-pandemic levels, leads us to believe that the impact of COVID-19 is persistent and has been detrimental to the utilisation of oyster cards in London.

Conducting the same review on contactless payments does not yield the same result. As shown in Figure 4.4, we found that the histograms filtered for before and after the introduction of contactless payments both displayed the same abnormal distribution as Figure 2.5. This further supports the argument that COVID-19 had a much more significant impact on the utilisation of oyster cards. Despite this, post 19th September 2014 the average daily use was 1.99 million, which is 67% of the usage before the introduction of contactless payments, implying that contactless payments did have *some*, if little, significance.

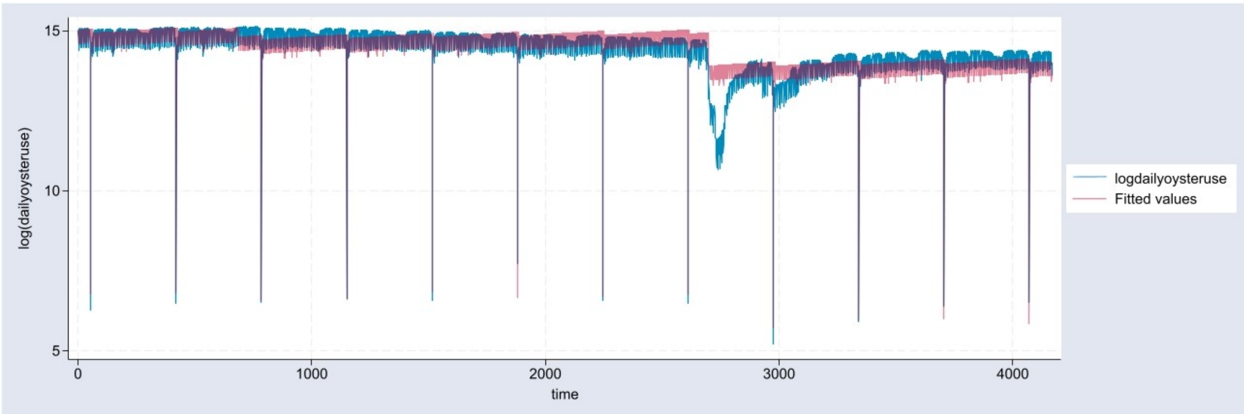


Figure 4.1: Line chart comparing daily oyster card uses to the fitted values from Model 3.2

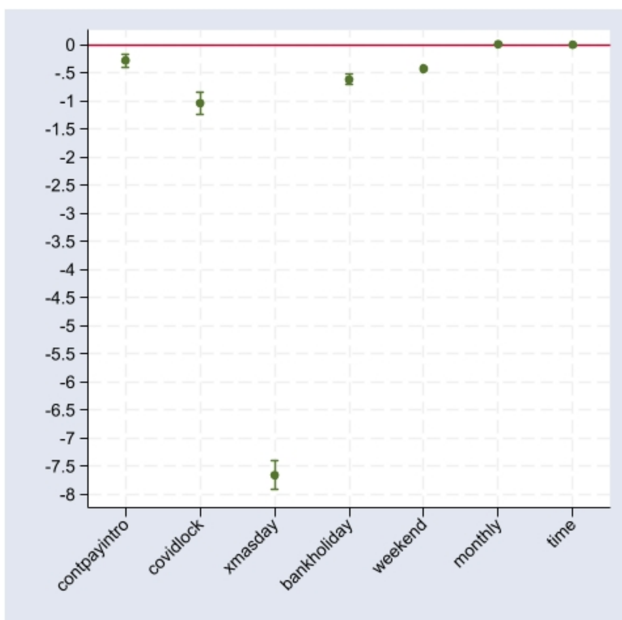
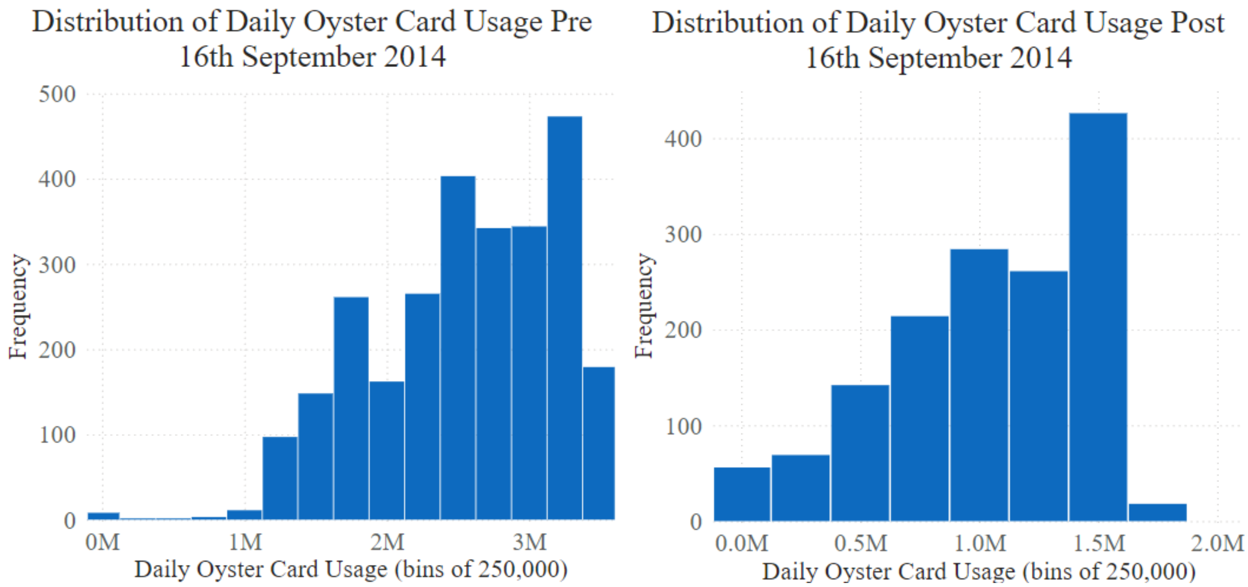
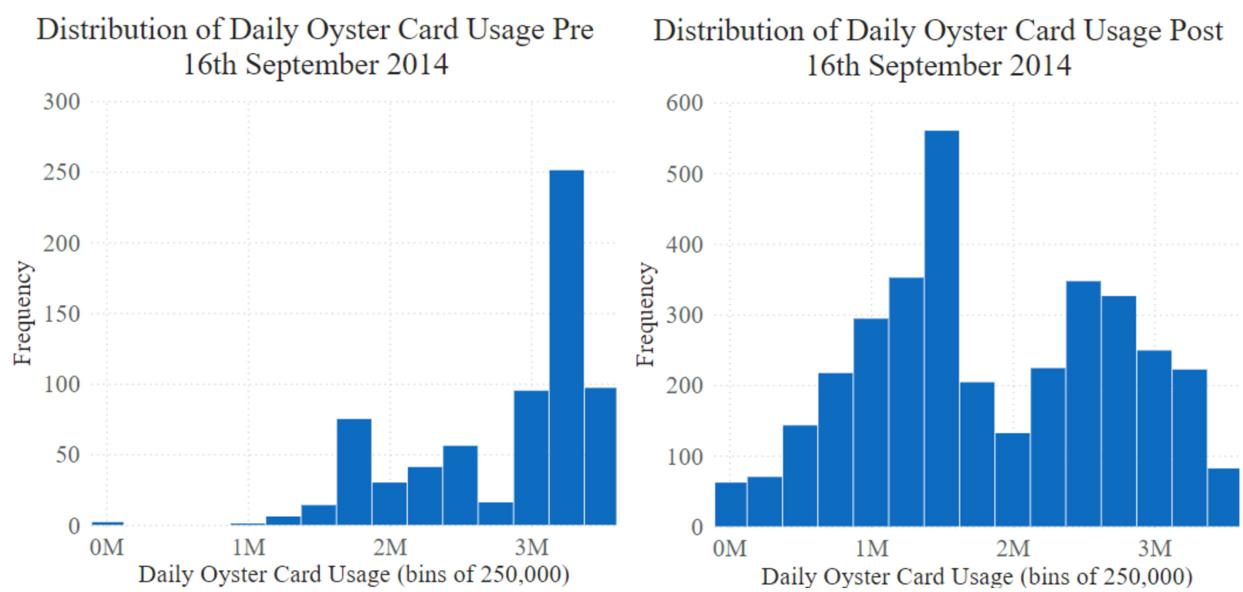


Figure 4.2: Box plot illustrating point estimates and confidence intervals for Model 3.2 coefficients



Produced using Power BI

Figure 4.3: Histograms illustrating the distribution of daily oyster card uses before and after the start of the first COVID-19 Lockdown



Produced using Power BI

Figure 4.4: Histograms illustrating the distribution of daily oyster card uses before and after the introduction of contactless payments on TfL services

Chapter 5

Conclusion

Overall, we conclude that COVID-19 had a more significant effect on the utilisation of oyster cards in London. The evidence from the joint hypothesis tests as well as analysis on the distribution of daily oyster card utilisation led us to this conclusion. We also conclude that the introduction of contactless payments had a smaller but still significant effect on the utilisation of oyster cards in London.

The introduction of contactless payments reduced oyster card use by 24.27%, which is explained by TfL customers switching to the more convenient method of payment. The drop in oyster card use of 64.65% due to the start of the first COVID-19 lockdown is significant, and highlights the severe impact the pandemic had on TfL services. Christmas Day has an even larger effect on daily uses, with a 99.95% drop each year. This can be explained by the closure of most TfL services on the Christian holiday, but also by less people commuting or travelling on the day. Months have a positive effect on daily oyster card use, with months later in the year having a larger effect. This could be due to more travellers during the festive season, as well as bad weather meaning less people choose to walk/cycle/run. More interestingly, time also exhibits a positive upward trend. Although minimal, it implies that our model estimates an underlying upward trend in the use of oyster cards. This contradicts most literature on the subject, but is an interesting interpretation. Finally, both bank holidays and weekends have a negative impact on rider-ship in our model, supporting our prediction that less people travel around the weekends and bank holidays due to holidays and maximising leave.

We believe that Model 3.2 provides a good fit for the chosen dataset. Employing the Newey-West standard error barely changed any coefficient values, but did change the standard errors and t-statistics for most variables. We concur that it was a good choice of a standard-error estimator for this model. Using a log linearised dependent variable was good for analysis, allowing us to compute exact percentage changes in Table 4.1 rather than large integers. Conducting the joint hypothesis tests let us reach conclusions with a high level of confidence. Improvements for future study could involve the collection of panel data to conduct a DiD analysis for policy changes around TfL services. Policy changes are not limited to the introduction of contactless payments or the start of the first COVID-19 lockdown, and may involve fare changes, ULEZ boundary expansions, the completion of the Elizabeth Line, and more. Future investigations could consider more explanatory variables and their impact. Forecasting the trajectory of oyster card use would be an interesting extension and may aid TfL on making a decision to retire the scheme, as well as evaluating how daily trends are weakened post-COVID and the factors affecting those.

Appendix A

A.1 Data Inputs

Date	Day	Card Usage
31/03/2024	SUN	808548
30/03/2024	SAT	1200366
29/03/2024	FRI	1131592
28/03/2024	THU	1610124
27/03/2024	WED	1661576
26/03/2024	TUE	1675084
25/03/2024	MON	1611321
24/03/2024	SUN	986345
23/03/2024	SAT	1270691
22/03/2024	FRI	1662611
21/03/2024	THU	1710846
20/03/2024	WED	1698095
19/03/2024	TUE	1690702
18/03/2024	MON	1630169
17/03/2024	SUN	956155

(a)

Date	Day	Card Usage
16/03/2024	SAT	1266858
15/03/2024	FRI	1671013
14/03/2024	THU	1711807
13/03/2024	WED	1694871
12/03/2024	TUE	1633947
11/03/2024	MON	1607331
10/03/2024	SUN	915767
09/03/2024	SAT	1303806
08/03/2024	FRI	1698011
07/03/2024	THU	1713180
06/03/2024	WED	1706709
05/03/2024	TUE	1694150
04/03/2024	MON	1637163
03/03/2024	SUN	983531
02/03/2024	SAT	1226436
01/03/2024	FRI	1649971

(b)

Table A.1: Oyster card usage data

Date	Day	Card Usage
04/03/2017	THU	2281790
04/03/2017	FRI	2282044
04/03/2017	SAT	2281790
03/03/2017	MON	3093224
03/03/2017	TUE	3093437
03/03/2017	WED	3093224
02/03/2017	FRI	3109302
02/03/2017	SAT	3109468
02/03/2017	SUN	3109302
01/03/2017	TUE	3088772
01/03/2017	WED	3088952
01/03/2017	THU	3088772
28/02/2017	SAT	3100720
28/02/2017	SUN	3100864
28/02/2017	MON	3100720
27/02/2017	WED	2980896
27/02/2017	THU	2981035
27/02/2017	FRI	2980896
26/02/2017	SUN	1725713
26/02/2017	MON	1725799
26/02/2017	TUE	1725713

Table A.2: Raw oyster card usage data with duplicates February-March 2017

Date	Day	Card Usage
04/03/2017	SAT	2281790
03/03/2017	FRI	3093224
02/03/2017	THU	3109302
01/03/2017	WED	3088772
28/02/2017	TUE	3100720
27/02/2017	MON	2980896
26/02/2017	SUN	1725713

Table A.3: Cleaned oyster card usage data February-March 2017

Date	Day	Card Usage
30/06/2017	FRI	3007078
30/06/2017	FRI	3007078
29/06/2017	THU	3013748
29/06/2017	THU	3013748
28/06/2017	WED	2965700
27/06/2017	TUE	2906163
27/06/2017	TUE	2906163
26/06/2017	MON	2825783
26/06/2017	MON	2825783
25/06/2017	SUN	1704745
25/06/2017	SUN	1704745
24/06/2017	SAT	2168143
24/06/2017	SAT	2168143

Table A.4: Raw oyster card usage data with duplicates June 2017

Date	Day	Card Usage
30/06/2017	FRI	3007078
29/06/2017	THU	3013748
28/06/2017	WED	2965700
27/06/2017	TUE	2906163
26/06/2017	MON	2825783
25/06/2017	SUN	1704745
24/06/2017	SAT	2168143

Table A.5: Cleaned oyster card usage data June 2017

A.2 Outliers

Date	Day	Card Usage
25/12/2012	TUE	526
25/12/2013	WED	660
25/12/2014	THU	668
25/12/2015	FRI	750
25/12/2016	SUN	714
25/12/2017	MON	2241
25/12/2018	TUE	718
25/12/2019	WED	644
25/12/2020	FRI	183
25/12/2021	SAT	368
25/12/2022	SUN	603
25/12/2023	MON	666

Table A.6: Card usage on December 25th, Christmas Day, for all available years

Appendix B

B.1 Regression Results

Table B.1: Regression results for Model 3.2 with Newey-West std. err.

(a)

Number of Observations	**4,168**
F(7, 4160)	**1152.09**
Prob > F	**0.0000**

(b)

log(dailyoysteruse)	**Coefficient**	**Newey-West Std. Err.**	**t**	**P> |t|**	**95% Conf. Interval**
contpayintro	-0.278	0.0587	-4.73	0.000	[-0.393, -0.162]
covidlock	-1.040	0.105	-9.91	0.000	[-1.246, -0.834]
xmasday	-7.655	0.131	-58.61	0.000	[-7.911, -7.399]
bankholiday	-0.620	0.0482	-12.88	0.000	[-0.715, -0.526]
weekend	-0.427	0.00743	-57.48	0.000	[-0.442, -0.413]
monthly	0.0107	0.00351	3.05	0.002	[0.00384, 0.0176]
time	0.0000973	0.0000421	2.31	0.021	[0.0000149, 0.0001798]
constant	14.91	0.0323	461.02	0.000	[14.848, 14.975]

Figures rounded to 3.sf if below 1, otherwise rounded to 2.dp.

Table B.2: Regression results for Restricted Model One

(a)

Number of Observations	**4,168**
F(6, 4161)	**1039.09**
Prob > F	**0.0000**

(b)

		Newey-West			**95%**
log(dailyoysteruse)	**Coefficient**	**Std. Err.**	**t**	**P> |t|**	**Conf. Interval**
contpayintropluscovidlock	-0.617	0.0876	-7.05	0.000	[-0.789 , -.446]
xmasday	-7.65	0.149	-51.32	0.000	[-7.940, -7.355]
bankholiday	-0.628	0.0545	-11.53	0.000	[-0.735, -0.522]
weekend	-0.427	0.00816	-52.36	0.000	[-0.443, -.411]
monthly	0.0115	0.00412	2.79	0.005	[0.00340, 0.0196]
time	0.0000256	0.0000423	0.60	0.545	[-0.0000573, 0.000109]
constant	15.19	0.0339	448.73	0.000	[15.13, 15.26]

Figures rounded to 3.sf if below 1, otherwise rounded to 2.dp.

Table B.3: Regression results for Restricted Model Two

(a)

Number of Observations	**4,168**
F(5, 4162)	**1050.46**
Prob > F	**0.0000**

(b)

		Newey-West			**95%**
log(dailyoysteruse)	**Coefficient**	**Std. Err.**	**t**	**P> |t|**	**Conf. Interval**
xmasday	-7.63	0.161	-47.38	0.000	[-7.943, -7.312]
bankholiday	-0.634	0.0601	-10.55	0.000	[-0.752, -0.516]
weekend	-0.427	0.00886	-48.20	0.000	[-0.444, -0.409]
monthly	0.00982	0.00434	2.26	0.024	[0.00130, 0.0183]
time	-0.000299	0.0000113	-26.42	0.000	[-0.000322, -0.000277]
constant	15.15	0.0314	482.66	0.000	[15.08, 15.21]

Figures rounded to 3.sf if below 1, otherwise rounded to 2.dp.

References

Glickenstein, Harvey (2013). "Contactless payment debuts on London's buses, new and improved railways around the world [Transportation Systems]". In: *IEEE Vehicular Technology Magazine* 8.2, pp. 19–25. DOI: 10.1109/MVT.2013.2252268.

Greene, W.H. (2003). *Econometric Analysis*. Prentice Hall. ISBN: 9780130661890. URL: https://books.google.co.uk/books?id=JJkWAQAAMAAJ.

Low, Harry (Sept. 2021). "London's Oyster card: Are its days numbered?" In: URL: https://www.bbc.co.uk/news/uk-england-london-58197631.

Schwerdt, Guido and Ludger Woessmann (2020). "Chapter 1 - Empirical methods in the economics of education". In: *The Economics of Education (Second Edition)*. Ed. by Steve Bradley and Colin Green. Second Edition. Academic Press, pp. 3–20. ISBN: 978-0-12-815391-8. URL: https://www.sciencedirect.com/science/article/pii/B978012815391800001X.

TfL (2014). "Contactless payments set to launch". In: *tfl.gov.uk*. URL: https://tfl.gov.uk/info-for/media/press-releases/2014/july/contactless-payments-set-to-launch.

— (2020). "TfL Commissioner's report". In: *Commissioner's Reports, TfL Publications and Reports*. URL: https://tfl.gov.uk/corporate/publications-and-reports/commissioners-report.

UK, GOV (2020). "Staying at home and away from others (social distancing)". In: *Gov. uk*. URL: https://www.gov.uk/government/publications/full-guidance-on-staying-at-home-and-away-from-others.

Wooldridge, Jeffrey M. (2020 - 2020). *Introductory Econometrics: A Modern Approach*. Seventh edition. Australia: Cengage. ISBN: 9781337671330.