Social Norms as a Cost-Effective Measure of Managing Transport Demand:

Evidence from an Experiment on the London Underground*

Kingsley Offiaeli[1], Firat Yaman[2]

Abstract

Demand for public transport in cities has been and is projected to increase, putting existing transport networks under increasing strain. Nudging passengers to behave in certain ways through the creation of a salient social norm has the potential to be a cost-effective mechanism to manage transport demand. Transport for London (TfL) implemented in the second half of 2017 an experiment on one of its busiest metro train platforms. The platform surface was painted to highlight the exact location of the train doors once it comes to a full stop and to direct passengers to wait in parts of the platform that would not obstruct passengers from alighting from the train and leaving the platform. Drawing on millions of individual train waiting times, we estimate the effect of this intervention to change passenger behaviour on the platform on train waiting and delay times. We use different sets of assumptions about what the counterfactual change in waiting and delay times would have been in the absence of the intervention. The intervention has reduced train waiting times by up to 6.6%. This reduction came about mainly through reducing delay times of trains once they are delayed. The reductions tend to occur during peak traffic hours. The implied cost-savings amount to a return of £6 per £1 investment.

JEL classifications: C21, C24, D91, R41

Keywords: Social norm, train waiting times, train delay times, public transport

* We would like to thank Andrew Hyman, Toby Goodwin, Alexander Anhwere-James, Hayley Oberlander, and James Cockerton, all from Transport for London, for their help and support in carrying out this work.
[1] City, University of London, and Transport for London.
[2] City, University of London.

1. Introduction

Transport is a key factor in any modern economy and its cost-effective management and supply are crucial if the ever-growing demand is to be met and development sustained (Duarte *et al.*, 2010). Many local and national authorities now implement policies to encourage people to patronise more environmentally friendly modes of transportation which mainly involve the use of public transport (Karekla and Tyler, 2012), adding importance to the reliability of urban public transport systems. Any delay to a high frequency metro network can have far reaching and knock-on effects, as delays to a train service quickly cascade to other trains on the same line. Perceptions of decreased reliability can have long-term adverse effects on customer good-will and – in the case of London – compromise the Mayor's aim of having 80 per cent of all trips in London to be made on foot, by bicycle or public transport by 2041 (TfL, 2018a).

Since the 1990s the rail market has exhibited strong growth (ATOC, 2013; RDG, 2016). A collaborative report by Infrastructure UK and the National Infrastructure Commission finds that passenger demand – if unconstrained by supply limitations – would increase between 84% and 110% between 2015 and 2040 (Preston, 2018). Network Rail (2013) predicts an increase in demand of 58% for non-London commuting and 115% for London commuting between 2011 and 2043. Preston (2018) provides evidence that a generational shift is occurring among 'millennials' who favour public transport over car ownership and predicts a capacity crunch.

Increases in demand will inevitably impact on the reliability of any transport network (Melo et al., 2011; Barron et al., 2013) and increase transport costs. This is reflected in the performance of many transport providers, e.g. declining punctuality in London and the South East counties (RDG, 2016). How can transport networks cope with these increases in demand in the face of capacity constraints?

Many metros now invest in acquiring more capacity for their networks through different means like technical and structural solutions to manage travel demand (Barron, 2016). To manage dwell times at platforms one can resort to hard and soft measures (Bamberg *et al*., 2011). Hard measures include investments in structures and infrastructures; for an old network like London Underground it entails retrofitting existing infrastructures which is usually very expensive and can sometimes pose an engineering challenge. Soft measures on the other hand broadly represent measures which cause conscious or subconscious behavioural change. Traditional methods involve campaigns, signposts, audio-visual cues on the platforms, fare structure manipulation and others (Bamberg *et al*., 2011; Loukopoulos, 2007; Fujii and Taniguchi, 2006). Therefore, where there exist capacity constraints, it is imperative that the providers of the transport network seek smarter and cost-effective ways of managing passenger flow and demand.

Many researchers now question the efficacy of using only traditional economic methods in controlling transport demand. Economists and transport policy makers have hitherto focused on prices to control passenger demand (e.g. charging more during peak hours), but the efficacy of price in the demand function of any public transport network depends critically on its price elasticity, transport policy and passengers' behavioural norms. In addition, there is evidence that non-price interventions can be very effective and relatively inexpensive in some scenarios (Allcott, 2011; Bertrand *et al.*, 2010).

There is limited quantitative evidence based on experimental settings in transport management showing causal effects from norms to behaviour (Metcalfe and Dolan, 2012), and existing work is confined to laboratory experiments. In this paper, we address the question of whether using social norms to nudge passengers into conformity has affected a reduction in dwell times on a key platform on one of the busiest metro networks in the world. We employ data from London Underground's Green Lane experiment at King's Cross station. The experiment was aimed at influencing passenger behaviour by laying green vinyl on the platform supported by audio and visual cues that encourage customers to pass along the platform until they find a non-green space to stand and wait for a train. The green vinyl was laid only on the Southbound Victoria line platform at King's Cross (providing the treated platform), owing to its vantage position as a pinch point location and a central hub on the network created by persistent increase in demand and connections to other inter-urban lines. The vinyl on the platform also shows where the doors would open when the arriving train completely stops at the platform thereby helping passengers know exactly where the doors would be. Data were retrieved for periods before, during, and after the experiment (after the vinyl was removed) for every London Underground train through King's Cross and adjacent stations using specialist software.

We observe a significant reduction in dwell time by 6.6% (2.3 seconds), which is a profound result for London Underground by its own standard. London Underground values one second savings in dwell time at King's Cross station at £68,000 worth of customer benefits (Goodwin, 2017). Therefore, our research indicates that the Green Lane intervention generated customer benefits of £156,400 per year for London Underground at a cost of £25,000 (total costs of materials and labour for installing and decommissioning the Green Lanes at one platform). If dwell times could be reduced by a total of 2.7 seconds, the frequency of trains going through the affected platform could be increased from 36 to 37 trains per hour, resulting in additional customer benefit of £3.6 million.[3] The dwell time reductions mainly occur during peak demand times, and can be mostly attributed to reductions in delayed departures. We further investigate whether the dwell time reductions are sustained after the removal of the signals which caused the behavioural change in the first place, thereby adding to the important discussion on how sustainable interventions suggested by the behavioural science literature are. Our results

[3] This figure comes from conversations we have had with TfL personnel. Operating more trains would decrease the waiting time between trains and the congestion in trains.

suggest that any beneficial effect of the Green Lanes on dwell times disappeared after the Green Lanes were removed.

Methodologically, our paper is distinct from the existing literature in two important respects: First, it is based on a real-world intervention at one of London's busiest stations, and exploits its quasi-experimental nature by comparing the dwell time changes on the treated platform to a number of different potential control platforms, accounting for the possibility of seasonal variations in dwell time, changes affecting the entire station, and changes affecting the entire service line. Our empirical strategy is to estimate the effect of the Green Lanes on dwell times under a set of different assumptions about how dwell times would have evolved in the absence of the experiment. We obtain statistically and economically significant effects in most of our models. The results show that while the correctness of any of those assumptions cannot be known, it seems very improbable that all our estimated treatment effects should be attributable to a factor other than the Green Lanes.

The second methodological contribution is to explicitly model the operating procedure for trains to stop and depart. Not every potential dwell time saving translates into actual savings. If the train is on schedule and needs to wait for its scheduled departure, its dwell time will not decrease, even if alighting and boarding terminates quicker. We model dwell and delay times as latent variables and allow for the Green Lanes to have different effects on dwell and delay times.

The rest of this paper is divided into nine further sections. Section 2 reviews the existing literature. Section 3 discusses the experiment, and section 4 describes the data and estimation methodology. A graphical analysis follows in section 5 and main results are presented in section 6. Sections 7 and 8 discuss heterogeneous trends and further results. Results from the removal of the Green Lanes are in section 9. Section 10 concludes.

2. Literature review

Our paper is closely related to two strands of literature. Behavioural economics and the literature on social norms have analysed the effectiveness of relatively cheap and non-intrusive measures to affect behavioural changes resulting in socially desirable outcomes. The transport literature has identified dwell times as one of the main factors in determining reliability in high frequency networks and found that crowding on the platform and train explains most of the variability in dwell times.

Managing congestion and capacity through the control of train platform dwell time is now receiving increased attention among researchers and operators (Avineri, 2011). Dwell time is an important variable that changes service level and reliability, so its extension or inconsistency can be detrimental to a network's capacity and ability to provide reliable service (Thoreau et al., 2016; Barron, 2016). Hard and soft measures can be distinguished in dwell time management.

Hard measures represent heavy investments which include capital expenditure on structural adjustments like platform expansion, installation of platform edge doors (PEDs), station restructuring, line re-signalling and procurement of new rolling stock. Given train and platform infrastructure, the amount of boarding and alighting passengers, and in-train occupancy have been shown to explain 70% or more of dwell time variation (Lin and Wilson, 1992; Puong, 2000; Rashidi, Ranjitkar, and Hadas, 2014). Examining passenger movement in a laboratory setting, Fujiyama *et al*. (2014) find that adjusting train width and platform step height improved boarding and alighting. This hard measure is a useful consideration in the construction of new metro stations or procurement of new rolling stock but would be a very expensive investment for an existing network to retrofit platforms or adjust train doors. Karekla and Tyler (2012) analysed the Victoria line in London to determine the possibility of reducing dwell time by making specific changes to train system hardware which entails both trains and platform. Using data from London Underground and on-site observation they conclude that adjusting the width of train doors and platform step height are most effective in reducing dwell time, like the conclusion arrived at by Thoreau *et al*. (2016). Again, this is an expensive hard measure especially for an existing network. It would cost London Underground approximately £1.5m per platform to adjust the height (Karekla and Tyler, 2012).

Turning to soft measures, these can be very cost effective and as efficient in controlling dwell time. These measures include non-intrusive interventions such as communication campaigns, pricing and other techniques that support and encourage behavioural change in a certain direction. (Brög *et al*., 2009; Cairns *et al*., 2008; Taylor, 2007; Avineri and Goodwin, 2010; Bamberg *et al*., 2011). Pricing to control demand and therefore platform dwell time is effective (Douglas *et al*., 2011; Liu and Charles, 2013; Currie and Delbosc, 2011; Qu *et al*., 2018). But it comes at a cost to the transport consumer, and it may encourage modal switch to less energy efficient modes. Increasing or decreasing the level of service provision (supply of transport) is a major variable in managing platform dwell time. Subject to line or network capacity, running more or less service helps in managing demand and dwell time in high frequency networks.

Platform communication systems can also encourage passengers to pass along platforms to get on less crowded carriages of an arriving train (Olaverri-Monreal *et al*. 2018; Moncrieff, 2015), but can become counter intuitive as they may encourage 'bunching up' of passengers around the doors of the supposedly emptier carriage. In a field experiment carried out at Schiphol Airport station in Amsterdam, van den Heuvel (2016) finds that adjusting train stopping positions decreased dwell time by 30 seconds during the peak, however Oliveira *et* *al*. (2019) opine that the method does not help with crowd control and is only effective in facilitating the boarding of less busy train carriages.

Another major variable in the dwell time management equation is passenger behaviour (Oliveira *et al*., 2019; Barron, 2016; De Ana Rodriguez et al., 2016; Harris, 2006; Wiggenraad, 2001). It is widely accepted that norms are important in influencing human behavior (Sherif, 1936; Cialdini *et al*., 1991; Merton, 1957; Coleman, 1990). Norms can be private or social. For a norm to be considered "social", it must be acceptable to and shared by the other members of the society. Its sustainability is a function of both the approval and disapproval of the members of the society (Elster, 1989). The distinguishing feature of a social norm is that it does not benefit any one individual but (parts of) society, and the punishment for non-conformity cannot be enforced legally but through social sanctions imposed by others. Private norms on the other hand result in self-imposed sanctions by individuals such as the feelings of embarrassment, shame, and guilt when they do not conform (Sugden, 1987; Elster, 1988 & 1989; Coleman, 1990; Young, 2008; Garnett, 2009). Both private or personal and social norms influence how passengers behave in public spaces. It has been argued that social norms yield pareto-efficient scenarios (Coleman, 1990; Ullman-Margalit, 1977) and are quite efficient and effective in the equitable regulation of social welfare (Akerlof, 1976; Bicchieri et al., 2018; Nolan, 2015). In a situation where a norm leads to a pareto-inefficient situation it is expected to disappear with time. This is the case with most, if not all gender, race, or sexual orientation bias norms (Bicchieri *et al.*, 2018).

Perhaps the closest interventions to the Green Lane experiment are platform communications and markings encouraging passengers to pass along the platform or not to hold train doors. Seriani and Fernandez (2015) evaluate some interventions, including a keep-out zone on platforms. While simulation results suggest a potential to reduce dwell times by 50%, their laboratory experiments found no effect of the keep-out zone, most likely due to observed non-compliance by passengers waiting on the platform. Our paper is thus one of the few which seeks to analyse the implementation of such soft interventions and its impact on dwell times. To the best of our knowledge it is the only one which does so in a real-world setting rather than in computer simulations or laboratory experiments.

3. Background

3.1 The London Underground (LU) Network.

London Underground is the oldest network in the world, which ran its first train service in 1863. Consequently, LU faces capacity and structural challenges as many of the modernisation works to the network are either impossible or very expensive to retrofit. The network consists of 17 different lines connecting 270 stations and extends to 250 miles of track making it the 7th largest (in served passengers) and 3rd longest (in kilometres of track) network in the world. In 2017 the network served about 4 million passenger journeys per day. The central zones are the busiest and connect passengers to many of London's landmarks and financial hubs.

Users of this network face overcrowding on the platforms in the peak times (0700 – 1000 hours and 1600 – 1900 hours). Critical congestion occurs particularly in the 'peak of the peak' (0800 – 0900 hours and 1645 – 1730 hours) when the network is busiest and operating close to its maximum capacity. At these times passengers face delays as they may be unable to board the first available train and even the subsequent ones depending on their position on the platform and level of congestion on both the platform and the arriving train. Consequently, LU is constantly exploring innovative and cost-effective methods of improving customer experience and reducing the generalised cost of travel. One of such ways is the Green Lane project designed to influence passenger travel in a certain way to aid the reduction of dwell times, travel time and costs.

3.2 The Experiment

The Green Lanes project was aimed at influencing passenger behaviour in a transport setting. It was an experiment performed at the platform level in a bid to reduce the generalised cost of travel for customers by decreasing dwell and journey times which would, consequently, increase the capacity for more service frequency at the treated station. This could be achieved through 'nudging' passenger behaviour using visual and audio cues, but without incurring high costs or imposing physical or financial impediments. The installation of the Green Lanes started on July the 18th, 2017 and was completed by the 1st of September 2017. The lanes remained in place until early 2018.

To perform the experiment the southbound platform of the *Victoria line* at London *King's Cross* station (henceforth referred to as the *treated platform*) was chosen because of its central location and the persistent increases in dwell times in recent years. King's Cross station is a major hub and terminal with connections to many parts of London and the United Kingdom. The Victoria line serves several central and important stations linking many of London's landmarks, central and suburban districts. Dwell times on the Victoria line have increased due to congestion brought about by persistent rise in travel demand over the years. King's Cross station constitutes a pinch point location and a bottle neck on the Victoria line. The southbound platform was chosen for treatment because the dwell times on this platform increased significantly from 35 seconds to 47 seconds between 2015/16 working timetables (Goodwin, 2017).

Figure 1: Green lanes at London King's Cross station. Adapted from Goodwin, 2017.

The experiment was simple. As can be seen in Figure 1, green vinyl was laid on the platform in such a way that encourages customers to pass along the platform until they find a non-green space to stand and wait for a train. The green vinyl also showed where the doors would open when the arriving service train comes to a complete stop at the platform edge. The choice for the green colour is in accordance with international convention for green indicating 'clearance to proceed' and made violation of the social norm that TfL was trying to establish (do not stand in the lane) highly salient.

This salience-of-violation feature also distinguishes the experiment from previous attempts by TfL to control passenger behaviour via conventional methods such as speaker announcements and encouragement by staff not to stand in front of the doors, which were effective only for the duration of one dwell time occurrence, but would have to repeated for the next arriving train. If passenger behaviour could be altered by appeals to personal norms only, then we would expect the desired behaviour to become established over time. The Green Lane experiment adds a social dimension by making the norm violation immediately visible – and costly – for other passengers waiting on the platform who can then express their disapproval by established signs (e.g. body language). For example, it is an established norm on the London Underground to stand on the right of a moving escalator to allow people to pass on the left. The salience of the norm violation might have contributed to the relative success in establishing this norm.

The passengers are expected and encouraged to keep moving on any green section of the platform and not to stop until they get to any non-green section. When a train arrives at the platform, the alighting passengers use the space on the Green Lanes spurs to exit the platform so that passengers waiting in the non-green sections can board quicker. In theory, this should eliminate or at least reduce the pushing, bumping and shoving that happens at peak times when the platforms are crowded.

The Green Lanes have been successful in reducing the number of waiting passengers who stand in front of the train doors before they open (TfL, 2018b). They might also have helped to reduce platform train interface (PTI) issues such as passengers or staff getting caught in the closing doors or having items caught between the doors at the platform edge. At all times during the experiment there were visual cues like posters and direction markers encouraging passengers not to stand or wait on any green section but to keep moving while on it. At peak times there would be a member of station staff on the platform giving audio messages, managing overcrowding on the platforms and assisting the train driver in ensuring safety on the platform during train arrival and departure. A close alternative to the Green Lanes are platform edge doors. As the London network is old and extensive, updating them by retrofitting platform edge doors on old and curved platforms would be an engineering challenge and significantly more expensive than the Green Lanes vinyl.

3.3 Operating Procedure.

The Victoria line runs an automatic train operation process, but all the trains have drivers in the front cab. The driver can intervene when automation breaks down as well as assist with doors opening and closing to minimise PTI issues resulting in injuries. There are 16 stations on the route and King's Cross station is somewhat in the middle in Zone 1 (centre) providing interchanges with other LU lines and National Rail services. When the train arrives at the platform, the driver opens the doors for alighting and boarding to commence. The driver monitors this process through the in-cab and platform CCTVs depending on location. Provided the station starting signal is clear, the driver then pushes the 'doors-close' button which effectively brings boarding and alighting to completion. This usually begins with an audible sound that only lasts for a couple of seconds before the doors begin to shut. The trains are fitted with sensitive edge technology; this is a safety device which ensures the doors are completely shut without which the train would be unable to proceed. A 'clear' signal indicates that the route ahead is clear of any train, secure, and safe for the train to proceed. As soon as the doors close the train departs. Once it clears the platform and section of track, it is then safe for the next train to arrive. The process is repeated at the arrival of the next train.

The station starting signal is automatically controlled by a computer preloaded with a predetermined running timetable which has the scheduled arrival and departure times of every train going through every platform on a line. If a train arrives before the scheduled arrival time, the starting signal remains at danger (red) and would only clear at the scheduled departure time. This always remains the case safe for when a failure occurs, at which point the signalling process would be controlled by a duty Signalling Operator and the service itself regulated by a duty Service Controller. The passengers are encouraged to stand at the non-green areas so as not to obstruct the flow of traffic especially at peak times. At both the AM and PM peak times there is usually a station staff member on the platform assisting with platform duties. They were tasked with encouraging passengers to acknowledge the Green Lane and to conform to the rule of not standing or waiting anywhere on the Lane. The Green Lane keeps the walkways clear for alighting passengers to easily disembark from the train so that boarding can commence quicker.

4. Data and methodology

4.1 Data

The Green Lane installation commenced on the 18th of July 2017 with the lanes laid on the inbound Victoria line platform.[4] As it was not completed until the 1st of September of the same year, for the main analysis we exclude the time period during which the installation was in progress (18.7. to 1.9.). Appendix A2 contains results for the full sample and where the start of the Green Lanes is assumed to be the 18th of July. The Green Lane experiment was not announced or made public to passengers before its start. The dwell time data were supplied by TfL using specific software that records detailed real time train movements to the second. The data includes dwell time counts for London Underground services from 5am to 1.30am from Monday to Friday (weekdays), but we restrict our sample to trains departing between 6am and 11pm, since platform crowding does not occur outside this time interval. The weekend data have been excluded owing to lower weekend service frequency, prevalence of engineering works leading to platform or station closures and decreased weekend demand. Any abnormally long (>70 seconds) or short (≤10 seconds) dwell time is disregarded, retaining 98% of dwell time observations. We discard the right tail of dwell times as these long times are likely to have been

[4] Our main analysis therefore only uses trains travelling *towards* the City centre. We repeat the analysis for trains travelling outwards as a placebo-experiment. See section 8.

caused by incidences such as train or signal failures. We exclude the left tail of dwell times as these short times could only be achieved with very low passenger numbers (both on the platform and in the carriage). The Green Lanes would not have any effect on dwell times in both of those scenarios. We have also repeated our analysis for dwell time observations between 5 and 75 seconds. This did not change any of our main results.

All trains arriving at the study platforms have an arrival time and a departure time. The time difference between these variables gives the dwell time (Dwell) in seconds. We extracted dwell times for all trains which pass through King's Cross station or an adjacent station, for the time periods from 21st of May to 30th of November for 2016, 2017 and 2018. For any train in our data we have the following information: its arrival and departure times, which station it dwelled at, which service line it served, and which direction it was travelling in.

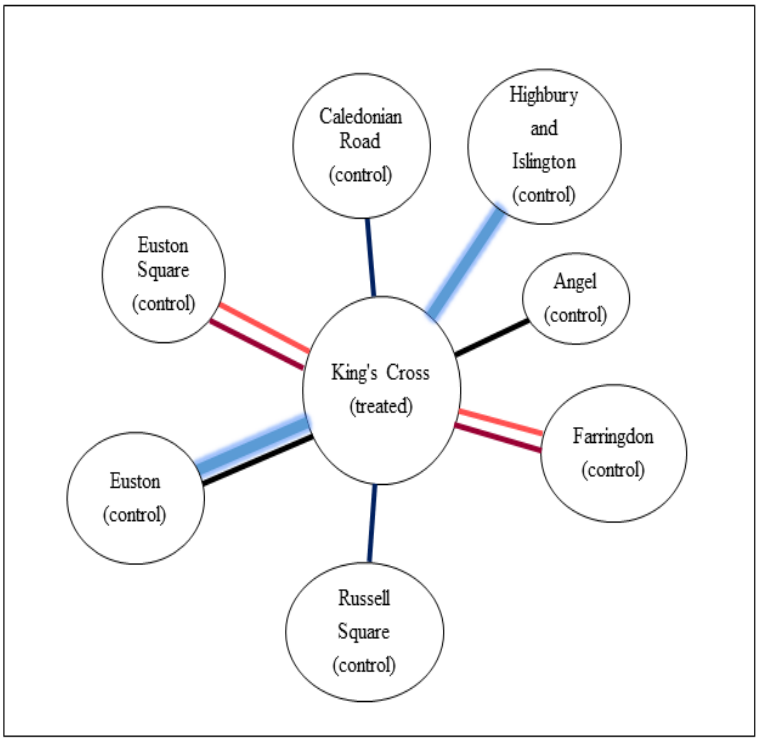


Figure 2: Stations and services lines of the study. King's Cross, the treated station, is served by the Victoria line (treated line, thick light blue), the Northern line (black), the Piccadilly line (dark blue), the Metropolitan line (dark red), and the Hammersmith and City line (light blue).

Figure 2 shows the scope of our data and highlights the treatment and the control stations and service lines. The treated platform is the Victoria line platform at King's Cross station. Service lines other than the Victoria line (thick line) are control lines, and stations other than King's Cross station (circle in the middle) are control stations. For each year, we have approximately 1 million dwell time observations.

<u>4.2 Empirical Strategy</u>

Comparing average dwell times at the treated platform after the Green Lanes installation to before their installation is unlikely to produce the true effect of the Green Lanes on dwell times. The Green Lane intervention deviates from an ideal experiment in three important ways:

- All dwell times under treatment only fall into the period of September 1st to November 30th,
- All dwell times under treatment are observed only on the Victoria line,
- All dwell times under treatment are observed only at King's Cross station.

Thus, treatment status is deterministic rather than random. Knowledge of date, station, and service line of an observation would perfectly reveal its treatment status. If we only compared observations on the treated platform after and before the intervention, then we would be picking up the effect of the intervention *and* any other factors which might cause dwell times to be different between September 1st and November 30th compared to the pre-treatment period (e.g. weather conditions, passenger demand, etc). For this reason, we rely on quasi-experimental methods which are frequently deployed in empirical economics.

The Difference-in-differences (henceforth 2D) estimator seeks to mimic experimental designs by identifying observations which would serve as an appropriate control group to the treated observations. If, for example, the above-mentioned factors affect all platforms equally, then the effect of the Green Lanes can be identified as the change in dwell times for the treated observations, *over and above* the change in dwell times of (a subset of) all other platforms. In practice, it is unlikely that all other platforms would have been affected by factors affecting dwell time in equal measure. We thus must be more careful in the selection of appropriate control observations. We consider the following scenarios:

1. The factors affecting dwell times on the treated platform (other than the Green Lanes) are the same in 2016 and in 2017. This would qualify observations on the treated platform *in the previous year* as appropriate control observations.
2. The factors affecting dwell times (other than the Green Lanes) affect all observations in King's Cross Station equally. This would qualify observations on non-treated service lines at King's Cross Station as appropriate control observations.
3. The factors affecting dwell times (other than the Green Lanes) affect all stations on the Victoria line equally. This would qualify observations on the Victoria line in stations other than King's Cross Station as appropriate control observations.

All three scenarios are much less restrictive and more plausible than saying that *all* other platforms would constitute a good control group. Each scenario requires that we assume that external factors affect dwell times for treated and control observations equally.[5] Since we cannot know or test which of the three scenarios is closest to reality, we estimate the Green Lane effect for all three scenarios.

The above-described scenarios highlight the limitations of the 2D estimator. What if, for example, the dwell times of treated observations are affected by the Green Lanes, *and* by time-of-the-year effects which affect dwell times equally in 2017 and in 2016, *and* by effects which affect all platforms in King's Cross Station equally? We would need to combine scenarios 1 and 2 to reflect this. The triple difference (henceforth 3D) estimator does precisely that. For this example, for observations at King's Cross Station, we would first obtain the change in dwell times (first difference) in 2017 and subtract from this (second difference) the change in dwell times in 2016, only for observations which are not on the Victoria line. This quantity would tell us by how much the time-of-year effect at King's Cross Station has changed from 2016 to 2017 on service lines other than the Victoria line. *If* we assume that this change in the time-of-year effect would affect the Victoria line equally, *then* the difference (third difference) in this change between the Victoria line and other service lines would recover the pure effect of the Green Lanes. Combining any two of the above three scenarios results in three different 3D models. Again, we estimate all three of those models.

Finally, all three confounding scenarios described above might be present. In a further generalisation of the 3D estimator we propose a quadruple difference (4D) estimator. If we followed the same steps as described in the previous paragraph, but this time for stations other than King's Cross Station, then we would have a 3D

[5] We can recover two quantities: The effect of other factors for control observations, and the combined effect of other factors for treatment observations *and* the Green Lanes. We can recover the Green Lanes effect only if we assume that other factors affected control and treatment observations equally.

estimate of combined service line and time-of-year effects on dwell times for other stations. Under the assumption that this combined effect is the same for King's Cross and other stations, the difference (fourth difference) between the 3D estimate for King's Cross Station and other stations would recover the effect of the Green Lanes. We summarise the different models that we estimate in table 1 along with the assumptions which are required to recover the effect of the Green Lanes. Appendix 1 contains the statistical models and their explanations. All models include, where applicable, the control variables described in the next section.

Table 1: Different estimators and assumptions

Estimator	*Control group*	*Identifying Assumption*
Simple difference (Before-after)	Treated platform before the Green Lanes	There would be no difference in dwell times on the platform between fall and summer
2D estimator	Treated platform in previous year	The difference in dwell times on the platform between fall and summer would be the same in 2016 and 2017
2D estimator	Other stations on the same service line	The difference in dwell times between fall and summer would be the same for all stations on the service line
2D estimator	Other service lines in the same station	The difference in dwell times between fall and summer would be the same for all service lines in the station
3D estimator	Treated platform in previous year and other stations on the same service line	The change from 2016 to 2017 in the difference in dwell times between fall and summer would be the same for all stations on the service line
3D estimator	Treated platform in previous year and other service lines in the same station	The change from 2016 to 2017 in the difference in dwell times between fall and summer would be the same for all service lines in the station
3D estimator	Other stations on the same service line and other service lines in the same station	The difference across stations in the difference in dwell times between fall and summer would be the same for all service lines
4D estimator	Treated platform in previous year and other stations on the same service line and other service lines in the same station	The difference across stations in the change from 2016 to 2017 in the difference in dwell times between fall and summer would be the same for all service lines

Notes: Summer refers to the pre-treatment period May 21st to July 17th, fall refers to the treatment period September 1st to November 30th.

4.3 Variables

The dependent variable in our analysis is the natural logarithm[6] of dwelling time denoted Y_{tsl} where t is time, s denotes the station, and l the service line. The main independent variables for the 2D, 3D, and 4D estimators are:

$Post_t$ = 1 if t is later than 6am, September 1st, in either year (2016, or 2017), and 0 otherwise,

$D2017_t$ = 1 if t is in 2017, and 0 if t is in 2016,

$Kings_s$ = 1 if s is King's Cross station, and 0 otherwise,

$Victoria_l$ = 1 if l is the Victoria line, and 0 otherwise.

An observation is identified as subject to the Green Lane treatment, if (and only if) all those indicator variables are equal to 1. Appendix A1 describes in detail how these variables are used in the consistent estimation of the Green Lanes effect on dwell times.

Where applicable, and following our discussion in section 2, we also include the following control variables: *Demand* is the sum of daily station entries and exits. We include it as a control variable since higher demand is likely to increase dwell times. Since the demand variable is a total daily count of the number of passengers through the gates of a station, it is difficult to apportion the passengers to individual platforms at the stations of entry or exit. We thus assume demand to be constant throughout the day and across station platforms. *Lines* is the number of service lines through a station: given a level of demand, more service lines would distribute

[6] Residuals from a dwell time regression exhibit a log-normal distribution. We therefore use the natural logarithm of dwell times as the dependent variable, which also results in a better model fit in terms of R^2.

station demand over more platforms and result in less crowding on the platform. In addition, *DemandPerLine* (*Demand* divided by *Lines*) has been added to account for average demand per platform in a station. *ServiceLevel* (in seconds) measures the time interval between two scheduled train arrivals and is expected to be negatively correlated to dwell time; higher service frequency reduces station dwell times. *ServiceLevelDemand* is a variable interacting demand and the level of train service on a line, allowing the effect of service frequency on dwell times to depend on demand. Since demand will not be uniformly distributed over the time of day, or over the days of the week, we also include dummies for each 15-minute interval of a day (from 6am to 23pm), and for each weekday. Finally, we include a linear time trend[7] which is restricted to be the same for all trains and both years (that is, we include a variable equal to x if the observation is for the x^{th} day of any year). We relax this restriction in section 7 when allowing for heterogeneous trends. For simplicity, we collect the control variables described in this paragraph in a vector X. Table 2 provides the complete list and description of variables.

<u>4.4 Delay Time Analysis</u>

Given scheduled departure times of trains, any intervention to speed up alighting and boarding times of passengers would affect dwell times only if the train exceeds or is close to exceeding its scheduled departure time. Otherwise, even if alighting and boarding completes faster, the train would have to wait on the platform until its scheduled departure time, thus only prolonging the time where it is idle.

[7] We have also estimated our models with quadratic time trends, but this had only a negligible impact on our estimates.

Table 2: Variables

Variable	Description
Dependent variable	
Y_{tsl}	Natural log of dwell time at time *t*, station *s*, and service line *l*
Main independent variables	
$Post_t$	Binary variable equal to 1 if *t* is between September 1 and November 30 (in either year)
$D2017_t$	Binary variable equal to 1 if *t* is in 2017
$Kings_s$	Binary variable equal to 1 if *s* is King's Cross station
$Victoria_l$	Binary variable equal to 1 if *l* is the Victoria line
Control variables	
$Demand_{ts}$	Number of entries and exits into station *s* (daily)
$Lines_s$	Number of service lines through station *s*
$DemandPerLine_{ts}$	$Demand_{ts}$ / $Lines_s$
$ServiceLevel_{tl}$	Scheduled time interval between two trains in seconds. Indexed also by *t* as service levels can vary over the day.
$ServiceLevelDemand_{tsl}$	$ServiceLevel_{tl} \times Demand_{ts}$
$D600_t$	Binary variable equal to 1 if *t* is between 6am and 6.15am
…	…
$D2245_t$	Binary variable equal to 1 if *t* is between 10.45pm and 11pm
$DMonday_t$	Binary variable equal to 1 if *t* falls on a Monday
…	…
$DFriday_t$	Binary variable equal to 1 if *t* falls on a Friday
$Trend_t$	Linear time trend: 0 for May 21 (in either year), then incrementing by 1 for each calendar day

We therefore extend the analysis to an investigation of whether the Green Lanes had a stronger effect on reducing delay times rather than dwell times in general. To do this, we have to take into account that 1) we observe delay times only once a train exceeds its scheduled departure time, and 2) we do not observe the train's regular dwell time once it is delayed (rather, the dwell time is censored). Consider the following empirical model. A train is scheduled to stay on the platform for $\bar{t}$ seconds. A latent dwell time variable Y^*_{tsld} and a latent delay time variable Z^*_{tsld} are given by

$$Y^*_{tsld} = \beta X_{tsld} + \epsilon^y_{tsld}$$

$$Z^*_{tsld} = \gamma X_{tsld} + \epsilon^z_{tsld}$$

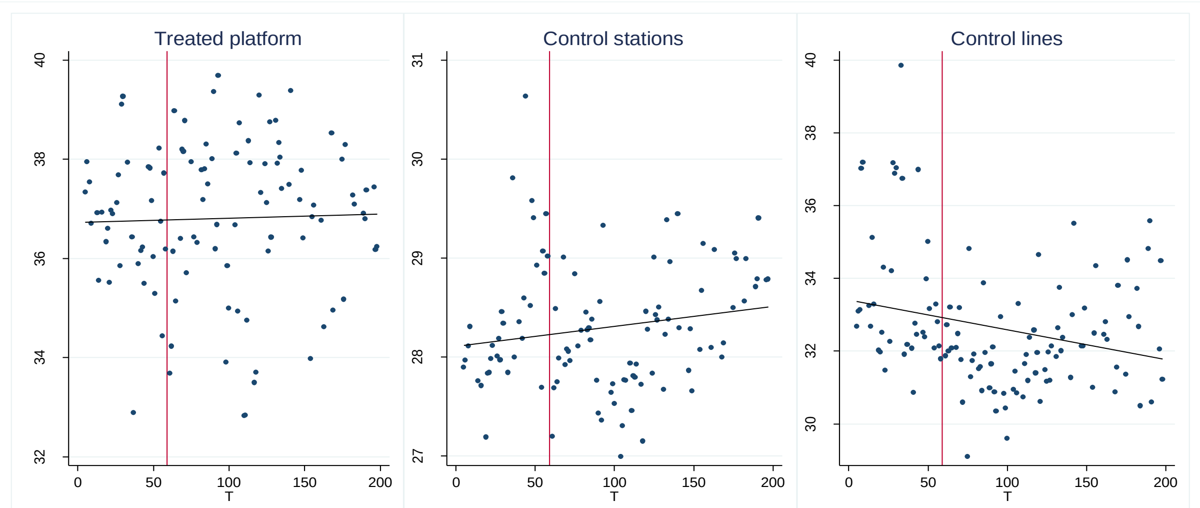
The actual dwell time Y_{tsld} is observed only if $Y^*_{tsld} \leq \bar{t}$, in which case $Y_{tsld} = Y^*_{tsld}$, and the delay time is unobserved. If, on the other hand $Y^*_{tsld} > \bar{t}$, then the actual delay time is given by $Z_{tsld} = Z^*_{tsld}$, and we know that the latent dwell time exceeds the schedule $\bar{t}$: $Y^*_{tsld} > \bar{t}$. We assume that the errors are jointly normally distributed.

Amemiya (1985) derives closed form solutions of the likelihood of this class of models (p. 386). We estimate this model with maximum likelihood. No restrictions are placed on the parameters within or across the latent variable equations. We thus allow the Green Lanes to have differential effects on regular dwell and delay times.

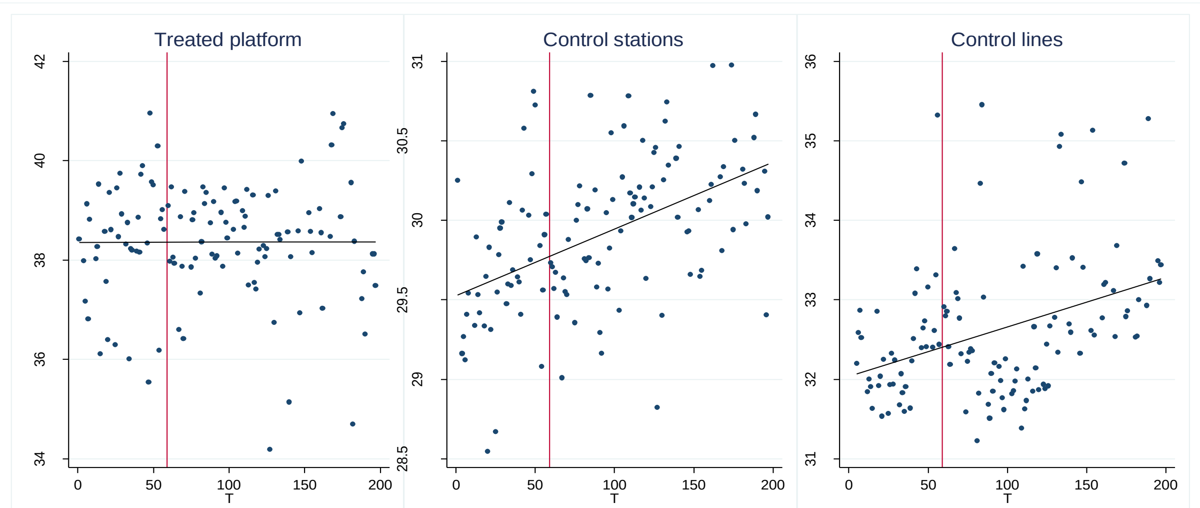
One constraint that we face is that we do not observe the scheduled arrival and departure times of all trains. However, from trains for which we do observe these data, we can impute a train's scheduled dwell time. Our delay time is thus based on the difference between scheduled and actual *dwell* times, rather than the scheduled and actual departure times. For example, if a train arrives 5 seconds behind schedule, and departs 5 seconds behind schedule, it would be classified as departing just on time according to our imputation. Because the service frequency is high (on the Victoria line a train runs every 2-3 minutes), this misclassification should not be of great concern, as passengers do not arrive at the platform with the intention of catching a particular scheduled train, but rather to get on the next available train regardless its scheduled departure (actual timetables for the Underground are not displayed on the platforms or on TfL's website).

5. Graphical analysis

We begin by visually inspecting the average daily dwell times of trains. The upper panel of figure 3 shows dwell times in 2016, and the lower panel dwell times in 2017. The vertical line corresponds to the 18th of July, the day at which the Green Lane installation began in 2017.[8]



Dwell time = 36.7 + 0.0008T R^2 = 0.1%   Dwell time = 28.1 + 0.002T R^2 = 2.8%   Dwell time = 33.4 - 0.008T R^2 = 6.4%



Dwell time = 38.4 + 0.000T R^2 = 0.0%   Dwell time = 29.5 + 0.004T R^2 = 23.0%   Dwell time = 32.0 + 0.006T R^2 = 14.3%

Figure 3: Average daily dwell times in 2016 (top) and 2017 (bottom) at treated platform (left), adjacent stations on the Victoria line (middle), and other service lines at King's Cross station. The vertical line corresponds to the 17th of July (when the installation of the Green Lanes in 2017 started).

[8] Figure 3 includes all days. For the regression analysis, we exclude observations from the installation period (18th July to 1st September) as described in section 4.1.

We observe that dwell times on the treated platform (graph on the left) are longer in 2017 than in 2016 by about a second. However, the dwell time trend over the year is flat in both years. Thus, since dwell times did not trend down in 2017 compared to 2016, one would conclude that the Green Lanes have not reduced dwell times. We stress here that the graphs only show unconditional regression lines of dwell times against a linear trend while our main analysis includes a rich set of explanatory variables.

If the comparison is made against other stations (graph in the middle) or other lines (graph on the right) in 2017, we see that while dwell times on the treated platform were flat, dwell times at other platforms increased over the same period. Thus, if the counterfactual trend of dwell times on the treated platform would be captured by either of those control platforms, one would conclude a successful reduction (or prevention of increase) of dwell times. The graphs also give an intuition about what we could expect from a triple difference estimation. The difference in the slope of the regression lines between 2017 and 2016 is close to zero on the treated platform. For other stations, we see a steep increase in 2017, and a less steep increase in 2016.

Thus, a difference in differences estimate involving other stations in 2017 should produce a strong negative effect. However, if we assume that the difference in slopes on the treated platform between 2017 and 2016 would have followed the same trend as the difference in slopes in the control stations, then the estimated treatment effect will be smaller.

The graphs show that choice of control observations is crucial. While inspecting trends in the pre-treatment period can be suggestive, uncertainty about the counterfactual evolution of dwell times on treated observations cannot be resolved. We therefore present in the next section results from alternative models.

6. Main results

The main results are presented in Table 3. The first column shows results from the linear dwell time model, while the second and third columns are results for the delay time model described in section 4.4. The second column shows the impact of the Green Lanes on latent dwell times, while the third column shows the effect on delay times. We start by comparing the dwell times before and after the Green Lane intervention on the Southbound Victoria line at King's Cross station. The result suggests a modest and statistically insignificant increase in dwell time by

Table 3: Treatment effect estimates of Green Lanes

	Dwell Analysis	Delay Analysis	
Model	Effect on dwell time	Effect on dwell time	Effect on delay time
Simple difference (before-after)	1.3*	5.0	-5.6
	(0.6)	(4.0)	(3.0)
Difference-in-differences (1)	0.2	3.9*	-4.6**
	(0.3)	(1.8)	(1.2)
Difference-in-differences (2)	-1.2**	0.5	-2.1
	(0.3)	(1.6)	(1.4)
Difference-in-differences (3)	-1.0**	1.7	-5.1**
	(0.3)	(2.1)	(1.1)
Triple difference (1, 2)	-2.0**	0.2	-8.7**
	(0.4)	(2.3)	(1.9)
Triple difference (1, 3)	-5.1**	-0.1	-12.6**
	(0.4)	(2.6)	(1.7)
Triple difference (2, 3)	-2.1**	1.4	4.4**
	(0.4)	(2.3)	(1.5)
Quadruple difference (1, 2, 3)	-6.6**	-1.5	-3.9
	(0.5)	(3.2)	(2.3)

Notes: The coefficients are percentage changes in dwell/delay times. Standard errors are in parentheses. * $p < 0.05$, ** $p < 0.01$. (1) uses the treatment platform in 2016 as control observations, (2) uses adjacent stations on the Victoria line as control observations, (3) uses other service line platforms at King's Cross station as control observations. All regressions control for *Demand*, *Lines*, *DemandPerLine*, *ServiceLevel*, *ServiceLevelDemand*, day of the week dummies, dummies for each 15-minute interval of the day, as well as a linear time trend. See also methodology section.

1.3% on the treated platform from pre to post treatment (*Simple difference*). We then use dwell time on the same platform and in the same period in 2016 as control to measure the possible treatment effects in 2017; this is intuitive as it accounts for any seasonal patterns on dwell time by comparing platform dwell times for both years (*Difference-in-differences (1)).* Here, we observe no significant reduction in dwell time on the treated platform compared to 2016. However, if we compare the dwell times in 2017 between King's Cross and its adjacent stations (*Difference-in-differences (2)*), we find a 1.2% reduction in dwell times. For the model that uses other service lines as a control group (*Difference-in-differences (3)*) we find a dwell time reduction of 1.0%.The next three rows display the estimated treatment effect for the triple difference estimator. *Triple difference (1, 2)* uses adjacent stations on the Victoria line and the treated platform in 2016 as the two groups of control observations. *Triple difference (1, 3)* uses other service lines at King's Cross station and the treated platform in 2016 as control observations. *Triple difference (2, 3)* uses adjacent stations on the Victoria line and other service lines at King's Cross station as control observations. The estimated reductions in dwell times range from 2.0% to 5.1%. Finally, the quadruple difference estimator – the most general model – suggests that dwell times were reduced by 6.6% – a reduction of 2.3 seconds.

The second and third columns explain whether these changes in dwell times came about through a general reduction of dwell times or through cutting the delay times of trains which were behind schedule. Since the service level and timetable in 2017 was not changed we would expect the reduction to come mainly through reduced delay times. For most specifications we observe reductions of delay time, ranging from an insignificant 2.1% to 12.6%. Only the *Triple difference (2, 3)* model finds a positive impact of the Green Lanes on delay times.

7. Heterogeneous trends

The identification of the treatment effect relies on the assumption that the dwell times of the treated observations would have changed by the same amount as the dwell times of the non-treated observations in the absence of the treatment.[9] While we cannot know whether this is the case, we can check whether treatment and control observations were trending in parallel before the treatment. We conduct tests for all our models in table 3 in the pre-treatment period with the null hypothesis of parallel trends for control and treatment observations. Unfortunately, in all cases this hypothesis is rejected (results not reported). All our tests strongly indicate that the treatment observations were trending up relative to the control observations in the pre-treatment period. How would this bias our coefficient of the Green Lane effect? If we extrapolated these trends, then the difference in dwell times between treatment and control observations would increase over time, resulting in a positive, but spurious, estimate of the Green Lanes. If after the treatment the trends run in parallel, we would still have a positively biased estimate, as the difference in dwell times between treatment and control observations after the treatment would still exceed the same difference before the treatment. A negative, and therefore compromising, bias occurs only if the growing gap in dwell times before the treatment reverts. In that case, if the difference in dwell times after the treatment is smaller than before the treatment, we would obtain a negative estimate for the Green Lane effect. The visual inspection of the dwell times does not suggest that this is the case.

[9] Analogous parallel trend assumptions can be formulated for the triple and quadruple difference models. See methodology section.

To analyse the robustness of our results with respect to different pre-existing trends we re-estimate our models allowing for separate trends between different observations. For example, for the first difference-in-differences model, we allow for different trends in 2016 and 2017, while for the triple difference model which uses both years and different stations, we allow for different trends for King's Cross station in 2016, King's Cross station in 2017, other stations in 2016, and other stations in 2017. The results of these alternative models are in table 4. One important difference compared to the main results is the reversal of the effect for model *Difference-in-differences (3)*. Since we can see that the dwell times of observations on other service lines are increasing over time, allowing for separate trends does not attribute the growing gap between other lines and the Victoria line to the treatment anymore and reverses the effect.

Table 4: Treatment effect estimates of Green Lanes - Heterogenous trends

	Dwell Analysis	Delay Analysis	
Model	Effect on dwell time	Effect on dwell time	Effect on delay time
Difference-in-differences (1)	-0.0	8.5	-24.9**
	(0.8)	(4.9)	(4.0)
Difference-in-differences (2)	-0.7	0.6	-1.4
	(0.7)	(4.6)	(3.8)
Difference-in-differences (3)	4.0**	11.0	-9.1**
	(0.8)	(5.7)	(2.9)
Triple difference (1, 2)	-7.6**	1.6	-38.5**
	(1.0)	(6.0)	(5.0)
Triple difference (1, 3)	-0.9	5.4	-40.3**
	(1.1)	(7.0)	(4.5)
Triple difference (2, 3)	-2.0*	3.1	-26.2**
	(0.9)	(6.3)	(4.0)
Quadruple difference (1, 2, 3)	-9.5**	0.6	-70.1**
	(1.4)	(8.5)	(6.1)

Notes: The coefficients are percentage changes in dwell/delay times. Standard errors are in parentheses. * $p < 0.05$, ** $p < 0.01$. (1) uses the treatment platform in 2016 as control observations, (2) uses adjacent stations on the Victoria line as control observations, (3) uses other service line platforms at King's Cross station as control observations. All regressions control for Demand, Lines, DemandPerLine, ServiceLevel, ServiceLevelDemand, day of the week dummies, dummies for each 15-minute interval of the day, as well as linear time trends for each combination of treatment and control platforms. See also methodology section.

The negative effects of the Green Lanes in the triple and quadruple difference models however are preserved for dwell times, and we estimate very substantial reductions for latent delay times.

8. Further results

We next turn to the analysis of dwell times differentiated by time of day. London Underground typically splits the day into AM peak (07.00 to 10.00), PM peak (1600 to 1800), Inter-peak (10.00 to 1600) and off peak (any time outside these times). Trains dwell longest in the AM and PM peaks due to demand as these are when commuters go to work, do school runs, etc. The purpose is to examine if the Green Lane policy had a differing effect in any part of the day. Figure 4 is a plot of the estimated effects and confidence intervals from the quadruple difference model by time of the day in 15-minute intervals. We observe reductions in dwell time throughout the day, but significant effects are concentrated around the morning and evening peak hours. As platform demand ramps up in both peak periods, it appears that passengers tend to conform to the platform norm by obeying the Green Lane policy; this in turn drives down dwell time by a fraction. During periods of less demand the Green Lane's effect is not so significant because the main driver of dwell time and a major cause of impedance to boarding and alighting is demand.

The graph suggests that the Green Lane had higher peak effects (both AM and PM) than on the inter-peak and off-peak periods. This concurs with the normative theory that passengers tend to conform more when they know their actions impact on others, which is more pronounced at peak times because of high demand for waiting area. At other times, when the supply of waiting area exceeds demand, passengers do not bother so much about how or where they wait as there is abundant space for alighting and boarding to take place, sometimes simultaneously.

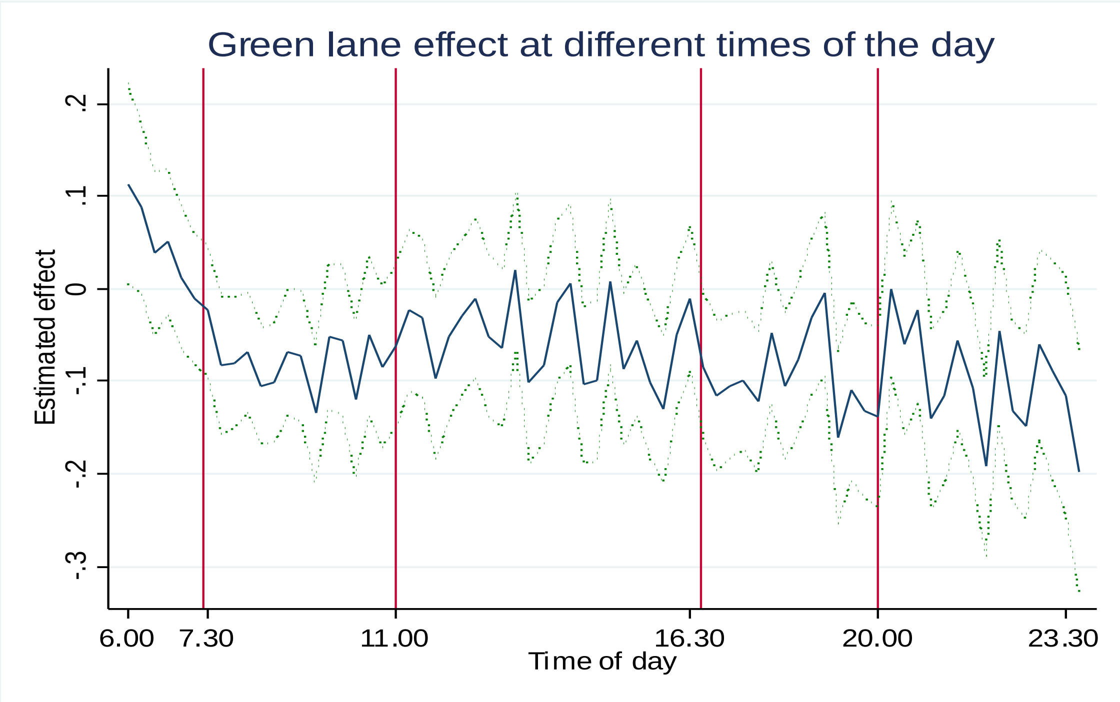


Figure 4: Estimated effects (solid blue) and confidence intervals (dashed greed) of Green Lanes on dwell times.

We next analyse whether the Green Lanes had any effect on the opposite direction of travel. It is tempting to think of the opposite travel direction as a placebo experiment because of the similarities of the two platforms. The outbound platform of the Victoria line at King's Cross station shares most characteristics with the inbound platform: it serves the same line, and therefore has the same service level; it is located at the same station (the two platforms are immediately connected, only separated by a passenger corridor; in particular, the platforms are not on opposite ends of the rails), and therefore has the same daily station demand and the same number of interchanges. The exact number of passengers served by a platform is not observed, but it is probable that the platforms serve roughly the same number of people within a day – e.g. commuters whose return journey is the reverse of their onward journey. The outbound platform differs from the inbound platform in two respects: First, the distribution of passengers over the day is probably different, e.g. most traffic on the inbound platform might be concentrated in the morning, while most of the traffic on the outbound platform might be concentrated in the evening. Unfortunately, we observe only daily station demand, but do not know the exact time of day, nor how it is distributed across the service lines. Second, the outbound platform was not treated with the Green Lanes.

However, there would be a great intersection between people using the platform in the inbound and the outbound direction, for example commuters. It is conceivable that changes to their behaviour on the treated platform extend to other platforms as well. Table 5 presents results for the outbound direction.

Table 5: Treatment effect estimates of Green Lanes - outbound direction

	Dwell Analysis	Delay Analysis	
Model	Effect on dwell time	Effect on dwell time	Effect on delay time
Simple difference	2.2**	9.3*	-5.6*
	(0.6)	(4.7)	(2.6)
Difference-in-differences (1)	0.4	1.7	2.7
	(0.3)	(2.1)	(1.5)
Difference-in-differences (2)	-0.9**	-0.8	-0.4
	(0.3)	(2.0)	(1.2)
Difference-in-differences (3)	1.4	3.2	-2.3*
	(0.3)	(2.3)	(1.1)
Triple difference (1, 2)	-1.0*	-2.4	2.0
	(0.4)	(2.4)	(1.9)
Triple difference (1, 3)	-1.0*	-0.0	5.1**
	(0.4)	(2.8)	(1.7)
Triple difference (2, 3)	-0.4	-1.5	1.4
	(0.4)	(2.7)	(1.3)
Quadruple difference (1, 2, 3)	-1.8**	-5.1	6.8**
	(0.5)	(3.5)	(2.2)

Notes: The coefficients are percentage changes in dwell/delay times. Standard errors are in parentheses. * $p < 0.05$, ** $p < 0.01$. (1) uses the treatment platform in 2016 as control observations, (2) uses adjacent stations on the Victoria line as control observations, (3) uses other service line platforms at King's Cross station as control observations. All regressions control for Demand, Lines, DemandPerLine, ServiceLevel, ServiceLevelDemand, day of the week dummies, dummies for each 15-minute interval of the day, as well as a linear time trend. See also methodology section.

The treatment effects on dwell times are much smaller in magnitude compared to the inbound observations, but we still find significant dwell time reductions on the order of 1.0% to 1.8% in the triple and quadruple difference models. In most specifications no significant effect on latent dwell and delay times are found. The result could mean two things: perhaps the outbound direction is an accurate scenario to describe what would have happened to the inbound direction in the absence of treatment, suggesting that our main results are somewhat negatively biased (they overestimate the reduction in dwell times). Alternatively, passengers might have extended their platform behaviour to the outbound platform as well, also reducing dwell times there, though not as strongly as on the treated platform.

9. Removal of the Green Lanes

Our final analysis is on whether the Green Lane effect has been permanent. The Green Lanes were removed in early 2018. Did passengers revert to non-compliant behaviour and thus cause a reversion in dwell times? To this end, we now define the year 2016 as pre- and the year 2018 as post treatment periods and omit the year 2017 altogether. This causes us to lose one dimension of control, so we have two difference-in-differences and one triple difference model. However, the concern that led us to include the year 2016 as one control dimension in the main section was the possibility of seasonal patterns in dwell time. Since we are now comparing across rather than within years, seasonality should not be a problem.

Table 6 shows what happened to dwell times after removal. Relative to 2016, dwell times are greater when compared against control stations, and when compared against control stations and control service lines. Only when compared against control service lines do we find a negative effect. However, the effect is statistically not distinguishable from zero. For delay times, two specifications support a reduction, while one supports an increase in dwell times. Overall, the evidence points to the disappearance of any beneficial effect of the Green Lanes after their removal.

Table 6: Treatment effect estimates of Green Lanes after removal

Model	Dwell Analysis Effect on dwell time	Delay Analysis Effect on dwell time	 Effect on delay time
Simple difference	2.6**	10.4**	-28.9**
	(0.2)	(1.4)	(1.1)
Difference-in-differences (2)	8.4**	21.2**	18.8**
	(0.2)	(1.3)	(1.1)
Difference-in-differences (3)	-0.1	-5.3**	-29.6**
	(0.2)	(1.3)	(0.8)
Triple difference (2, 3)	1.6**	2.6*	-4.3**
	(0.2)	(1.3)	(1.1)

Notes: The coefficients are percentage changes in dwell/delay times. Standard errors are in parentheses. * $p < 0.05$, ** $p < 0.01$. (2) uses adjacent stations on the Victoria line as control observations, (3) uses other service line platforms at King's Cross station as control observations. All regressions control for Demand, Lines, DemandPerLine, ServiceLevel, ServiceLevelDemand, day of the week dummies, dummies for each 15-minute interval of the day, as well as a linear time trend. See also methodology section.

10. Conclusion

Transport plays a key role in any economy. If demand continues to grow, then policy makers and train operating companies must investigate and invest in smarter ways of managing demand so that transport services can be supplied efficiently. One of such ways is the management of passenger behaviour through establishing norms that would encourage a paradigm shift and sustained behavioural change. This paper shows that norms can be very cost efficient when applied properly as was achieved at London King’s Cross station. The Green Lanes changed passenger behaviour as people conformed to the existing platform norm. Passengers knew where the doors would be when the train comes to a complete stop at the platform edge. This reduced the interactions at the train side as passengers queued at the space provided by the Green Lane, which allows for clear doorway for the alighting passengers. The reduction in dwell time was achieved because of the combination of passengers standing at the right place and allowing passengers off the arriving train. This is particularly relevant to established transport networks operating close to or at full capacity and for which infrastructural adjustments can be prohibitively expensive. For a fraction of the cost of procuring new rolling stock or adjusting station structure to meet growing demand, the Green Lane policy achieved a reduction, or at least a prevention in the increase of dwell times at the treated platform vis-à-vis the control platforms. More modern networks might not benefit as much (or indeed be under pressure to reduce dwell times) as platforms and trains might already have other mechanisms in place to control passenger positions and flows (e.g. platform edge doors, barriers at the train entrance, etc).

References

Akerlof, G., 1976. "The economics of caste and of the rat race and other woeful tales." *Quarterly Journal of Economics*, 90(4): 599-617.

Allcott, H., 2011. "Social norms and energy conservation." *Journal of Public Economics*. 95(9-10): 1082-1095.

Amemiya, T., 1985. Advanced Econometrics, Cambridge, MA: Harvard University.

ATOC (Association of Train Operating Companies), 2013. *Growth and prosperity. How franchising transformed the railways into a British success story*. Available at : http://www.atoc.org/download/clientfiles/files/ATOC%20Growth%20and%20Prosperity%20report.pdf.

Avineri, E., Goodwin, P., 2010. *Individual behaviour change: Evidence in transport and public health.* Project Report. Department for Transport, London. [online] Available from www.eprints.uwe.ac.uk [Accessed 2 February 2019].

Avineri, E., 2011. Applying Behavioural Economics in the Design of Travel Information Systems. In: Anon. The 43rd Annual UTSG (The Universities' Transport Study Group) Conference, Milton Keynes.

Bamberg, S., Fujii, S., Friman, M., Gärling, T. 2011. "Behaviour theory and soft transport policy measures." *Transport Policy*. 18(1): 228-235.

Barron, A., Melo, P.C., Cohen, J.M., Anderson, R.J. 2013 "Passenger-focused management approach to measurement of train delay impacts." *Transportation Research Record: Journal of the Transportation Research Board*. 2351(1): 46-53.

Barron, A., 2016. "Dwell time reductions: good design plus people Management." [online] www.railtechnologymagazine.com. Accessed 1st December 2018.

Bertrand, M., Karlan, D., Mullainathan, S., Shafir, E., Zinman, J. 2010. "What's advertising content worth? Evidence from a consumer credit marketing field experiment." *Quaterly Journal of Economics*. 125(1): 263-306.

Bicchieri, C., Dimant, E., Xiao, E., 2018. Deviant or Wrong? The Effects of Norm Information on the Efficacy of Punishment. PPE Working Paper 0016, University of Pennsylvania.

Brög W., Erl, E., Ker, I., Ryle, J., Wall, R., 2009. "Evaluation of voluntary travel behaviour change: experiences from three continents." *Transport Policy*, 16(6): 281-92.

Cairns, S., Sloman, L., Newson, C., Anable, J., Kirkbride, A., Goodwin, P., 2008. "Smarter choices: assessing the potential to achieve traffic reduction using soft measures." *Transport Reviews* 28(5): 593-618.

Cialdini, R.B., Kallgren, C.A., Reno, R.R. 1991. "A focus theory of normative conduct: a theoretical refinement and re-evaluation of the role of norms in human behaviour." *Advances in Experimental Social Psychology*. Vol. 24, pp201-234. Academic Press, Inc.

Coleman, J., 1990. *Foundations of Social Theory*. Belknap Press of Harvard University Press.

Currie, G., Delbosc, A., 2011. "Transport disadvantage: a review." In G. Currie (Ed.), *New Perspectives and Methods in Transport and Social Exclusion Research* (First ed., pp. 15 - 25). Bingley UK: Emerald Group Publishing Limited.

De Ana Rodriguez, G., Seriani, S., Holloway, C., 2016. "The impact of platform edge doors on passengers boarding and alighting time and platform behaviour." Paper presented at the *Transportation Research Board* 95th annual meeting, Washington, DC, January 10-14.

Douglas, N., Henn, L., Sloan, K., 2011. "Modelling the ability of fare to spread AM peak passenger loads using rooftops." *Paper presented at the 34th Australasian Transport Research Forum*, Adelaide, Australia, 28-30 September.

Duarte, A., Garcia, C., Giannarakis, G., Limão, S., Polydoropoulou, A., Litinas, N. 2010. "New approaches in transportation planning: happiness and transport economics." *NETECONOMICS: Economic Research and Electronic Networking*. 11(1): 5-32.

Elster, J., 1988. "Economic order and social norms." *Journal of Institutional and Theoretical Economics*, 144 (2): 357-366.

Elster, J., 1989. "Social norms and economic theory." *Journal of Economic Perspectives*. 3(4): 99-117.

Fujii, S., Taniguchi, A. 2006. "Determinants of the effectiveness of travel feedback programs: a review of communicative mobility management measures for changing travel behaviour in Japan." *Transport Policy*. 13: 339-348.

Fujiyama, T., Nowers, J., Tyler, N., 2014. "The effects of the design factors of the train-platform interface on pedestrian flow rates." *In Pedestrian and Evacuation Dynamics*, edited Weidmann, U., Kirsch, U., and Schreckenberg, M., 1163-1173. Cham, Switzerland: Springer International.

Garnett, N.S., 2009. "Private Norms and Public Places". 18 Wm. & Mary Bill Rts. J. 183 (2009-2010). Available at: https://scholarship.law.nd.edu/law_faculty_scholarship/302

Goodwin, T., 2017. The Green Lane Trial Project: Change Assurance Plan, Transport for London (Unpublished).

Harris, N.G., 2006. "Train boarding and alighting rates at high passenger loads." *Journal of Advanced Transportation*. 40: 249-263.

Heuvel, van den J., 2016. "Field experiments with train stopping positions at Schipol Airport Train Station in Amsterdam, Netherlands." *Transport Research Record*, 2546(4): 24-32.

Karekla, X., Tyler, N., 2012. "Reduced dwell times resulting from train-platform improvements: the costs and benefits of improving passenger accessibility to metro trains." *Transportation Planning and Technology*. 35(5): 525-543.

Lin, T.M. and Wilson, N.H., 1992. Dwell time relationships for light rail systems. *Transportation Research Record*, (1361).

Liu, Y., Charles, P., 2013. "Spreading peak demand for urban rail transit through differential fare policy: A review of empirical evidence." *Australasian Transport research Forum 2013, Australia, Brisbane.*

Loukopoulos, P., 2007. A classification of travel demand management measures. In: Gärling, T., Steg, L., (Eds.) Threats from Car Traffic to the Quality of Urban Life: Problems, Causes and Solutions. Elsevier, Amsterdam, pp. 275-292.

Melo, P.C., Harris, N.G., Graham, D.J., Barron, A., Anderson, R.J. 2011. "Determinants of delay incident occurrence in urban metros." *Transportation Research Record: Journal of the Transportation Research Board*. 2216: 10-18.

Merton, R., 1957. *Social Theory and Social Structure*. Free Press, Glencoe, IL.

Metcalfe, R., Dolan, P., 2012. "Behavioural economics and its implications for transport." *Journal of Transport Geography*. 24: 503-511.

Moncrieff, K., 2015. Designing Passenger Information for Dwell Time to Support Thameslink High Capacity Infrastructure. Rail Human Factors Conference, Fifth Edition.

Network Rail, 2013. *Long term planning process*. Available at: https://www.networkrail.co.uk/running-the-railway/long-term-planning

Nolan, J., 2015. "Social norms and their enforcement." *The Oxford Handbook of Social Influence*, (eds.) Harkins S, Williams KD, Burger J. Oxford University Press, Oxford, England.

Olaverri-Monreal, C., Çapalar, J., Nemec, A., Zahradnik, C., 2018. Optimisation of Passenger Distribution at Metro Stations Through a Guidance System. In: Moreno-Diaz, R., Pichler, F., Quesada-Arencibia (Eds), Springer Berlin Heidelberg, pp 397-404.

Oliveira, L., Fox, C., Birrell, S., Cain, R., 2019. "Analysing passengers' behaviours when boarding trains to improve rail infrastructure and technology." *Robotics and Computer Integrated Manufacturing*. 57: 282-291.

Puong, A., 2000. Dwell time model and analysis for the MBTA red line. *Massachusetts Institute of Technology Research Memo*, pp.02139-4307.

Preston, J., 2018. *The UK Passenger Rail System: how and why is it changing?* Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/761942/Passengerrailtransport.pdf

Qu, X., Wang, S., Zhang, W., 2018. "Trial-and-error train fare design scheme for addressing boarding/alighting congestion at CBD stations." *Transportation Research Record: Journal of the Transportation Research Board*, 118: 318-335.

Rashidi, S., Ranjitkar, P. and Hadas, Y., 2014. Modeling bus dwell time with decision tree-based methods. *Transportation Research Record*, *2418*(1), pp.74-83.

RDG (Rail Delivery Group), 2016. *Rail's transformation in numbers*. Available at: https://www.raildeliverygroup.com/files/Publications/2016-11_rdg_dataset.pdf

Seriani, S. and Fernandez, R., 2015. Pedestrian traffic management of boarding and alighting in metro stations. *Transportation research part C: emerging technologies*, *53*, pp.76-92.

Sherif, M., 1936. The psychology of social norms. Oxford, England: Harper.

Sugden, R., 1987. “The economics of rights, cooperation and welfare.” *The Economic Journal*. 97(387): 751-753.

Taylor, M., 2007. “Voluntary travel behaviour change programs in Australia: the carrot rather than the stick in travel demand management.” *International Journal of Sustainable Transportation*. 1(3): 173-192.

TfL, 2018a. Mayor’s Transport Strategy. [online] https://www.london.gov.uk/sites/default/files/mayors-transport-strategy-2018.pdf. Accessed 24th February, 2020.

TfL, 2018b. The Green Lane Trial. Unpublished report.

Thoreau, R., Holloway, C., Bansal, G., Gharatya, K., Roan, T-R., Tyler, N., 2016. “Train design features affecting boarding and alighting of passengers.” Journal of Advanced Transportation. 50: 2077-2088.

Ullmann-Margalit, E., 1977. The Emergence of Norms. Oxford: Oxford University Press.

Wiggenraad, P.B.L., 2001. Alighting and Boarding Times of Passengers at Dutch Railway Stations. TRAIL Research School, Delft University of Technology, Holland.

Young H.P., 2008. “Social Norms.” In: Durlauf, S.N., Blume, L.E. (eds) The New Palgrave Dictionary of Economics. Palgrave Macmillan, London.

Appendix A1 – Methodology Details

The difference-in-differences estimator

Let Y_{tsl} represent the natural logarithm of dwell time of a train at time t, at station s, on service line l. We consider only inbound trains. Here, t is any moment in time from 6am on May the 21st to 11pm on November 30th in 2016 and 2017. We define the following variables:

$Post_t$ = 1 if t is later than 6am, September 1st, in either year (2016, or 2017), and 0 otherwise.

$D2017_t$ = 1 if t is in 2017, and 0 if t is in 2016.

$Kings_s$ = 1 if s is King’s Cross station, and 0 otherwise.

$Victoria_l$ = 1 if l is the Victoria line, and 0 otherwise.

An observation is identified as subject to the Green Lane treatment, if (and only if) all those indicator variables are equal to 1. The error term ε_{tsl} is assumed to be independent of the independent variables. Robust standard errors are used to account for heteroscedasticity. We estimate the following equations:

$$Y_t = \alpha + \beta Post_t + \gamma D2017_t + \delta(Post_t \cdot D2017_t) + \rho X_t + \varepsilon_t$$

where the sample is restricted to trains on the Victoria line at King's Cross station. X is a vector of station and service line characteristics as described in section 4.3 and contains a linear time trend. A negative δ indicates a decrease in dwell times at the southbound Victoria line at Kings Cross in 2017 compared to the same time period in the preceding year when no Green Lanes were in place. To see this, take the difference in expected dwell time conditional on X between the post- and pre-treatment in 2017 ($D2017_t$ =1) to obtain

$$\Delta_{2017} = \mathrm{E}(Y_t \,|\, Post_t = 1, X_t) - \mathrm{E}(Y_t \,|\, Post_t = 0, X_t) = \beta + \delta$$

while the same difference in 2016 would simply be β. The difference in the two differences is thus

$$\Delta_{2017} - \Delta_{2016} = \delta.$$

Note that this result depends critically on assuming the time-of-year effect to be the same in both years – β has the same value in both years. Without this assumption, the 2D estimate would only identify the joint effect of the Green Lanes (δ) and the difference in the time-of-year effects between 2017 and 2016 (e.g. $\beta_{2017} - \beta_{2016}$).

Next, we restrict the sample to observations in 2017 and on the Victoria line to estimate

$$Y_{ts} = \alpha + \beta Post_t + \gamma Kings_s + \delta(Post_t \cdot Kings_s) + \rho X_{ts} + \varepsilon_{ts}$$

By the same argument as above, for King's Cross Station we obtain

$$\Delta_{\mathrm{Kings}=1} = \mathrm{E}(Y_t \,|\, Post_t = 1, X_t) - \mathrm{E}(Y_t \,|\, Post_t = 0, X_t) = \beta + \delta$$

And for other stations

$$\Delta_{\mathrm{Kings}=0} = \mathrm{E}(Y_t \,|\, Post_t = 1, X_t) - \mathrm{E}(Y_t \,|\, Post_t = 0, X_t) = \beta$$

The resulting 2D estimate is

$$\Delta_{\mathrm{Kings}=1} - \Delta_{\mathrm{Kings}=0} = \delta$$

The last 2D estimator only considers observations in King's Cross station in 2017:

$$Y_{tl} = \alpha + \beta Post_t + \gamma Victoria_l + \delta(Post_t \cdot Victoria_l) + \rho X_{tl} + \varepsilon_{tl}$$

The change in dwell times for observations on the Victoria line is

$$\Delta_{\text{Victoria}=1} = E(Y_t \,|\, Post_t = 1, X_t) - E(Y_t \,|\, Post_t = 0, X_t) = \beta + \delta$$

The change in dwell times for observations on other service lines is

$$\Delta_{\text{Victoria}=0} = E(Y_t \,|\, Post_t = 1, X_t) - E(Y_t \,|\, Post_t = 0, X_t) = \beta$$

The resulting 2D estimate is

$$\Delta_{\text{Victoria}=1} - \Delta_{\text{Victoria}=0} = \delta$$

The triple difference estimator

The first 3D estimator considers time-of-the-year and station specific effects and is restricted to observations on the Victoria line. We can obtain a 3D estimate with a model that is saturated in its interactions of the key variables, *Post*, *Kings*, and *D2017*.

$$Y_{ts} = \alpha + \beta Post_t + \gamma Kings_s + \varsigma D2017_t + \delta(Post_t \cdot Kings_s) + \vartheta(Post_t \cdot D2017_t)$$

$$+ \iota(D2017_t \cdot Kings_s) + \eta(Post_t \cdot D2017_t \cdot Kings_s) + \rho X_{ts} + \varepsilon_{ts}$$

We obtain dwell time changes (between fall and summer) for King's Cross station, separately for 2017 and 2016:

$$\Delta_{\text{Kings},2017} = \beta + \delta + \vartheta + \eta$$

$$\Delta_{\text{Kings},2016} = \beta + \delta$$

With the resulting difference

$$\Delta_{\text{Kings},2017} - \Delta_{\text{Kings},2016} = \vartheta + \eta \qquad \text{(A1)}$$

Repeating this for other stations, we obtain

$$\Delta_{\text{other},2017} - \Delta_{\text{other},2016} = \vartheta \qquad \text{(A2)}$$

Thus, under the assumption that the year to year change in the dwell time difference between fall and summer is the same at King's Cross station as in other stations which serve the Victoria line, the difference between equations (A2) and (A1) identify the effect of the Green Lanes. The remaining two triple difference models can be constructed in analogous fashion.

Quadruple difference estimator

The 4D estimator naturally extends the 3D estimator. Now, all three scenarios need to be taken into account, and we have to distinguish observations by year, station, and service line. The model to be estimated is

$$Y_{tsl} = \alpha + \beta_0 Post_t + \beta_1 D2017_t + \beta_2 Kings_s + \beta_3 Victoria_l + \gamma_0(Post_t \cdot D2017_t) + \gamma_1(Post_t \cdot Kings_s) + \gamma_2(Post_t \cdot Victoria_l) + \gamma_3(D2017_t \cdot Kings_s) + \gamma_4(D2017_t \cdot Victoria_l) + \gamma_5(Kings_s \cdot Victoria_l) + \delta_0(Post_t \cdot D2017_t \cdot Kings_s) + \delta_1(Post_t \cdot D2017_t \cdot Victoria_l) + \delta_2(Post_t \cdot Kings_s \cdot Victoria_l) + \delta_3(D2017_t \cdot Kings_s \cdot Victoria_l) + \eta(Post_t \cdot D2017_t \cdot Kings_s \cdot Victoria_l) + \rho X_{ts} + \varepsilon_{ts}$$

We verify that η is the 4D estimator. If we considered only observations on the Victoria line, we would obtain

$$\Delta_{Kings,2017} - \Delta_{Kings,2016} = \gamma_0 + \delta_0 + \delta_1 + \eta \qquad \text{(A3)}$$

$$\Delta_{other,2017} - \Delta_{other,2016} = \gamma_0 + \delta_1 \qquad \text{(A4)}$$

And the resulting difference between (A3) and (A4) is

$$\Delta\Delta_{Victoria} = \delta_0 + \eta.$$

Repeating this now for observations which are not on the Victoria line we would obtain simply $\Delta\Delta_{not\ Victoria} = \delta_0$. Thus, the resulting quadruple difference is:

$$\Delta\Delta_{Victoria} - \Delta\Delta_{not\ Victoria} = \eta.$$

Appendix A2 – Results for complete sample period

Tables A1 to A3 correspond to tables 1 to 3 in the main text, but they are based on the entire sample, whereas the results in the main text exclude all observations which fall into the installation period of the Green Lanes. The 18th July marks the first day of the treatment period (that is, the variable Post is 1 for observations from the 18th of July to the 30th of November).

While quantitative differences to the main results exist, the general conclusions about the effectiveness of the Green Lanes hold for this alternative sample selection as well.

Table A1: Treatment effect estimates of Green Lanes

	Dwell Analysis	Delay Analysis	
Model	Effect on dwell time	Effect on dwell time	Effect on delay time
Simple difference	1.4**	4.7*	-4.3**
	(0.3)	(2.1)	(1.6)
Difference-in-differences (1)	0.2	3.9*	-7.2**
	(0.3)	(1.7)	(1.4)
Difference-in-differences (2)	-0.8**	0.4	-2.0
	(0.2)	(1.5)	(1.3)
Difference-in-differences (3)	-0.3	3.1	-5.2**
	(0.3)	(1.9)	(1.0)
Triple difference (1, 2)	-1.9**	0.4	-7.6**
	(0.3)	(2.0)	(2.0)
Triple difference (1, 3)	-5.3**	-0.3	-10.1**
	(0.4)	(2.4)	(1.5)
Triple difference (2, 3)	-1.9**	1.9	-0.4
	(0.3)	(2.1)	(1.4)
Quadruple difference (1, 2, 3)	-7.6**	-1.7	-12.4**
	(0.5)	(2.9)	(2.1)

Notes: The coefficients are percentage changes in dwell/delay times. Standard errors are in parentheses. * $p < 0.05$, ** $p < 0.01$. (1) uses the treatment platform in 2016 as control observations, (2) uses adjacent stations on the Victoria line as control observations, (3) uses other service line platforms at King's Cross station as control observations. All regressions control for *Demand*, *Lines*, *DemandPerLine*, *ServiceLevel*, *ServiceLevelDemand*, day of the week dummies, dummies for each 15-minute interval of the day, as well as a linear time trend. See also methodology section.

Table A2: Treatment effect estimates of Green Lanes - Heterogenous trends

Model	Dwell Analysis Effect on dwell time	Delay Analysis Effect on dwell time	 Effect on delay time
Difference-in-differences (1)	0.1	4.8	-9.6**
	(0.4)	(2.7)	(2.2)
Difference-in-differences (2)	0.8*	3.3	-1.9
	(0.4)	(2.5)	(2.1)
Difference-in-differences (3)	2.3**	8.1**	-6.7**
	(0.4)	(3.1)	(1.6)
Triple difference (1, 2)	-3.1**	1.1	-13.8**
	(0.5)	(3.3)	(2.7)
Triple difference (1, 3)	-4.5**	0.7	-11.1**
	(0.6)	(3.8)	(2.5)
Triple difference (2, 3)	-1.7**	2.5	-19.0**
	(0.5)	(3.4)	(2.2)
Quadruple difference (1, 2, 3)	-10.2**	-1.7	-46.9**
	(0.8)	(4.6)	(3.3)

Notes: The coefficients are percentage changes in dwell/delay times. Standard errors are in parentheses. * $p < 0.05$, ** $p < 0.01$. (1) uses the treatment platform in 2016 as control observations, (2) uses adjacent stations on the Victoria line as control observations, (3) uses other service line platforms at King's Cross station as control observations.All regressions control for Demand, Lines, DemandPerLine, ServiceLevel, ServiceLevelDemand, day of the week dummies, dummies for each 15-minute interval of the day, as well as linear time trends for each combination of treatment and control platforms. See also methodology section.

Table A3: Treatment effect estimates of Green Lanes - outbound direction

Model	Dwell Analysis Effect on dwell time	Delay Analysis Effect on dwell time	 Effect on delay time
Simple difference	0.6	4.7	-2.9*
	(0.3)	(2.6)	(1.4)
Difference-in-differences (1)	-0.3	1.6	-0.4
	(0.3)	(1.9)	(1.3)
Difference-in-differences (2)	-0.5*	-0.2	0.8
	(0.2)	(1.8)	(1.1)
Difference-in-differences (3)	0.3	3.0	-2.8**
	(0.3)	(2.1)	(1.0)
Triple difference (1, 2)	-1.4**	-1.7	-0.2
	(0.4)	(2.2)	(1.8)
Triple difference (1, 3)	-2.2**	-1.2	3.1*
	(0.4)	(2.5)	(1.6)
Triple difference (2, 3)	-1.2**	-3.0	-0.3
	(0.3)	(2.4)	(1.2)
Quadruple difference (1, 2, 3)	-3.2**	-6.9*	0.8
	(0.5)	(3.1)	(2.0)

Notes: The coefficients are percentage changes in dwell/delay times. Standard errors are in parentheses. * $p < 0.05$, ** $p < 0.01$. (1) uses the treatment platform in 2016 as control observations, (2) uses adjacent stations on the Victoria line as control observations, (3) uses other service line platforms at King's Cross station as control observations. All regressions control for Demand, Lines, DemandPerLine, ServiceLevel, ServiceLevelDemand, day of the week dummies, dummies for each 15-minute interval of the day, as well as a linear time trend. See also methodology section.