THE IMPACT OF REAL-TIME INFORMATION ON BUS RIDERSHIP IN NEW YORK CITY

Dr. Candace Brakewood (Corresponding Author)
Department of Civil Engineering
The City College of New York
160 Convent Ave, New York, NY 10031
Email: cbrakewood@ccny.cuny.edu
Phone: +1-(212)-650-5217

Dr. Gregory S. Macfarlane
Parsons Brinckerhoff
424 Fayetteville Street, Ste 1500, Raleigh, NC 27601
Email: macfarlaneg@pbworld.com
Phone: +1-(919)-836-4063

Dr. Kari Watkins
School of Civil and Environmental Engineering
Georgia Institute of Technology
790 Atlantic Drive, Atlanta, GA 30332
Email: kari.watkins@ce.gatech.edu
Phone: +1-(206)-250-4415

ABSTRACT

In the past few years, numerous mobile applications have made it possible for public transit passengers to find routes and/or learn about the expected arrival time of their transit vehicles. Though these services are widely used, their impact on overall transit ridership remains unclear. The objective of this research is to assess the effect of real-time information provided via web-enabled and mobile devices on public transit ridership. An empirical evaluation is conducted for New York City, which is the setting of a natural experiment in which a real-time bus tracking system was gradually launched on a borough-by-borough basis beginning in 2011. Panel regression techniques are used to evaluate bus ridership over a three year period, while controlling for changes in transit service, fares, local socioeconomic conditions, weather, and other factors. A fixed effects model of average weekday unlinked bus trips per month reveals an increase of approximately 118 trips per route per weekday (median increase of 1.7% of weekday route-level ridership) attributable to providing real-time information. Further refinement of the fixed effects model suggests that this ridership increase may only be occurring on larger routes; specifically, the largest quartile of routes defined by revenue miles of service realized approximately 340 additional trips per route per weekday (median increase of 2.3% per route). Although the increase in weekday route-level ridership may appear modest, on aggregate these increases exert a substantial positive effect on farebox revenue. The implications of this research are critical to decision-makers at the country's transit operators who face pressure to increase ridership under limited budgets, particularly as they seek to prioritize investments in infrastructure, service offerings, and new technologies.

KEY WORDS

Real-time information; public transit; panel regression; bus ridership

1 INTRODUCTION

Public transit plays an important role in urban transportation systems. Transit can help to combat roadway congestion, decrease gasoline consumption, and reduce carbon dioxide emissions in metropolitan areas (Schrank et al., 2012). Other benefits of transit include providing personal mobility options for those who cannot or choose not to drive (American Public Transportation Association, 2014) and supporting active mobility and its subsequent positive health impacts (Besser and Dannenberg, 2005; Litman, 2014a). Despite these benefits, transit agencies in many American cities struggle to increase (and in some cases, maintain) ridership levels as they compete with other modes of passenger transportation, particularly single-occupancy motor vehicles.

In order for public transit to be a viable option for travelers, it must be reliable, accessible, and presented in an understandable manner, among other things (Walker, 2012). These factors can potentially be improved with new customer information systems, which transit agencies are rapidly implementing. The widespread adoption of mobile devices by transit passengers has led to growing reliance on these devices and increased expectations for transportation information provided in personalized formats. Moreover, these applications are frequently more cost-effective for transit agencies than alternative methods of displaying this information, such as dynamic message signs. Consequently, the availability of web and mobile "apps" providing transit information – particularly real-time vehicle location/arrival information – has increased at an unprecedented pace over the last decade (Schweiger, 2011).

Given the rapid increase in the availability of transit apps, quantifying the impact of real-time transit information on actual travel behavior is essential for transit operators to make responsible decisions regarding the implementation of these systems and for planning agencies to properly plan for future scenarios. Because transit travel is affected by numerous factors, such as macroeconomic conditions and the weather, previous studies have had difficulty isolating changes in transit ridership due to real-time information.

This paper relies on a natural experiment that occurred in New York City beginning in 2011, when the transit agency began to gradually deploy real-time information on its buses operating in each borough of New York City on a by-borough basis. This deployment pattern enables use of regression techniques that control for unobserved heterogeneity in route-level ridership and time-dependent effects, isolating the effect of real-time information. The results of this analysis indicate that real-time information is associated with an increase of approximately 118 unlinked trips per route on an average weekday, although the ridership increase appears to be occurring primarily on the largest bus routes.

This paper proceeds as follows. First, prior research into the impacts of traveler information systems on transit passengers is presented to provide a basis for the contribution of this research. Next, the methodology for data collection and econometric analysis is discussed, followed by results and revenue implications. The final section contains a discussion of study limitations, opportunities for future research, and concluding remarks.

2 PRIOR RESEARCH

Real-time information (RTI) refers to the tracking of transit vehicle locations and/or predicted arrival times for vehicles at stops and/or stations, which is typically updated at least once per minute. One area of prior research pertains to RTI displayed on signage at stops or in stations (Hickman and Wilson, 1995; Dziekan and Kottenhoff, 2007; Kamga et al., 2013). Recently, the practice of providing RTI to transit riders via web-enabled and mobile devices has become increasingly ubiquitous (Schweiger, 2011), and a growing body of literature aims to understand the rider impacts of RTI provided via personal devices. Some of these studies have utilized simulation modeling techniques (Fries et al., 2011) and others have employed stated preference techniques (Tang and Thakuriah, 2010), in which researchers pose hypothetical scenarios to survey participants as opposed to directly observing their behavior. The following brief literature review focuses on research that evaluates actual transit rider behavior (as opposed to simulation or stated preference methods) because these studies are most likely to provide the concrete conclusions needed for decision-makers at transit agencies. Section 2.1 briefly summarizes key rider benefits of using RTI, and section 2.2 provides a detailed review of literature pertaining to changes in transit travel associated with the availability of RTI.

2.1 Prior Research on the Rider Benefits of Real-Time Information

Previous studies of transit riders using RTI have found some important benefits. First, RTI can help passengers adapt to unreliability of transit service, which was an important finding of a recent survey of current and former transit riders in the San Francisco Bay Area (Carrel et al., 2013). Second, RTI users can time their departure from their origin to minimize wait times at stops/stations; moreover, real-time information can help to reduce the perception of time while waiting at stops/stations. In Seattle, Washington, a recent survey of bus riders using RTI found that their actual wait times were almost two minutes less than those of non-users, and perceived wait times of RTI users were approximately 30% less than those who did not use RTI (Watkins et al., 2011). Other passenger benefits of RTI include increased perception of personal security and increased satisfaction with transit service (Zhang et al., 2008; Ferris et al., 2010; Gooze et al., 2013).

2.2 Prior Research on the Ridership Impacts of Real-Time Information

If RTI users can adapt to unreliable service more easily, spend less time waiting, feel safer, and/or are more satisfied with overall service, it follows that they may make more trips on the transit system, either by choosing transit over alternative modes or making trips that they would not have made otherwise. Therefore, a few recent studies have aimed to understand the impacts of RTI on transit travel.

A panel study conducted from 2006 to 2007 on the University of Maryland campus measured changes before and after the implementation of an RTI system on the university shuttle bus network (Zhang et al., 2008). Based on a fixed effects ordered probit model of individual travelers' monthly shuttle trips, the authors concluded that RTI did not significantly affect shuttle bus trip frequency. One possible explanation the authors identify is that the number of shuttle trips was measured only two weeks after an extensive marketing campaign of the new RTI system, and there may have been insufficient time for adjustments of travel behavior (Zhang et al., 2008). Another possibility is that the population under study was an academic community with potentially inelastic travel behavior; class and activity schedules may be relatively fixed, and would not therefore be substantially affected by new information.

A behavioral experiment conducted in Tampa, Florida randomly divided participants into a RTI user group and a control group without RTI (Brakewood et al., 2014). Both groups were asked to complete a survey before the study began and a second survey after a three month study period to assess changes in travel behavior. On both surveys, participants were asked to self-report the number of bus trips that they had made in the last week. The authors found that the change in trips from the before to the after survey was not significantly different between the two groups. However, the authors noted that many bus riders in the study were dependent on transit and had limited ability to increase their trips, as they were already using transit for all or a majority of their trips. Also, the study participants were recruited from among people already in the sphere of influence of the transit provider; thus, there was no opportunity to analyze the potential of RTI for attracting new riders.

Conversely, two studies of bus riders in Seattle, Washington provide some evidence that use of mobile RTI may lead to an increase in trips made on transit. In 2009, a web-based survey of over 400 RTI users asked respondents if their average number of transit trips per week changed as a result of RTI (Ferris et al., 2010). Approximately 31% of users reported increases in non-commute trips, while a smaller percentage reported increases in commute trips on transit. A follow-up web-based survey of RTI users in 2012 found similar results (Gooze et al., 2013). However, the authors identified two important caveats for these studies: the survey results were all self-reported and did not include a control group of non-RTI users.

The most relevant prior study in the context of this paper is an empirical evaluation of the real-time bus tracking system in Chicago (Tang and Thakuriah, 2012). The authors modeled average weekday route-level bus ridership for each month from 2002 until 2010, during which time Chicago's real-time vehicle tracking system was incrementally rolled out between August 2006 and May 2009. Controlling for unemployment levels, weather, gas prices, population, and transit service attributes (such as fares and frequency of service), Tang and Thakuriah (2012) showed a significant but "modest" increase of 126 average weekday trips per route attributable to RTI, which was an increase of approximately 1.8-2.2%. However, the authors identified some limitations to their study that could have contributed (favorably or otherwise) to their results:

1. **Number of Real-Time Information Interfaces**: The ways riders received information from the original RTI system changed greatly since the basic technology was implemented in 2006, which began with a simple web interface and later expanded to include smartphone applications.

2. **Technology Adoption:** RTI was only available to those who had the devices needed to access it (e.g., computers or handheld devices with internet); thus, riders who did not have these technologies could not use it. This is noteworthy in the beginning of the study period, when levels of mobile technology adoption were lower. For comparison, the Apple iPhone debuted in June 2007 (Apple Computer, 2007), and a public release of Google's Android software followed in late 2008 (Morrill, 2008); only near the end of the study period had modern smartphones achieved widespread market penetration.

3. **Awareness of Real-Time Information**: It is possible that many travelers were unaware of RTI during the period of analysis.

This leads to three noteworthy items that could be improved in future research. The quasi-experimental design used in the Chicago study would be more suitable in a transit system launching RTI under the following three conditions: (1) a simultaneous launch on multiple interfaces (i.e. website, SMS, and smartphone applications), (2) a passenger population with high levels of technology adoption (particularly mobile devices), and (3) a coordinated marketing campaign to increase awareness. These three characteristics describe another major metropolitan area that recently launched a real-time bus customer information system: New York City.

3 METHODOLOGY

This section describes the methodology used to evaluate the ridership impacts of the bus RTI system in New York City. First, some background information about the New York City transit system and the launch of the bus RTI system is presented. This is followed by the results of an on-board survey supporting the assumptions of high levels of awareness and adoption of RTI. Next, a description of the data used in the ridership analysis is provided, and finally, the specific modeling approach, panel regression, is discussed.

3.1 Background on New York City Transit

In New York City, most local bus service is operated by New York City Transit (NYCT) under the umbrella organization of the Metropolitan Transportation Authority (MTA). NYCT operates both the largest heavy rail system (the Subway) and bus system in the country with an annual ridership of approximately 2.50 billion unlinked rail trips and approximately 800 million unlinked bus trips, respectively (Neff and Dickens, 2013). The bus system, which is the focus of this analysis, includes approximately 200 fixed routes that serve the five boroughs of New York City: Manhattan, Queens, Brooklyn, Staten Island, and the Bronx (Metropolitan Transportation Authority, 2014a).

3.2 Roll-out of Real-Time Bus Information

In 2009, the MTA executive leadership team made providing RTI a strategic priority, and the agency rapidly began to roll-out real-time bus information through a platform known as Bus Time (Rojas et al., 2012). Bus Time was initially launched on a single bus route in Brooklyn (the B63) on February 1, 2011 (Metropolitan Transportation Authority, 2011). After this 'pilot' route, Bus Time was expanded on a borough-by-borough basis with a few strategic single routes between major borough releases. On January 11, 2012, Bus Time was launched on all NYCT bus routes operating in the borough of Staten Island (Metropolitan Transportation Authority, 2012a). This was followed by the availability of Bus Time on a single route in Manhattan (the M34) in April 2012 and another route in Brooklyn (the B61) in July 2012. The second borough-wide launch occurred in the Bronx on November 9, 2012, and nearly one year later, on October 7, 2013, Bus Time became available for all routes in Manhattan (Metropolitan Transportation Authority, 2013). On March 9, 2014, Bus Time was launched on all remaining bus routes in Queens and Brooklyn (Metropolitan Transportation Authority, 2014b). The gradual roll-out of Bus Time is summarized in Figure 1; notably, this launch timeline creates a natural experiment in which routes with RTI can be compared to routes without RTI during an equivalent time period, while simultaneously controlling for other factors that could affect ridership.

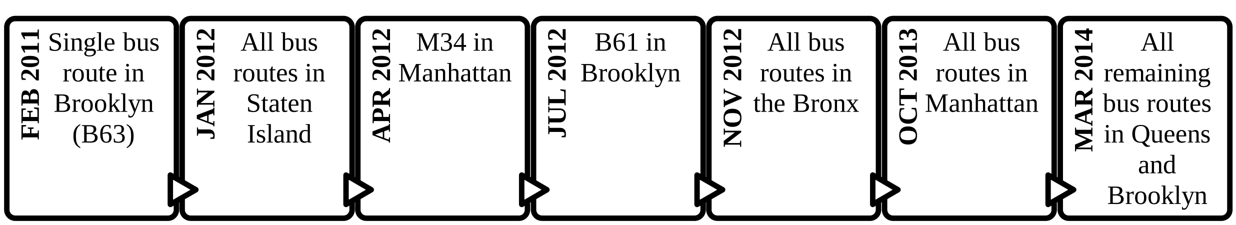

Figure 1: Timeline of the Real-time Information System Launch in New York City

3.3 Real-Time Information Interfaces, Technology Adoption, and Marketing

New York City advantageously has three characteristics that may improve upon the natural experiment previously used by Tang and Thakuriah (2012) in Chicago: (1) a simultaneous launch of RTI on multiple interfaces, (2) a transit riding population with high levels of technology adoption, particularly mobile, and (3) a coordinated marketing campaign to increase awareness of RTI.

The first characteristic – a simultaneous launch of RTI on multiple interfaces – occurred in New York City in two primary ways. First, each Bus Time launch included three MTA-managed interfaces: a desktop website, a mobile website, and SMS/text messaging (Metropolitan Transportation Authority, 2014c). Additionally, the MTA freely released the real-time bus tracking data to software developers in parallel to the launch of the MTA-managed interfaces since the initial pilot route launched in Brooklyn. This “open data” approach resulted in the availability of dozens of smartphone and web applications created by independent third party developers (Metropolitan Transportation Authority, 2014d).

The second characteristic – a transit riding population with high levels of technology adoption – is important to assure that passengers have access to the digital tools necessary to use RTI. The MTA invested significant efforts in customer research to understand levels of technology adoption by transit riders prior to the first borough-wide launch of real-time information in Staten Island. In December 2011, approximately one month prior to the launch, the MTA conducted an on-board rider survey on both local and express bus routes in Staten Island, in which a total of 1,536 paper surveys were collected. Riders were asked which technologies or devices they had used in the last 30 days, and of the 1,304 replies to this question, 62% stated that they had used text messaging, 62% had used internet on a computer, 52% had used a smartphone, and 51% had used the internet on a mobile phone. These survey results indicate that a majority of riders had one or more means to access RTI prior to its first borough-wide launch.

Third, the MTA conducted a targeted marketing campaign to increase awareness of RTI in coordination with each launch. This included posting instructions about how to use Bus Time on the poles at (almost) every bus stop (known as Guide-a-Rides) to alert riders of this new service as they wait for the bus. In summary, the combination of these three characteristics is likely to have led to high levels of RTI utilization, and consequently, may also result in ridership impacts in a relatively short time period.

3.4 Awareness and Utilization of Real-Time Information in Staten Island

To understand actual levels of rider awareness and utilization, the MTA conducted a passenger survey a few months after the first borough-wide launch of RTI in Staten Island. The agency administered an on-board paper survey for local bus routes in Staten Island in mid-May 2012 and for express routes in early June 2012.[1] A total of 1,496 surveys were collected, and the results are shown in Table 1. The three rightmost columns show the number of survey participants selecting each answer (column labeled “Count”), the percentage of respondents selecting each answer that excludes those participants who did not answer that particular question (column labeled “Responded %”), and the percentage of respondents selecting each answer including the non-respondents for that particular question (column labeled “Total %). Because this was a self-completed on-board survey, questions had differing response rates. Additionally, the sampling methodology focused on existing riders and did not try to assess how RTI may attract new riders.

[1] The data from the May/June 2012 Staten Island bus rider survey were provided by the MTA to the lead author.

Of the 1,404 respondents who answered the survey question about awareness, 73% stated that they had read about or heard about Bus Time in Staten Island. Two thirds (66%) of participants who were aware of Bus Time had used it, which equates to nearly half (44%) of all riders surveyed. Also, riders who said they had used Bus Time were asked how frequently they use it, and 55% of them stated that they use Bus Time on “most or all” of their Staten Island Bus trips (not shown in the table). In summary, only a few months after the first borough-wide launch, there were high levels of awareness and utilization of real-time information in Staten Island, and it is likely that other boroughs achieved similar levels of awareness and utilization because similar outreach campaigns were used with each launch.

Table 1: Awareness and Utilization of Real-Time Information in Staten Island

Topic	**Question**[a]	**Answers**	**Count**	**Responded %**[b]	**Total %**[b]
Awareness	Have you read or heard about MTA Bus Time in Staten Island, a new way for riders to get information about how many stops or miles aware the next bus is?	Yes	1028	73%	69%
		No	278	20%	19%
		Not sure	98	7%	7%
		Total Respondents	*1404*	*100%*	*94%*
		No Answer	92	-	6%
Utilization	Have you ever used Bus Time in Staten Island?	Yes	658	66%	44%
		No	343	34%	23%
		Total Respondents	*1001*	*100%*	*67%*
		Unaware/Not sure/No Answer	495	-	33%
		All Respondents	1496	100%	100%

[a]*Question wording is identical to the survey.* [b]*Percentages rounded to the nearest whole percent.*

3.5 Data Collection and Assembly

This section describes the data that was assembled for the ridership analysis and begins with a description of the dependent variable followed by a discussion of numerous independent variables used in the analysis.

3.5.1 Dependent Variable

The primary variable of interest in this analysis is bus ridership, and this serves as the dependent variable in the following analysis. Because real-time information was rolled out on different routes at different times (typically in the same borough), bus ridership was assessed at the route level. Average weekday route-level unlinked bus trips per month was selected as the unit of analysis because this is regularly tabulated by NYCT using data from the fare collection system and is commonly used for long term transportation planning analyses. A total of 185 bus routes (or groups of routes) operated by NYCT were considered in the analysis. Routes operated by the MTA Bus Company were not included in the analysis because the data was not available to the authors. A small number of routes were grouped due to joint scheduling/counts, which occasionally occurs for routes operating in the same corridor (e.g. M101/2/3, BX40/42, etc.).

Average weekday route-level unlinked trips were compiled for each month during a three year period from January 2011 until December 2013 (36 months), which begins shortly before the launch of real-time information on the pilot route in Brooklyn and continues through the borough-wide launches in Staten Island, the Bronx, and Manhattan. Notably, there were no major weekday service changes during the study period, though a major service cut occurred in

June 2010, which was approximately six months before the start of the study period (Grynbaum, 2010).

Figure 2 shows the average weekday bus trips per month with routes aggregated by borough. Boroughs were assigned based on the MTA route name, which is determined by the primary borough of travel. Routes that begin in B were assigned to Brooklyn, BX to the Bronx, M to Manhattan, S to Staten Island, and Q to Queens. As can be seen in Figure 2, the data exhibit strong seasonal trends, with the highest levels of ridership typically occurring in March and May and the lowest usually in August. Brooklyn has the highest overall average weekday ridership and Staten Island has the lowest. It should be noted that Queens has relatively low ridership compared to Brooklyn, Manhattan, and the Bronx because numerous bus routes in Queens are not operated by NYCT and therefore, were not considered in this analysis.

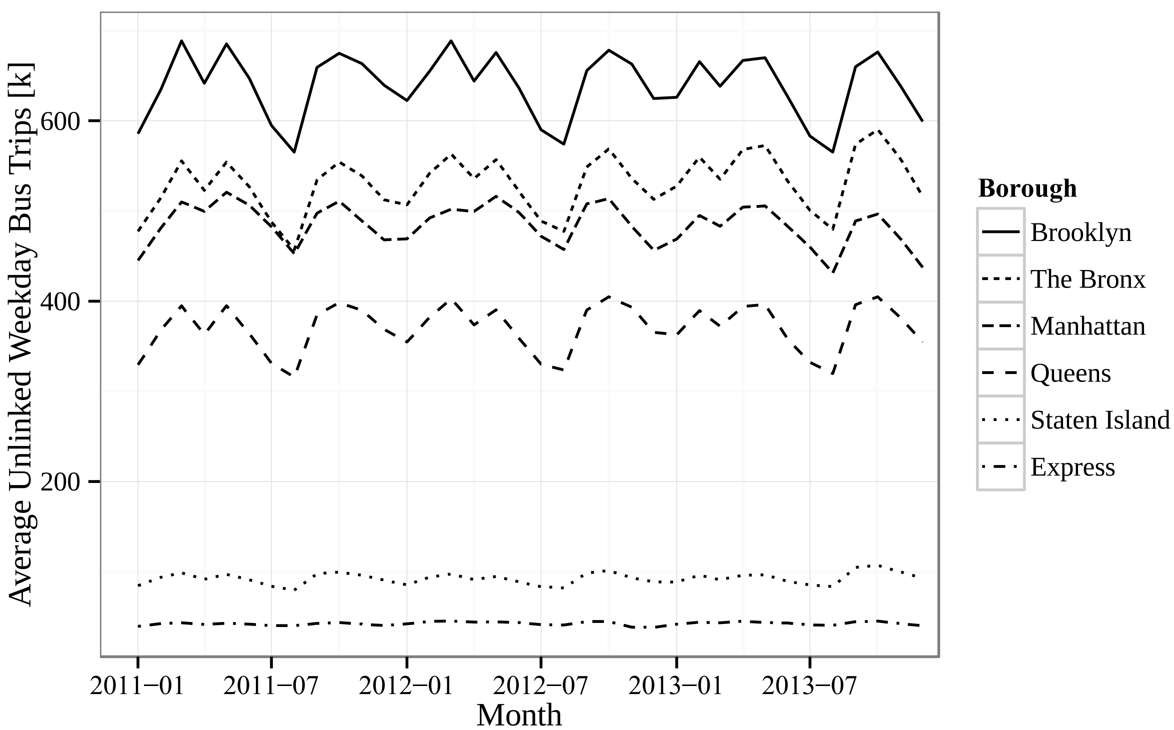

Figure 2: Average Weekday NYCT Bus Ridership by Borough

To isolate the impact of RTI on ridership, other factors that may have affected NYCT bus ridership during the three year study period were taken into account. Table 2 provides a brief description of each variable considered in the analysis. The explanatory variables were classified into two groups – transit-related factors and external factors – based on the categorization used by Tang and Thakuriah (2012). Transit-related variables were those that were largely under the control of the transit agency (such as fares and service provision), whereas external factors were mostly outside the influence of the transit provider. To the right of the categorization, a brief description, geographic unit, source of data of each variable and descriptive statistics are provided. The following paragraphs describe first the transit-related variables and then the external factors.

Table 2: Variables, Data Sources and Descriptive Statistics

Category	Variable Description	Geographic Unit	Variable Type	Data Source	Minimum	Maximum	Median
Dependent Variable	Average Weekday Unlinked Bus Trips	Route	Continuous	New York City Transit	227	62,383*	9,218
Explanatory Variables (Transit-related)	Availability of Real-Time Information	Route	Binary	MTA Press Releases	0	1	0
	Bus Average Weekday Scheduled Revenue Miles	Route	Continuous	New York City Transit	144	10,816	1,287
	Select Bus Service	Route	Binary	MTA Press Releases	0	1	0
	Bus and Rail Base Fare ($)	City	Continuous	MTA Press Releases	$2.25	$6.00	$2.25
	Rail Actual Vehicle Revenue Miles	City	Continuous	New York City Transit	25,254,835	30,340,771	28,791,719
	Rail Scheduled Vehicles Operating in Peak Service	City	Continuous	New York City Transit	5,154	5,272	5,231
Explanatory Variables (External Factors)	Availability of Bike-sharing	Borough	Binary	Citi Bike Website	0	1	0
	Borough Population (only annual estimates available; linear interpolation by month)	Borough	Continuous	US Census Bureau	469,628	2,603,360	1,625,126
	Gas Price ($/gallon)	City	Continuous	US Energy Information Administration	$3.25	$4.09	$3.75
	Unemployment Rate (percent)	City	Continuous	US Bureau of Labor Statistics	6.6%	9.4%	8.5%
	Total Monthly Snowfall (mm; measurement at Central Park)	City	Continuous	National Oceanic & Atmospheric Administration	0	913	0
	Total Monthly Precipitation (mm; measurement at Central Park)	City	Continuous	National Oceanic & Atmospheric Administration	9	481	85
	Hot Month (average temperature above 20 degrees Celsius)	City	Binary	National Oceanic & Atmospheric Administration	0	1	0
	Cold Month (average temperature below 10 degrees Celsius)	City	Binary	National Oceanic & Atmospheric Administration	0	1	0
	Hurricane Sandy	City	Binary	NYU Rudin Center Report	0	1	0

**A small number of routes were combined due to joint scheduling/counts. 62,383 is the M15 local/M15 SBS joint ridership count*.

3.5.2 Transit-Related Variables

The first transit-related explanatory variable listed in Table 2, real-time information, was modeled as a binary variable for any route with real-time information during each month in the three year study period. Initially in January 2011, no routes had real-time information, and this gradually changed until all routes in Staten Island, the Bronx, and Manhattan had real-time information. Most routes in the remaining two boroughs (Brooklyn and Queens) constitute a control group for the duration of the study period.

The second transit-related independent variable listed in Table 2 is average weekday scheduled revenue miles per bus route, and this was provided directly by the transit agency. This variable is commonly used in the transit literature (e.g., Evans, 2004) and is intended to represent the total amount of service on each bus route because it takes into account differences in frequency, span of service, and route length. Because NYCT bus schedules are modified approximately once per quarter, a total of twelve changes in scheduled revenue miles were included in the three year panel dataset.

Next, the availability of Select Bus Service (SBS) on a route was considered. SBS service includes bus rapid transit (BRT) features, such as off-board fare collection. A total of six bus routes either began as SBS or were upgraded to SBS during the three year study period, and this was modeled with a binary variable.

The literature commonly cites price as a factor that can cause changes in transit ridership (e.g., McCollom and Pratt, 2004). Hence, the base full fare is included as an independent variable. There was only one fare change during the period of analysis, which occurred in March 2013 and was an increase from $2.25 to $2.50 in the base bus and rail fare (Metropolitan Transportation Authority, 2012b).

Two variables to represent the level of service on the rail system were also included: monthly system-wide rail revenue miles and the number of vehicles operated in peak service. These variables were included because bus riders might be choosing between rail and bus service, and consequently, significant changes in the provision of rail service might result in changes in bus ridership (Tang and Thakuriah, 2012). The effect of rail might differ from the peak periods compared to the off-peak, and for that reason, the second variable pertaining to peak service was included.

3.5.3 External Factors

Numerous factors external to the transit system were also considered in the analysis. First, a new bike-sharing program, known as Citi Bike, was introduced in sections of two boroughs (Manhattan and Brooklyn) during the last six months of the study period. Because this represents a new form of transportation not previously available in New York City and one that might compete for bus passengers directly, it was hypothesized that this could influence bus ridership in areas where bike-share facilities were available. Consequently, the availability of bike-sharing was modeled as a binary variable for all bus routes in Manhattan and Brooklyn after the program commenced.

Prior research has shown that transit ridership can be dependent on changes in population (e.g., Taylor and Fink, 2003). To account for this, annual estimates of borough-level population were gathered from the US Census Bureau for 2010 and 2012, and monthly estimates were created by linear interpolation. Similarly, gas prices can influence transit demand, although the short run cross elasticity of transit demand and gas price is typically low (Litman, 2014b).

Regardless, monthly average retail gasoline price in New York City was included, and this was obtained from the US Energy Information Administration.

Research has also shown that variance in daily weather can impact transit ridership (e.g., Stover and McCormack, 2012; Singhal et al., 2014; Arana et al., 2014). Therefore, weather data were gathered from the National Oceanic and Atmospheric Administration (NOAA) for New York, NY. The measurements at Central Park were used as city-wide measurements, and temperature, precipitation, and snowfall were considered. Temperature was modeled as a binary variable to represent hot and cold months, where a hot month was defined as one with an average temperature above 20 degrees Celsius (68 degrees Fahrenheit) and a cold month was one with an average temperature below 10 degrees Celsius (50 degrees Fahrenheit). Precipitation was modeled as the total monthly precipitation in millimeters, and total monthly snowfall in millimeters was also included.

Last, a special variable was included to account for the effects of Hurricane Sandy, which occurred during the last week of October 2012 and significantly affected transit service in early November 2012. Hurricane Sandy was modeled as a binary variable for all bus routes regardless of their location for November 2012. It should be noted that the hurricane was also taken into account in the route-level bus ridership figures. On the day the hurricane struck, transit service was suspended. Approximately 24 hours after the hurricane stuck, bus service resumed and was provided free of charge (Kaufman et al., 2012). NYCT did not include these days in the average weekday ridership data, since the method of tabulating average weekday bus trips is based upon fare collection data. A few days after the hurricane (in early November 2012), bus service resumed with usual fare collection while some subway service remained suspended; these days are included in the average weekday ridership data.

3.6 Modeling Approach

Regression techniques were used to assess the relationship between route-level bus ridership and the previously discussed independent variables over the three year panel. First, ordinary least squares regression is considered, followed by fixed and random effects models, as well as other methods to account for serial correlation.

3.6.1 Ordinary Least Square Regression

Average weekday route-level bus trips per month (y) is considered as a function of the route- and time-level attributes (x) described in the previous sections. Using it as an indicator of the route ($i \in 1, \dots N$) at time ($t \in 1, \dots T$), a linear regression model was estimated by ordinary least squares (OLS),

$$y_{it} = \alpha + \beta x_{it} + \epsilon_{it} \qquad [1]$$

where β is a vector of estimated coefficients, and ϵ_{it} is an error term assumed to be independently and identically distributed (IID) with a normal distribution of mean 0 and variance σ.

The estimates resulting from this model may be inconsistent due to unobserved route-level effects (violating the IID assumption). For example, routes passing through neighborhoods of greater density or socioeconomic activity will consistently have higher ridership than routes in less dense or active areas. In this case, the error term ϵ_{it} is actually composed of two unobservable pieces, an individual effect u_i and an idiosyncratic error μ_{it}. There are two common econometric techniques, known as fixed effects and random effects, that attempt to separate u_i from ϵ_{it}, which are discussed in the following section.

3.6.2 Random Effects and Fixed Effects Regression

The random effects (RE), or random intercept, model can be written as:

$$y_{it} = \alpha_i + \beta x_{it} + \mu_{it} \qquad [2]$$

In the RE model, separate estimates of the variance in individual effects σ_u and idiosyncratic error σ_ϵ are obtained; this allows for route-level intercepts $\alpha_i = \alpha + u_i$, where u_i is distributed normally with mean 0 and variance σ_u (Wooldridge, 2009).

A potential weakness of the RE model is that estimates obtained in this manner are inconsistent if the route-level effects u_i are correlated with the route-level attributes x_{it}. For example, if high ridership routes are more or less affected by changes to fare, weather, or RTI, then RE estimates are potentially unreliable. A consistent but less efficient model in this case is the fixed effects (FE) model,

$$y_{it} - \overline{y}_i = \beta(x_{it} - \overline{x}_i) + \mu_{it} \qquad [3]$$

where the route-level unobserved effects u_i are deleted entirely from the model by demeaning the data. The model is less efficient as a result of sacrificing N degrees of freedom to estimate the individual means $\bar{y}_i$ and $\bar{x}_i$, but no interference from unobserved route-level effects remains.

To assess which of the RE or FE models is appropriate for the situation, a Hausman test can be used (Hausman and Taylor, 1981). This formally tests the differences in the coefficients of the FE and RE model. If the coefficients are sufficiently different from each other, then the RE model is inconsistent and the FE model should be used; on the other hand, if the estimates are similar then the RE model is both consistent and efficient.

3.6.3 Addressing Serial Correlation

Another potential threat to model inference is the possibility that the error terms μ_{it} are serially correlated, which commonly occurs in time series analyses. If this is the case, then hypothesis tests on the significance of the β estimates will be invalid. There are multiple ways to account for serial correlation if it exists, and three commonly used methods were considered. First, an autoregressive AR(1) term can be introduced into the error generating process:

$$\mu_{it} = \rho\mu_{i,t-1} + \nu_{it} \qquad [4]$$

where ρ is an estimated coefficient of the first order autoregressive process, and ν_{it} is a residual error term assumed IID normal. The significance of ρ can indicate the necessity of an AR structure, as a finding that $\rho = 0$ suggests that there is no autocorrelation.

A second, similar method is the introduction of an autoregressive moving average (ARMA 1,1) term, and this was used by Tang and Thakuriah (2012) in their random effects model. The error generating process in this case has the following structure:

$$\mu_{it} = \rho\mu_{i,t-1} + \nu_{it} + \theta\nu_{i,t-1} \qquad [5]$$

where θ is another estimable coefficient.

There remains a risk, however, that serially correlated residuals from an RE or FE model follow neither an AR nor an ARMA data generating process. In such a case, hypothesis tests would be invalid. Perhaps a more natural method of addressing issues of serial correlation in panel regression models is to use robust standard errors, such as those calculated using the Huber/White/sandwich estimator (StataCorp, 2013); these standard errors are robust to serial correlation within the panel, as well as heteroskedasticity.

4 RESULTS

In this section, the process to identify a statistically preferred model is discussed, and this is used to infer the relationship between RTI and observed route-level bus ridership. The estimated models are presented in Tables 3, 4, and 5.

4.1 Model Identification

First, an elementary OLS model was estimated, which is shown in the left column of Table 3. The results of a Lagrange multiplier test (Breusch and Pagan, 1980) indicated that the error term in the OLS model exhibited systematic effects, and consequently it is necessary to account for route-level effects using either a RE or FE model.

Guided by the methodology of Tang and Thakuriah (2012) in Chicago, RE models were estimated, including two that incorporate defined patterns of serial correlation. Specifically, an AR(1) and ARMA(1,1) error generation process were considered, and the model estimates are shown in Table 3. The RE models were estimated in R (R Core Team, 2013) using the package nlme (Pinheiro et al., 2014). The results of a likelihood ratio test indicate that the RE ARMA(1,1) model is preferred to the simple RE model.

The models in Table 3 seek to replicate as close as was feasible the specifications of the Chicago model, though some changes were necessitated by constraints of data availability. For example, in the Chicago model, weighted hourly frequency of bus service per route was included, but in New York City, revenue miles on each route were more readily accessible to measure transit service provision. Other variables deemed necessary to adapting the framework from Chicago to New York City, such as Hurricane Sandy and the introduction of bike-sharing, were also included. In summary, the model shown in the rightmost column of Table 3 is intended to follow that estimated for the city of Chicago in light of unavoidable constraints.

The next analysis departs considerably from the previous work by also considering the FE model. Recall from the previous econometric presentation that RE estimates are inconsistent when unobserved route-level effects are correlated with predictor variables, and that a Hausman test can be used to identify the proper model. Table 4 presents a RE model (with no adjustments for serially correlated errors) in the leftmost column and a FE model with identical variables adjacent. Estimates were obtained using the application Stata. It should be noted that some minor specification changes from the RE models shown in Table 3 were made to better fit the model to the New York City dataset, including dividing the coefficients of bus service and bike-sharing by borough.

A Hausman test on the two models in Table 4 rejects that the RE model is consistent, and therefore the FE model should be selected (Hausman $\chi^2_{29} = 306.7$, p-value < 0.001). To account for residual serial correlation, the robust standard errors (RSE) were estimated for each of the models shown in Table 4. Since the robust standard errors differ from the regular standard errors, the robust standard errors are relied on for statistical inference on the model. In summary, econometric theory and statistical tests advise that an FE model with robust standard errors (RSE) is preferred to the other models previously estimated in terms of statistical reliability and validity. Therefore, the FE models with RSEs are relied on to draw conclusions about the impact of real-time information on ridership.

The models shown in Table 4 include the availability of real-time information as a single binary variable. Because the 185 bus routes in this dataset varied greatly in terms of average weekday ridership from smaller local routes to major trunk routes, the FE model shown in Table 5 was also estimated, which divides the real-time information variable into four quartiles based on the level of bus service (in revenue miles) per route. All other variables in this model presented in Table 5 were estimated in the same manner as the FE models shown in Table 4.

Table 3: Ordinary Least Squares and Random Effects Regression Results

	OLS Estimate (SE)	**RE Estimate** (SE)	**RE AR(1) Estimate** (SE)	**RE ARMA(1,1) Estimate** (SE)
Real–Time Information	–582.17**	104.99***	59.84	70.53
	(261.23)	(35.56)	(52.86)	(49.34)
Bus Service (Revenue Miles)	5.51***	3.36***	3.92***	3.57***
	(0.06)	(0.11)	(0.13)	(0.13)
Select Bus Service	13008.62***	–473.38***	–877.71***	–682.37***
	(594.20)	(165.24)	(250.60)	(234.13)
Fare ($)	–3380.71***	–1670.53***	–2711.14***	–2880.50***
	(92.71)	(200.39)	(219.60)	(206.28)
Rail Revenue Miles (thousands)	0.04	0.06***	0.07***	0.08***
	(0.20)	(0.02)	(0.02)	(0.02)
Rail Vehicles Operated in Maximum Service	–6.14*	–2.84***	–4.23***	–5.22***
	(3.49)	(0.49)	(0.56)	(0.54)
Citi Bike	1233.57***	–471.84***	–284.69***	–278.74***
	(345.81)	(46.21)	(65.24)	(61.45)
Unemployment Rate	–227.23	–368.81***	–446.63***	–484.69***
	(241.45)	(49.69)	(51.43)	(49.45)
Population (thousands)	1.73***	2.55***	2.66***	2.51***
	(0.14)	(0.60)	(0.65)	(0.65)
Gas Price ($)	–523.63	–219.15**	–318.52***	–264.02**
	(768.91)	(104.74)	(104.12)	(104.82)
Cold Month	–150.51	–270.90***	–187.16***	–145.72***
	(483.44)	(59.25)	(43.70)	(40.52)
Hot Month	–214.94	–237.38***	–135.18**	–101.51*
	(619.71)	(76.12)	(56.45)	(54.85)
Total Monthly Snowfall (mm)	–0.76	–0.84***	–0.60***	–0.45***
	(0.69)	(0.09)	(0.07)	(0.07)
Total Monthly Precipitation (mm)	–0.06	–0.04**	–0.04***	–0.05***
	(0.14)	(0.02)	(0.01)	(0.01)
Hurricane Sandy	83.9	198.77**	44.94	–75.1
	(828.02)	(101.05)	(71.03)	(64.88)
R^2	0.62	-	-	-
Adj. R^2	0.62	-	-	-
AIC	-	108876.69	107160.24	106943.78
BIC	-	109073.88	107364.23	107154.57
Log Likelihood	-	–54409.34	–53550.12	–53440.89

**p < 0.1, **p < 0.05, ***p < 0.01*

Balanced panel with N=185 routes, T=36 months and NT=6660 total observations.

Monthly dummy variables are shown in the appendix.

Table 4: Single Real–Time Information Variable Regression Results

	Random Effects Estimate (SE)	Random Effects Estimate (Robust SE)	Fixed Effects Estimate (SE)	Fixed Effects Estimate (Robust SE)
Real–Time Information	104.954		118.278	
	(35.760)***	(52.029)**	(35.162)***	(52.695)**
Bus Service in Brooklyn	5.804		5.381	
	(0.188)***	(0.543)***	(0.241)***	(0.693)***
Bus Service in Bronx	6.059		5.073	
	(0.227)***	(0.865)***	(0.263)***	(0.935)***
Bus Service in Manhattan	5.819		3.051	
	(0.264)***	(1.088)***	(0.374)***	(1.227)**
Bus Service in Queens	3.127		2.765	
	(0.159)***	(0.926)***	(0.179)***	(1.275)**
Bus Service in Staten Island	0.574		0.212	
	(0.183)***	(0.254)**	–0.238	–0.301
Select Bus Service	–331.617		–262.039	
	(166.210)**	–443.946	–165.009	–461.757
Fare ($)	–1,030.88		–862.884	
	(164.240)***	(103.282)***	(184.457)***	(121.641)***
Rail Revenue Miles (thousands)	0.079		0.072	
	(0.021)***	(0.009)***	(0.021)***	(0.008)***
Rail Vehicles in Maximum Service	–2.925		–2.566	
	(0.452)***	(0.428)***	(0.453)***	(0.398)***
Citi Bike in Manhattan	–467.602		–556.237	
	(62.827)***	(126.536)***	(62.135)***	(143.921)***
Citi Bike in Brooklyn	–376.546		–375.308	
	(54.936)***	(97.277)***	(53.857)***	(96.701)***
Unemployment Rate	–275.806		–243.379	
	(45.289)***	(41.964)***	(48.215)***	(40.208)***
Cold Month	–249.481		–249.223	
	(58.040)***	(30.536)***	(56.868)***	(30.778)***
Hot Month	–258.168		–246.906	
	(75.470)***	(38.447)***	(73.991)***	(35.622)***
Total Monthly Snowfall (mm)	–0.833		–0.819	
	(0.081)***	(0.071)***	(0.079)***	(0.070)***
Total Monthly Precipitation (mm)	–0.387		–0.366	
	(0.158)**	(0.063)***	(0.155)**	(0.060)***
Hurricane Sandy	212.891		206.319	
	(100.157)**	(51.822)***	(98.172)**	(51.793)***
σ_u	3569.71		6425.35	
σ_μ	758.52		758.52	
R^2	0.47		0.47	

**p<0.1; **p<0.05; ***p<0.01.*
Balanced panel with N =185 routes, T =36 months and NT =6660 total observations.
Monthly dummy variables shown in the appendix.
Huber-White robust standard error.

Table 5: Quartiles of Bus Service Real-Time Information Variable Regression Results

	Fixed Effects Estimate (SE)	Fixed Effects Estimate (Robust SE)
Real-Time Information on Small Routes (Q1)	16.256 (61.568)	(62.551)
Real-Time Information on Smaller Medium Routes (Q2)	147.101 (61.415)**	(106.412)
Real-Time Information on Larger Medium Routes (Q3)	–35.114 (64.971)	(106.778)
Real-Time Information on Large Routes (Q4)	340.466 (63.655)***	(124.803)***
Bus Service in Brooklyn	5.376 (0.240)***	(0.693)***
Bus Service in Bronx	5.017 (0.263)***	(0.945)***
Bus Service in Manhattan	3.153 (0.375)***	(1.229)**
Bus Service in Queens	2.762 (0.179)***	(1.274)**
Bus Service in Staten Island	0.03 (0.243)	(0.329)
Select Bus Service	–326.825 (165.544)**	(458.593)
Fare ($)	–868.031 (184.201)***	(123.463)***
Rail Revenue Miles (thousands)	0.073 (0.021)***	(0.008)***
Rail Vehicles in Maximum Service	–2.564 (0.453)***	(0.393)***
Citi Bike in Manhattan	–535.102 (62.646)***	(152.800)***
Citi Bike in Brooklyn	–375.586 (53.781)***	(96.759)***
Unemployment Rate	–244.935 (48.153)***	(40.397)***
Cold Month	–247.74 (56.788)***	(30.635)***
Hot Month	–245.322 (73.890)***	(35.529)***
Total Monthly Snowfall (mm)	–0.82 (0.079)***	(0.070)***
Total Monthly Precipitation (mm)	–0.366 (0.155)**	(0.061)***
Hurricane Sandy	204.454 (98.027)**	(51.790)***
σ_u	6393.18	
σ_μ	757.37	
R^2	0.47	

** p<0.1; ** p<0.05; *** p<0.01;*
NT=6660 total observations. Monthly dummy controls in appendix. Huber-White robust standard error.

4.2 Model Inference

This section provides interpretation of the models presented in Tables 4 and 5. First, the ridership impact of the variable of interest, real-time information, is discussed. This is followed by inferences from the other transit-related independent variables and finally the ridership impacts of factors external to the transit system.

4.2.1 Real-Time Information

As shown in Tables 4 and 5, the variable of interest, real-time information, was significant to some degree in both robust fixed effects models. The coefficient of the real-time information variable can be interpreted as the number of additional bus trips per route on an average weekday attributable to the availability of real-time information. The coefficient of 118.278 in the single variable FE model indicates that real-time information yields, on average, an increase of approximately 118 daily trips on routes where real-time information was available, which is a median increase of 1.7% of route-level ridership.

However, the quartile model's robust standard errors reveal that real-time information only caused a significant increase in ridership on large routes, but that the improvement was larger than the single variable model indicated. On the largest quartile of routes (Q4, defined as having more than 1,900 revenue miles of service on an average weekday), real-time information increased ridership, on average, by about 340 trips per weekday, which represents a median increase of 2.3% of route-level ridership on the largest routes. For the three smaller quartiles (Q1, Q2, Q3), the coefficients for the real-time information variables were not significant. In summary, this model suggests that real-time information is only affecting ridership on the largest routes, which could be occurring for various reasons. One possible explanation is that routes with lots of service may see a larger change because the existing level of service is highest, so they are more likely to attract "choice" trips (such as non-commute trips). Another explanation may be that the ridership numbers are simply high enough to actually realize a quantifiable change; on small routes, a 1-2% change may only be a handful of trips per day, which may be lost to measurement error or overcome by statistical noise.

4.2.2 Other Transit-Related Variables

As can be seen in Tables 4 and 5, most transit-related independent variables were significant in the FE model. The level of bus service per route was significant, and it can be interpreted as the change in average weekday ridership resulting from an increase in revenue miles of service. The coefficients for bus service per route were separated into five different variables based on the borough of each route, and these coefficients indicate significantly different effects on ridership by borough. Staten Island had the only insignificant coefficient in the FE models. This may be because it has the lowest current availability of transit service; therefore, changing the level of service may have little impact on ridership in this more automobile-dependent borough.

The dummy variable for Select Bus Service (SBS) was not significant in the FE models when robust standard errors are observed. It should be noted that SBS routes were modeled as having joint ridership with their corresponding local route (e.g. the B44 and B44 SBS were modeled as a single route) due to data constraints, and this may have been one reason why there was little predicted impact on ridership.

The coefficient for the fares variable was significant. The value of the coefficient (–862.884 in the single variable model and –868.031 in the quartile model) can be interpreted as the change in average weekday route-level ridership associated with a one dollar increase in fares.

The two variables representing system-wide rail service were both significant. The total number of rail revenue miles (in thousands) operated per month had a coefficient of approximately 0.07, and this positive value suggests that as the level of overall rail service increases, bus ridership increases. Perhaps this can be interpreted as overall rail service having a complementary relationship with bus service; for example, as rail service increases, travelers in New York City become more reliant on transit, and consequently, increase both their rail and bus trips. On the other hand, the variable for system-wide peak rail service, which was vehicles operated in peak service, had a negative coefficient of approximately –2.56. This suggests that increasing rail service in the peak hour may decrease bus ridership. This substitution effect may be because commuters choose rail service over bus in peak periods.

4.2.3 External Factors

For the external factors, the Citi Bike bike-sharing program had a significant, negative effect on route-level bus ridership. The availability of bike-sharing may have decreased route-level bus ridership by over 500 trips per route in Manhattan – which has more bike-sharing stations – and approximately 375 trips per route in Brooklyn on an average weekday. The decrease may be because bike-sharing provides an alternative mode of transportation to bus service, particularly for short trips that might be made on local bus routes. However, the magnitude of this coefficient appears to be unrealistically large. Performing a back of the envelop calculation to assess if all NYCT bus routes in Manhattan and Brooklyn experienced this level of ridership decrease reveals that a very large percentage (almost all) of Citi Bike's ridership on an average weekday in 2013 would be from former bus riders. Therefore, further study is recommended to better understand the complex relationship between bus ridership and bike-sharing.

Three commonly used socioeconomic variables were included in the analysis: unemployment rate, population and gas prices. The unemployment rate had a significant negative effect on bus ridership. Both models suggest that as unemployment rate increases 1%, route-level bus ridership decreases by approximately 244 trips on an average weekday. This aligns with previous research showing that decreasing employment rates can have a negative impact on transit ridership (e.g., Taylor and Fink, 2003). The two other socioeconomic variables, gas prices and population, did not have a significant impact on bus ridership in the fixed effects model results, and consequently, they were removed from the final specification. Prior research has shown that the cross price elasticity of gas prices and transit ridership in the short run is inelastic (Litman, 2014b), so it is unsurprising that this variable was insignificant in the model. Regarding population, the data available were not at a granular level (only annual estimates by borough were available), and if there were more accurate reflections of population changes, this could have a more substantial impact on ridership.

Numerous weather variables were included in the model. Both cold and hot temperatures appear to have caused declines in ridership, with a decrease of approximately 240 to 250 trips per route on an average weekday if the month were either cold or hot. Perhaps this is because transit riders forgo unnecessary trips if the weather is particularly hot or cold, or they instead use other modes (such as a taxi or private automobile) to ensure that the entire trip was air conditioned or heated. Both total monthly snowfall and total monthly precipitation had a negative impact of ridership, which aligns with previous literature. The last weather variable, Hurricane Sandy, had a significant positive coefficient. The two models indicate that the occurrence of the hurricane increased route-level bus ridership by approximately 205 to 206 trips per route on average weekdays in November 2012. This is likely because sections of the rail system remained shut down in the immediate aftermath of the disaster, and transit riders instead used the bus system to travel (Kaufman et al., 2012).

Monthly dummy variables were included in the model to control for seasonal trends following common econometric practice and the prior study of Chicago (Tang and Thakuriah, 2012). The coefficients for these variables are shown in the appendix.

Finally, the goodness-of-fit across all models is comparable, as shown by the similar R-squared values.

5 REVENUE IMPLICATIONS

Next, the farebox revenue impacts of the ridership increase found from the regression models was assessed. For this analysis, NYCT provided the authors with the average fare per bus trip for each month over the three year study period. Because of the large difference between local and express bus fares, the average fare per trip per month was divided into average monthly local bus fare and average monthly express bus fare by NYCT. Average fares were utilized for this analysis because they take into account the differences in price between various fare types, such as full fares, period passes, and discounted fares for students.

In the previous section, the average weekday increase in bus ridership per route associated with the availability of real-time information was estimated from the regression models and, based on the results of two different models, found to be either 118 trips per route on all routes or 340 trips per route on the largest quartile of bus routes. If it is assumed that each bus route with real-time information experienced this weekday increase each day in which real-time information was available, then the total weekday increase in bus trips can be calculated for the three year study period. Moreover, multiplying this increase in route-level ridership by the average fare per trip for each route (either local or express) and summing the results for all routes with real-time information provides a simple approximation for the associated increase in farebox revenue over the three year period.

The results of this revenue analysis are shown in Figure 3. The top part of the figure shows the estimated total additional weekday farebox revenue per month for all bus routes (both local and express) over the three year study period for both the linear and quartile models. As can be seen in this section of the graphic, this ranges from zero dollars during January 2011 prior to the availability of RTI to approximately $400,000 in additional weekday farebox revenue in for December 2013 when three boroughs had RTI available. The middle part of Figure 3 shows the revenue implications for both models (linear and quartile) for express bus routes only, which have a significantly higher average fares (ranging from $4.47 to $5.07 over the three year study period). The bottom section of Figure 3 displays the revenue results for local bus routes only (average fare ranging from $1.19 to $1.47 over the study period). Summing the revenue from all months in 2011, 2012, and 2013 results in a weekday farebox revenue total of approximately $6.3 million attributed to additional trips from real-time information based on the results of the first (linear) regression model. The total weekday farebox revenue was estimated to be approximately $5.6 million using the results of the second regression model with real-time information modelled by quartile.

There are four important caveats about this revenue analysis. First, during the months when real-time information was first launched on each bus route, it was assumed that real-time information was available for all weekdays during that month. However, not all of the launches occurred on the first day of the month, so this simplifying assumption may have led to a slight overestimate of the revenue impacts. Second, this analysis only includes weekdays during the three year study period and does not consider potential farebox revenue increases for weekend ridership because the regression analysis did not assess weekend ridership. This could cause the total revenue impacts to be underestimated, since it was likely that weekend ridership also experienced an increase in trips as real-time information became available. Third, this revenue analysis uses monthly average fare that includes period passes (such as unlimited monthly passes). However, it is unknown to what extent the additional bus trips attributable to real-time information are being made by riders paying the full fare or by riders who purchased a period pass (in which case, the marginal revenue of an additional ride would be zero). Last, it is important to note that this period of analysis ends when real-time information was only available in three of the five boroughs that comprise New York City. In 2014, real-time information was launched in the remaining two boroughs, and if the revenue model was extended through the end of 2014 to include all routes having real-time information, it is likely that the system-wide farebox revenue impacts would be significantly larger. Moreover, extending the analyses of ridership and revenue through the full, city-wide deployment period would allow for comparison of total revenue with the costs to deploy real-time information, which was procured on a system-wide basis.

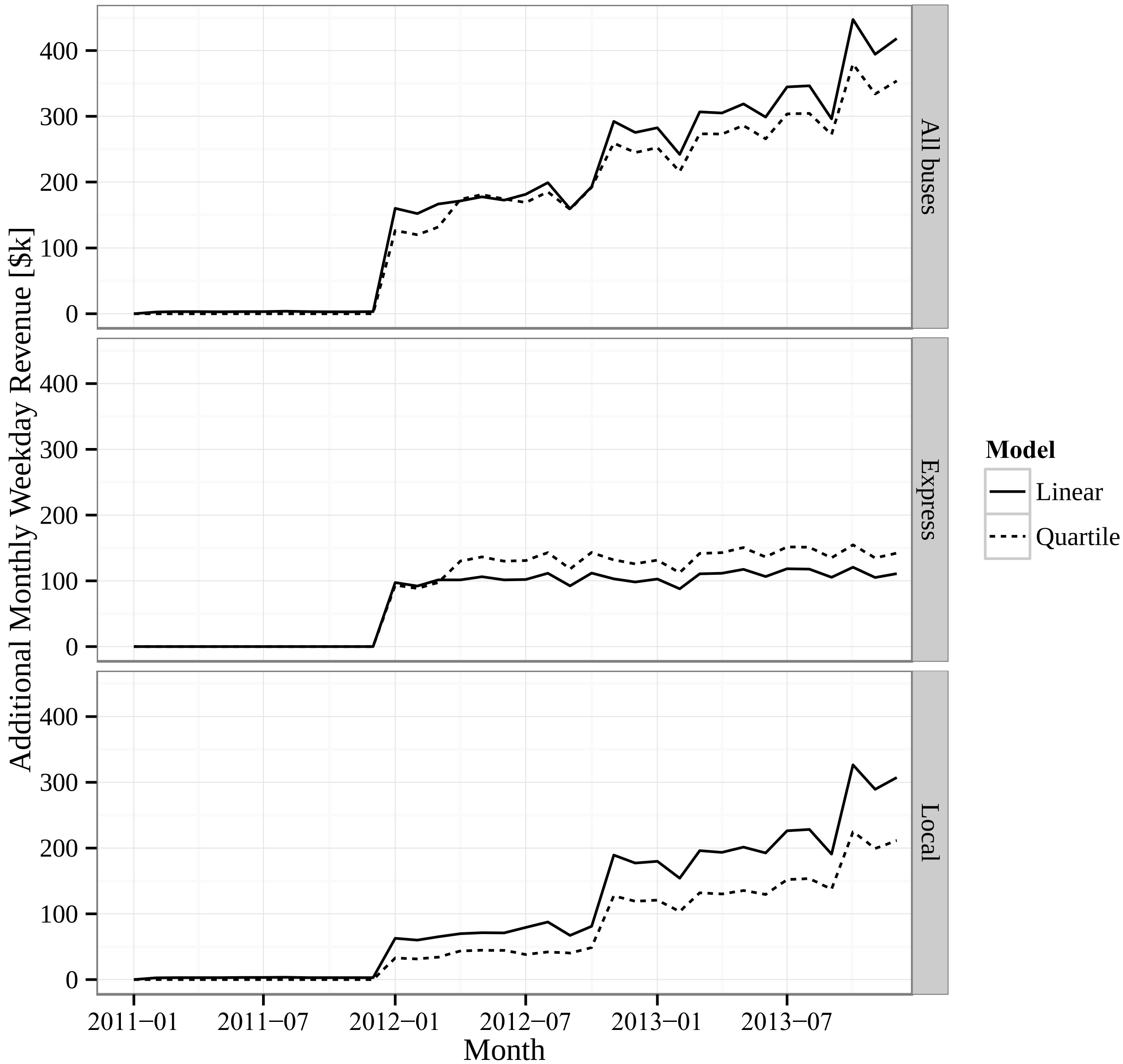

Figure 3: Additional Estimated Weekday Farebox Revenue per Month Attributable to RTI

6 AREAS FOR IMPROVEMENT AND FUTURE RESEARCH

There are a number of notable limitations to this study. One of the most challenging aspects of this research design was controlling for all of the factors that affected route-level bus ridership during the three year study period. For example, minor changes to the transportation network in New York City (e.g. road closures, bridge repairs, etc.) could have influenced the level of bus ridership but were not accounted for in the regression models.

Additionally, numerous independent variables in the panel regression models were represented with binary variables. This assumes that the variable has the same "effect" on ridership during any month that the variable was present. For example, Hurricane Sandy was modeled as affecting transit service in November 2012; however, there could have been lingering effects that were not captured in the model. Similarly, real-time information was modeled as a dummy variable, but it also could have had varying effects on ridership over time. More sophisticated fading effects models could aim to capture this in future research.

An interesting avenue for future research that emerged from the regression models pertains to the impact of bike-sharing programs on public transit ridership. In this analysis, the availability of bike-sharing was simply modeled as a binary variable, despite varying levels of bike-sharing service along bus routes (in terms of station location and number of bikes), and the magnitude of the impact of the bike-sharing program on bus ridership appears to be unrealistically large. Further research in this area is recommended, and additional studies could also evaluate the impacts on rail ridership, which may differ from that on bus service.

In terms of the overall modeling approach, there could be opportunities to utilize more sophisticated emerging techniques that consider both temporal and spatial autocorrelation. Routes that intersect or parallel each other may see their ridership counts move together as a result of transferring passengers or unobserved changes in local activity patterns. Similarly, routes that have significant sections paralleling or intersecting the subway network may see similar ridership changes that were not accounted for in this modeling framework. Controlling for these endogenous or unobserved effects will be an important challenge for future research.

This analysis focused on the overall impacts of bus real-time information on route-level weekday bus ridership, but there are many areas for additional refinement in the future. For example, future research could compare the ridership impacts of real-time information on weekdays with weekends, since weekend travel typically includes more discretionary trips. In New York City, numerous bus routes had substantial changes in weekend service as part of the MTA's 2012-2013 Service Enhancement Package, which makes this case study less suited for an analysis of weekend ridership (Metropolitan Transportation Authority, 2015). Another possible refinement is segmenting ridership impacts between high and low frequency routes or peak and off-peak periods. Additionally, expansions to understand the impact of real-time information on train ridership could be conducted, since real-time information also became available for some of the rail system during the study period (Mann, 2012).

Last, the study period for the ridership and revenue analyses concluded when real-time information was only available on bus routes in three of the five boroughs that comprise New York City. Extending the study period through 2014 so that real-time information is available on all NYCT bus routes would likely significantly increase the results of the revenue analysis. Then, the total farebox revenue attributable to real-time information could be compared to the system-wide costs of deploying real-time information in New York City.

7 CONCLUSIONS

In this study, an empirical analysis of the ridership impacts of real-time bus information in New York City was conducted. Panel regression techniques were used to evaluate bus ridership over a three year period, while controlling for changes in transit service, fares, local socioeconomic conditions, weather, and other factors. Two fixed effects models with robust standard errors were selected for final presentation. The first model, which included real-time information as a single binary variable, showed an average increase of approximately 118 trips per route per weekday (median increase of 1.7% of weekday route-level ridership) attributable to providing real-time information. The second model, which divided the real-time information coefficient based on quartiles of bus service per route, suggests that the ridership increase occurred on the largest routes, which have 1900 revenue miles or more of average weekday service. Specifically, the model implied that real-time information increased ridership by about 340 trips per weekday on the largest quartile of routes, which is a median increase of 2.3% of route-level ridership on these routes.

Although both models present plausible results, the second model may be preferable for two reasons. One possible explanation why the largest routes experience a significant increase in ridership is that they may be more likely to attract “choice” trips (such as trips to go shopping or to recreational activities). For example, when a traveler is considering taking a bus trip versus an alternative mode, checking real-time information for the bus routes with the highest service levels may reveal that a vehicle is only a few minutes away, and consequently, the traveler chooses to take that extra trip on the bus. On bus routes with lower levels of service, the traveler may be presented with the information that he or she would have to wait for a longer period of time, and in that situation, the traveler may choose an alternative mode or forgo the unnecessary trip. An alternative explanation may be that the ridership numbers are simply high enough to realize a quantifiable change; on small routes, a change less than 2% may only be a handful of trips per day, which may escape data capture or significance in the model.

While the second model presents a somewhat more plausible explanation of what is occurring in the real world, the striking similarity that the first model (with a single real-time information variable) has with the results of the Chicago study should be noted. The same unit of analysis for the dependent variable in the regression model (monthly average weekday bus trips per route) was utilized, which allows for direct comparison between this model and the Chicago model. Tang and Thakuriah (2012) found a significant increase of 126 average weekday trips per route (approximately 1.8-2.2% of route-level ridership) attributable to RTI. The single real-time information variable fixed effects model showed an average increase of approximately 118 trips per route per weekday (median increase of 1.7% of weekday route-level ridership). While a few limitations of the natural experiment in Chicago were previously noted, this study of New York City also had limitations; for example, the study period was only three years, and extending the panel – particularly to include the launch of real-time information in the remaining two boroughs – could potentially impact the final results. Perhaps the similarity in these findings, despite limitations in each of the studies, suggests that bus ridership may increase one or two percent (holding all else equal) when passengers are provided with real-time information via web-enabled and mobile devices. In light of the finding regarding greater impacts on bus routes with high levels of service, the potential generalization of this result could be limited to large bus systems, since NYCT and the CTA are the first and third largest bus systems, respectively, in the country based on unlinked passenger trips (Neff and Dickens, 2013). These results, concurrent with the previous findings in Chicago, suggest that investments in customer information systems have had a significant impact on bus ridership levels, particularly for two of the country's largest bus systems. Therefore, this research has immediate implications for leaders in the transit industry making important decisions on how to improve public transportation systems, particularly those agencies that face pressure to increase ridership.

8 ACKNOWLEDGEMENTS

This research project was funded by a US DOT Eisenhower Graduate Fellowship. The authors are very grateful to the Metropolitan Transportation Authority for their support of this project, particularly Peter Harris, Mike Frumin, Tony Laidig, Sunil Nair, Rob Hickey, Bill Amarosa, and Mike Chubak.

9 REFERENCES

1. American Public Transportation Association. (2014). Facts at a Glance. Retrieved January 30, 2014, from http://www.publictransportation.org/news/facts/Pages/default.aspx

2. Arana, P., Cabezudo, S., & Peñalba, M. (2014). Influence of Weather Conditions on Transit Ridership: A Statistical Study using Data from Smartcards. *Transportation Research Part A: Policy and Practice*, 59, 1–12.

3. Apple Computer (2007) Apple Reinvents the Phone with iPhone. Apple Computer, press release 9 Jan. 2007. Retrieved on 3 June, 2014, from https://www.apple.com/pr/library/2007/01/09Apple-Reinvents-the-Phone-with-iPhone.html.

4. Besser, L. M., & Dannenberg, A. L. (2005). Walking to Public Transit Steps to Help Meet Physical Activity Recommendations. *American Journal of Preventive Medicine*, 29(4).

5. Breusch, T. S., & Pagan, A. R. (1980). The Lagrange Multiplier Test and its Applications to Model Specification in Econometrics. *The Review of Economic Studies*, 239-253.

6. Brakewood, C., Barbeau, S., & Watkins, K. (2014). An Experiment Evaluating the Impacts of Real-Time Transit Information on Bus Riders in Tampa, Florida. *Transportation Research Part A: Policy and Practice*, 69C, 409-422.

7. Carrel, A., Halvorsen, A., & Walker, J. L. (2013). Passengers' Perception of and Behavioral Adaptation to Unreliability in Public Transportation. *Transportation Research Record: Journal of the Transportation Research Board*, 2351, 153–162.

8. Dziekan, K., & Kottenhoff, K. (2007). Dynamic at-stop real-time information displays for public transport: effects on customers. *Transportation Research Part A: Policy and Practice*, 41(6), 489–501.

9. Evans, J. E. (2004). TCRP Report 95: Traveler Response to Transportation System Change. Chapter 9 – Transit Scheduling and Frequency. Transit Cooperative Research Program of the Transportation Research Board, Washington, D.C.

10. Ferris, B., Watkins, K., & Borning, A. (2010). OneBusAway: Results from Providing Real-Time Arrival Information for Public Transit. *Proceedings: CHI*, 1807–1816.

11. Fries, R. N., Dunning, A. E., & Chowdhury, M. A. (2011). University Traveler Value of Potential Real-Time Transit Information. *Journal of Public Transportation*, 14(2), 29–50.

12. Gooze, A., Watkins, K. E., & Borning, A. (2013). Benefits of Real-Time Transit Information and Impacts of Data Accuracy on Rider Experience. *Transportation Research Record: Journal of the Transportation Research Board*, 2351, 95–103.

13. Grynbaum, Michael M. (2010). Transit Agency Approves Cuts, and More Bad News Looms. *New York Times*. New York, NY.

14. Hausman, J. A. and Taylor, W. E. (1981) Panel Data and Unobservable Individual Effects. *Econometrica*, 49(6), 1377–1398.

15. Hickman, M. & Wilson, N.H.M. (1995). Passenger Travel Time and Path Choice Implications of Real-Time Transit Information. *Transportation Research Part C: Emerging Technologies*, 3(4), 211-226.

16. Kamga, C., Yazici, M. A. & Singhal, A. (2013). Implementation of Interactive Transit Information Kiosks at New York City Transit Facilities: Analysis of User Utilization and Lessons Learned. *Transportation Research Part C: Emerging Technologies*, 35, 218-231.

17. Kaufman, S., Qing, C., Levenson, N., & Hanson, M. (2012). Transportation During and After Hurricane Sandy. New York University Report, New York, NY.

18. Litman, T. (2014a). Evaluating Public Transportation Health Benefits. Victoria Transport Policy Institute. Retrieved January 21, 2014, from http://www.vtpi.org/tran_health.pdf

16. Litman, T. (2014b). Transit Price Elasticities and Cross-Elasticities. Victoria Transport Policy Institute. Retrieved June 26, 2014, from http://www.vtpi.org/tranelas.pdf

20. Mann, T. (2012). Next Stop for New York Subway: An App to Track Trains. *Wall Street Journal*. New York, NY. Retrieved January 30, 2014 from http://online.wsj.com/news/articles/SB10001424127887323984704578205870642642436

21. McCollom, B. E., & Pratt, R. H. (2004). TCRP Report 95: Traveler Response to Transportation System Changes. Chapter 12 - Transit Pricing and Fares. Transit Cooperative Research Program of the Transportation Research Board, Washington, D.C.

22. Metropolitan Transportation Authority. (2011). MTA Bus Time Offers Real-Time Bus Location Information for B63 Customers. Retrieved March 9, 2014, from http://new.mta.info/press-release/nyc-transit/mta-bus-time-offers-real-time-bus-location-information-b63-customers

23. Metropolitan Transportation Authority. (2012a). MTA Launches “Bus Time” on Staten Island. Retrieved March 9, 2014, from http://new.mta.info/press-release/mta-headquarters/mta-launches-bus-time-staten-island

24. Metropolitan Transportation Authority. (2012b). MTA Adopts Fare Increases for 2013. Retrieved October 15, 2014 from http://www.mta.info/news/2012/12/19/mta-adopts-fare-increases-2013

25. Metropolitan Transportation Authority. (2013). MTA Bus Time Arrives in Manhattan. Retrieved March 9, 2014, from http://new.mta.info/news/2013/10/07/mta-bus-time-arrives-manhattan

26. Metropolitan Transportation Authority. (2014a). Schedules. Retrieved March 9, 2014, from http://web.mta.info/schedules

27. Metropolitan Transportation Authority. (2014b). MTA Real-Time Bus Tracking Arriving in Brooklyn and Queens on March 9. Retrieved March 9, 2014, from http://new.mta.info/press-release/nyc-transit/mta-real-time-bus-tracking-arriving-brooklyn-and-queens-march-9

28. Metropolitan Transportation Authority. (2014c). MTA Bus Time. Retrieved March 9, 2014, from http://bustime.mta.info

29. Metropolitan Transportation Authority. (2014d). App Gallery. Retrieved May 14, 2014, from http://web.mta.info/apps

30. Metropolitan Transportation Authority. (2015). 2012 - 2013 Service Enhancement Package. Retrieved January 24, 2015, from http://web.mta.info/nyct/service/2012_2013BusServiceEnhancements.htm

31. Morrill, D. (2008) Announcing the Android 1.0 SDK, release 1 Android Developers Blog. Retrieved June 3, 2014, from http://android-developers.blogspot.com/2008/09/announcing-android-10-sdk-release-1.html

32. Neff, J., & Dickens, M. (2013). Public Transportation Fact Book. American Public Transportation Association, Washington, DC. p. 25.

33. Pinheiro, J., Bates, D., DebRoy, S., & Sarkar, D. (2014). Linear and Nonlinear Mixed Effects Models. Retrieved from http://cran.r-project.org/package=nlme

34. R Core Team (2013). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

35. Rojas, B. F. M., Weil, D., & Graham, M. (2012). Transit Transparency: Effective Disclosure through Open Data. Cambridge, MA. Retrieved from http://www.transparencypolicy.net/assets/FINAL_UTC_TransitTransparency_8 28 2012.pdf

36. Schrank, D., Eisele, B., & Lomax, T. (2012). Urban Mobility Report. Texas A&M.

37. Schweiger, C. (2011). TCRP Synthesis 91: Use and Deployment of Mobile Device Technology for Real-Time Transit Information. Transit Cooperative Research Program of the Transportation Research Board, Washington, D.C.

38. Singhal, A., Kamga, C., and Yazici, A. (2014). Impact of Weather on Urban Transit Ridership. *Transportation Research Part A: Policy and Practice*, A69, 379–391.

39. StataCorp. (2013). Xtreg Manual. College Station, TX. Retrieved from http://www.stata.com/manuals13/xtxtreg.pdf

40. Stover, V., & McCormack, E. (2012). The Impact of Weather on Bus Ridership in Pierce County, Washington. *Journal of Public Transportation*, 15(1), 95–110.

41. Tang, L., & Thakuriah, P. (2010). Will the Psychological Effects of Real-time Transit Information Systems Lead to Ridership Gain? Proceedings at the Transportation Research Board 90th Annual Meeting.

42. Tang, L., & Thakuriah, P. (2012). Ridership Effects of Real-time Bus Information System: A Case Study in the City of Chicago. *Transportation Research Part C: Emerging Technologies*, 22, 146–161.

43. Taylor, B., & Fink, C. (2003). The Factors Influencing Transit Ridership: A Review and Analysis of the Ridership Literature. Working Paper. Los Angeles, CA.

44. Walker, J. (2012). *Human Transit.* Washington, D.C.: Island Press.

45. Watkins, K. E., Ferris, B., Borning, A., Rutherford, G. S., & Layton, D. (2011). Where Is My Bus? Impact of Mobile Real-time Information on the Perceived and Actual Wait Time of Transit Riders. *Transportation Research Part A: Policy and Practice*, 45(8), 839–848.

46. Wooldridge, J. (2009). *Introductory Econometrics* (4th Edition). South-Western Cengage Learning. p. 493.

47. Zhang, F., Shen, Q., & Clifton, K. J. (2008). Examination of Traveler Responses to Real-Time Information About Bus Arrivals Using Panel Data. *Transportation Research Record: Journal of the Transportation Research Board*, 2082, 107–115.

A APPENDIX

Table A.1 shows the monthly dummy variables that were estimated with the RE models shown in Table 3. The reference month is January.

Table A.1: Monthly Dummy Variables from the OLS and RE Regression Results

	OLS Estimate (SE)	**RE Estimate** (SE)	**RE AR(1)** (SE)	**RE ARMA(1,1)** (SE)
February	1117.39	872.98***	1058.51***	1081.55***
	(814.61)	(100.66)	(76.33)	(71.84)
March	1081.64**	676.46***	863.35***	864.16***
	(541.61)	(66.96)	(62.05)	(63.10)
April	829.6	198.64*	488.10***	480.49***
	(824.24)	(102.93)	(91.92)	(88.09)
May	1373.21*	743.60***	1047.96***	1069.85***
	(757.16)	(92.59)	(84.38)	(82.82)
June	1022.78	623.58***	943.37***	953.61***
	(1043.13)	(129.99)	(109.41)	(102.67)
July	484.85	23.39	362.85***	352.92***
	(1003.85)	(125.88)	(105.06)	(98.34)
August	-40.73	-527.98***	-219.18**	-242.25**
	(989.45)	(122.20)	(109.17)	(105.53)
September	1315.32	874.47***	1220.39***	1245.98***
	(915.17)	(112.78)	(96.91)	(92.38)
October	1064.12	801.73***	1015.85***	1052.95***
	(754.18)	(93.50)	(74.46)	(72.52)
November	588.67	344.29***	516.01***	521.14***
	(775.44)	(100.62)	(85.96)	(82.87)
December	-239.36	-391.04***	-323.73***	-328.60***
	(539.89)	(76.74)	(66.17)	(64.49)
Constant	44582.18**	27100.01***	35657.75***	40104.43***
	(18511.29)	(3623.85)	(3600.21)	(3419.68)

$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table A.2 shows the monthly dummy variables that were estimated with the RE and FE models shown in Tables 4 and 5. The reference month is January.

Table A.2: Monthly Dummy Variables from the FE and RE Regression Results

	Single Bus Time Variable						**Quartiles of Bus Service**		
	Random Effects Estimate			Fixed Effects Estimate			Fixed Effects Estimate		
		(SE)	(Robust SE)		(SE)	(Robust SE)		(SE)	(Robust SE)
February	963.28	(98.005)***	(90.000)***	930.15	(96.332)***	(86.758)***	930.94	(96.192)***	(86.908)***
March	628.99	(63.218)***	(60.755)***	644.20	(62.467)***	(60.056)***	643.57	(62.376)***	(60.304)***
April	254.00	(102.169)**	(64.004)***	262.45	(100.852)***	(62.648)***	262.40	(100.708)***	(63.118)***
May	727.75	(89.436)***	(67.570)***	731.56	(88.017)***	(66.341)***	731.46	(87.888)***	(66.696)***
June	654.83	(127.133)***	(91.465)***	616.36	(124.846)***	(85.429)***	614.30	(124.685)***	(85.754)***
July	50.44	(122.35)	(80.68)	-13.55	(120.27)	(80.22)	-17.33	(120.12)	(80.21)
August	-518.61	(116.292)***	(77.471)***	-564.39	(114.041)***	(81.299)***	-568.93	(113.889)***	(81.207)***
September	934.50	(110.822)***	(92.589)***	904.09	(108.646)***	(88.388)***	903.46	(108.497)***	(88.725)***
October	857.57	(92.464)***	(70.366)***	863.97	(90.882)***	(68.681)***	866.10	(90.757)***	(68.523)***
November	509.55	(94.518)***	(72.477)***	511.21	(93.272)***	(69.672)***	512.27	(93.139)***	(70.164)***
December	-282.22	(74.077)***	(57.864)***	-255.92	(75.108)***	(55.307)***	-256.31	(75.004)***	(55.336)***
Constant	25313.46	(2,542.899)***	(3,022.184)***	-			-		

** p<0.1; ** p<0.05; *** p<0.01*
Robust Standard errors calculated using the Huber/White/sandwich estimator.