UNIVERSITY OF MINNESOTA

Statistical Analysis of Fare Compliance

FINAL REPORT

Diwakar Gupta and Yibin Chen
Department of Industrial and Systems Engineering
UNIVERSITY OF MINNESOTA

CTS 14-08

CENTER FOR
TRANSPORTATION STUDIES

UNIVERSITY OF MINNESOTA

Statistical Analysis of Fare Compliance

Final Report

Prepared by:

Diwakar Gupta
Yibin Chen

Department of Industrial & Systems Engineering
University of Minnesota

August 2014

Published by:

Center for Transportation Studies
University of Minnesota
200 Transportation and Safety Building
511 Washington Ave. S.E.
Minneapolis, MN 55455

Table of Contents

List of Tables

List of Figures

Executive Summary

Metro Transit uses a Proof-of-Payment (PoP) system with barrier-free stations on its Hiawatha Light Rail Line (HLR). In a PoP system, trains and station platforms are paid zones. Patrons are expected to carry proof of payment at all times when riding the trains or on the platform. Metro Transit police officers enforce compliance by riding selected trains and demanding patrons to show proof of payment. They check cardholders' payment record with the help of Mobile Phone Validator (MPV) devices. Paper tickets and transfers are checked manually. Police officers issue citations or warnings if patrons are found not to be in possession of the proof of payment of valid fares. Officers maintain a daily log of their activities, which is referred to as the Patrol Activity Log (PAL). Metro Transit and other transit agencies that use the PoP system measure fare compliance to estimate missed revenue and to determine the effectiveness of their efforts to improve compliance. They also perform manual counts to estimate ridership. Ticket sales and tagged rides' data collection is automated.

In this project, researchers employ a suite of statistical methodologies to estimate compliance by utilizing ticket sales and tagged rides' data as well as data from MPV devices and PALs. Specifically, the following types of data are used in this project. Note that the Cubic data does not include noncompliant riders and valid transfers of paper-ticket carrying patrons.

1) *TRR data (2012)*: Passenger count data, manually collected by two employees of the Ridership and Revenue (TRR) group. Approximately 20% of all weekday trips of the HLR were counted.
2) *Cubic data (2012)*: Fares paid data from Ticket Vending Machines (TVM) and smart cards tagged at a Rail Validator (RSV).
3) *Special events' data (2012)*: Estimates of ridership for major sports events, concerts, etc.
4) *MPV data (01/2012-03/2012)*: Data obtained from mobile phone validators that were used to verify payments made by smart cards.
5) *PAL data (01/2012-03/2012)*: Data recorded by police officers containing the number of riders checked, citations, and warnings.

As a first step in their analysis, researchers confirmed that ridership varied by travel direction, station, stratum, and day-of-week. Each unique combination of these factors is referred to as a *group*. Just as ridership varied by group, researchers found that noncompliance rates also varied by group, which suggested that an ideal inspection strategy would also vary by group. However, the Metro Transit Police Department used several different objectives to select inspection strategies, which included reducing fare evasion and making transit services safe for both passengers and staff. The accurate estimation of evasion rate was not the primary driver of their choice of inspection strategy. This required researchers to account for noncompliance rate differences by group (by calculating weighted average of group-level estimates) because a simple average across all groups would lead to a biased estimate.

Researchers found that the average hourly ridership counts from the Cubic and the TRR data were highly correlated. Therefore, Metro Transit could use Cubic data to estimate ridership with appropriate inflation factors to account for riders who transferred from buses and those who were noncompliant. This would substantially reduce the frequency of manual (TRR) counts. Manual

counts would be needed in the future largely to verify that the relationships between TRR and cubic data remained intact.

The overall evasion rate from the analysis was found to be 0.55% for weekdays and 0.7% for weekends. Also, researchers found that evasion rate estimates from PAL (0.55% to 0.7%) and MPV data (4.36% to 5.52%) were quite different. Further investigation revealed the following possible reasons for these differences.

1) Estimates from PAL data were "actual" citations and warnings calculated as percentage of all passengers checked. Estimates from MPV data were based on a rule, which was applied to all data. The former was subject to officers' discretion, but the latter was not. For example, if the passenger last tagged at 2 hours and 35 minutes earlier than the checking time, the rule when applied to MPV data would deem that as evasion, but the police officer upon observing that might not issue either a citation or a warning.
2) It is possible that evasion was more prevalent among magnetic cardholders. That is, passengers who carry paper tickets and transfers from buses were in fact more likely to carry valid proof of paid fare.
3) It is possible that some police officers do not document warnings in their logs.
4) Another possibility is that police officers make mistakes in recalling how many patrons they checked, relative to the number of citations or warnings they issued. Because these numbers are based on officers' recollections, they could be inaccurate.

Researchers recommend further investigation of the calculated differences in non-compliance rates from PAL and MPV data. One area of improvement is better documentation of citations and warnings' data.

The analysis also showed that the current inspection protocol was more than adequate for Metro Transit to achieve a 10% relative precision (95% CI being no more than 10% of the mean evasion rate) on weekdays (current relative precision was about 9%), but that the current effort was inadequate for weekends for the 10% precision target (current relative precision for weekends was about 12%). Metro Transit Police could reduce their weekday effort to 70% of current values, without changing proportion of shifts in each time window and direction/station inspection strategy. Concomitantly, they would need to increase their weekend effort by a factor of approximately 1.5 to achieve a similar 10% estimation accuracy. The analysis also showed that going to a 5% precision would be prohibitively expensive, requiring almost tripling of the inspection effort on weekdays and more than quintupling of the inspection effort on weekends.

Researchers recommend Metro Transit repeat analysis similar to that presented in this report each year with new data set. The introduction of new fare types, new validator devices and new rail lines will change patterns observed in this report. Two areas for future research emerge from this report. First, Metro Transit will benefit from an integrated inspection strategy design that simultaneously improves noncompliance estimation and reduces noncompliance rate. Second, data collected by Metro Transit can be used analyze how patrons use different types of fare products (e.g. different types of passes), which could be used to improve design of fare products in the future, i.e. simultaneously promote transit usage and improve revenues.

Chapter 1 Introduction

On its Hiawatha Light Rail Line (HLR), Metro Transit uses a Proof of Payment (PoP) system with barrier-free stations. Trains and station platforms are paid zones and patrons are expected to carry proof of payment at all times when riding the trains or on the platform. Compliance is enforced by Metro Transit police officers who perform spot checks and issue citations or warnings if patrons are found not to be in possession of proof of payment of valid fares. Metro Transit also performs random sampling to estimate non-compliance and ridership.

Metro Transit needs to estimate fare compliance to calculate missed revenue and undertake efforts to increase revenue via better compliance. In this project, researchers develop a suite of methodologies for estimating compliance by utilizing data collected for other purposes (ticket sales and tagged rides) as well as data that pertains directly to non-compliance and evasion (e.g. data from the mobile phone validator device used by Metro Transit police officers, and patrol activity logs). Statistical methodologies are employed to develop point and interval estimates and to determine sample sizes needed to realize a desired degree of precision.

Accurate measurement of fare compliance is important for Metro Transit for a variety of reasons. For example, accurate estimates can help assess the effectiveness of different rates of inspection and citations. Metro Transit can potentially improve effectiveness of the use of its resources for estimating fare compliance by targeting its resources and adopting a multi-pronged approach of utilizing different types of data. This research can benefit the staff of the Revenue Operations Department of Metro Transit and other transit agencies that use PoP and barrier-free systems.

Chapter 2 Methodology

2.1 Data

The methodology we used in this report involves analyses of five types of data. These are:

1) *The Ridership and Revenue Group (TRR) passenger sampling data (2012)*: data obtained from two passenger samplers who count rides on approximately 20% of all weekday trips of HLR.
2) *Cubic data (2012)*: data includes fares paid at Ticket Vending Machines (TVM) and smart cards tagged on a Rail Validator (RSV). Cubic data contains ridership information except noncompliant riders and valid transfers of paper-ticket carrying patrons.
3) *Special-events data (2012)*: data includes estimation of ridership for special events such as major sports events. Counts are done at the beginning of each major sport's season to establish the percentage of attendees riding HLR for the games. These percentages are multiplied by each game's attendance during the rest of the year to estimate special-event rides.
4) *Mobile Phone Validator (MPV) data (01/2012-03/2012)*: data obtained from mobile phone validators that are used to verify payments made by smart cards.
5) *Patrol Activity Log (PAL) data (01/2012-03/2012)*: data recorded by police officers, containing the number of riders checked, citations, and warnings.

2.2 Ridership Estimation

2.2.1 Ridership estimation using TRR data.

The TRR data is first grouped by hour of day and direction of trip. Then within each group, we calculate average ridership per trip and count how many trips are scheduled for every hour and direction. Next, we multiply average ridership per trip by the number of scheduled trips for every hour and direction to obtain an estimate of total ridership by hour and direction. Adding ridership over all hours and both directions obtains the daily ridership estimate. The mathematical formulae used in these calculations are shown below.

Let $C^d_{\{i,j\}}$ denote a grouped-sample count of rides where

> d = N or S, denotes direction,
> i = 0, 1, 2, …, 23, denotes hour,
> j = 1, 2, 3, …, n^d_i, is the index of sample within each group, and
> n^d_i denotes the number of observations in the group of hour i and direction d.

Then average ridership per trip in hour i and direction d equals $\bar{C}^d_i = \frac{1}{n^d_i}\sum_{j=1}^{n^d_i} \mathrm{C}^{\mathrm{d}}_{\{\mathrm{i,j}\}}$. Let k^d_i denote the number of scheduled trips in the group of hour i and direction d, then total ridership in hour i and direction d equals $\frac{1}{n^d_i}\sum_{j=1}^{n^d_i} \mathrm{C}^{\mathrm{d}}_{\{\mathrm{i,j}\}} \times k^d_i$. Upon summing the ridership over all i's and both directions, we obtain

$$\text{daily average ridership} = \sum_{d=\{N,S\}} \sum_{i=0}^{23} \left(\frac{1}{n_i^d} \sum_{j=1}^{n_i^d} \mathrm{C}_{\{i,j\}}^{d} \times k_i^d \right),$$

$$\text{and standard error} = \sqrt{\sum_{d=\{N,S\}} \sum_{i=0}^{23} \left[\frac{1}{n_i^d} \times (k_i^d)^2 \times Var(\mathrm{C}_{\{i,j\}}^{d}) \right]},$$

where $Var(\mathrm{C}_{\{i,j\}}^{d})$ is the variance of ridership in hour *i* and direction *d*, i.e., $Var\left(\mathrm{C}_{\{i,j\}}^{d}\right) = \frac{1}{n_i^d} \sum_{j=1}^{n_i^d} (\mathrm{C}_{\{i,j\}}^{d} - \bar{C}_i^d)^2$.

Finally, a 95% confidence interval is

$$\sum_{d=\{N,S\}} \sum_{i=0}^{23} \left(\frac{1}{n_i^d} \sum_{j=1}^{n_i^d} \mathrm{C}_{\{i,j\}}^{d} \times k_i^d \right) \pm 1.96 \times \sqrt{\sum_{d=\{N,S\}} \sum_{i=0}^{23} \left[\frac{1}{n_i^d} \times (k_i^d)^2 \times Var(\mathrm{C}_{\{i,j\}}^{d}) \right]}.$$

When sample size is large enough, the critical point for 95% confidence interval approximately equals $\Phi^{-1}(0.975) = 1.96$, where $\Phi^{-1}(\cdot)$ is the inverse cumulative normal distribution function.

2.2.2 Ridership analysis using Cubic data.

Cubic data is analyzed by month, week, product type, stratum, and stratum and product. A weekday is divided into four strata (time periods): AM peak (6:00 – 9:00 am), midday (9:00 am – 3:00 pm), PM peak (3:00 – 6:30 pm), night (6:30 pm – 0:00 am and 0:00 – 6:00 am). Products are categorized into four types: TVM tickets, Passes, SV's and Group tickets. TVM tickets include all tickets paid at TVM. SV's are single tickets paid by stored value cards. Group tickets are tickets paid by stored value cards for more than one passenger. Passes include all other types of smart cards.

In the analysis, total ridership is first counted by month and week. Second, ridership is counted for every day and averaged for every day of week to estimate a daily ridership. Furthermore, rides are grouped first by stratum, then by product, and finally by stratum and product. For each group, we count the total ridership for every day and take average over all days to estimate the daily ridership in that group. In the Cubic data analysis, confidence intervals are obtained by 95% t-distribution C.I. calculation.

2.2.3 Total ridership adjustment using special-events data.

Special-events riders are grouped by weekday and weekend/holiday. Weekday special-events ridership is added up to the TRR ridership estimation to estimate the total weekday ridership.

2.3 Fare Compliance Estimation

The evasion or non-compliance in PAL data includes all instances of citations and warnings issued by police officers for reason of fare evasion. Citations and warnings issued for other reasons are removed. In total, 15 citations and 7 warnings were not counted because they were issued for reasons such as drunken behavior, fighting, and smoking.

The evasion or non-compliance in MPV data includes the following instances:

1) Card expiration: this occurs when a cardholder's card expiration date is earlier than the date on which the card is checked.
2) Time expiration: this occurs when the last tagging time is more than 2.5 hours earlier than the checking time for SV cardholder, 10-rides cardholder and young-adult cardholder. Note that these cards are not unlimited ride cards.

There are several other outcomes in the MPV data that may be interpreted as instances of evasion. However, those outcomes were not included. Specifically, the following cases were considered valid: (1) Read time is earlier than last use time. (2) Pass not supported: the MPV is not updated to recognize the new cards. (3) SV cardholder with negative stored value, but is not time expiration: one time negative stored value is granted as a grace ride.

From this point forward, we use the term *evasion* to refer to the above-mentioned compliance violation events in PAL and MPV data. Results from each data are analyzed separately.

From the analysis of the TRR data, researchers confirmed that ridership varies by travel direction, station, stratum, and day-of-week. In this report, each unique combination of travel direction, station, stratum, and day-of-week is called a *group*. Just as ridership varies by group, it is possible that evasion rates also vary by group. Suppose the mean evasion rates are different, but the inspection strategy is not sensitive to different rates of ridership and we take a simple average of evasion rates across all groups in the data. Then, the simple average may lead to a biased estimate. In order to avoid such bias, we will need to either use weighted average of group-level estimates, or design inspection strategies that take possible group-level heterogeneity into account by matching inspection frequency to the relative volume of ridership in each group.

To the best of our knowledge, Metro Transit police department selects an inspection strategy based on its goal of reducing fare evasion and making transit services safe for both passengers and staff. Inspection strategy is also dependent on officers' shift schedules and officers' availability for the purpose of inspection. Note that inspection is one of a multitude of tasks that police officers are required to perform. Stated differently, the accurate estimation of evasion rate is not the primary driver of police officers' choice of inspection strategy. Also, researchers were informed by transit police leadership that it might not be possible to alter inspection strategy because of the multitude of constraints on possible shift schedules and multiple duties performed by police officers. Researchers did develop a framework for aligning inspection strategies with the goals of improving estimation as well as reducing non-compliance. These strategies were presented in two TAP meetings that took place on July 23, 2013 and September 16, 2013, respectively. Powerpoint slides from the September 16 meeting, at which transit police leadership were present, are provided in Appendix C. For these reasons and to remain consistent with research contract deliverables, we focus in this report on developing statistical estimation procedures that account for heterogeneous ridership rates across groups. We do not propose an alternate inspection strategy.

The methodology used in the execution of this project has the following steps:

1) Verify whether average evasion rate varies by direction, station, stratum, and day-of-week.

We use binomial regression model with *log* link function to verify whether average evasion rate varies by direction, station, and stratum for weekdays and weekends/holidays separately. The definitions of group variables are shown in Table 2.1.

Table 2.1 Group Variables

Variable	Index	Explanation	Indicator
Stratum	0	Night: 0:00 am-6:00 am & 18:30pm-0: 00 am	$I_{\{stratum=0\}}$
	1	AM Peak: 6:00 am- 9:00 am	$I_{\{stratum=1\}}$
	2	Midday: 9:00 am – 15:00 pm	$I_{\{stratum=2\}}$
	3	PM Peak: 15:00 pm – 18:30 pm	$I_{\{stratum=3\}}$
Direction	0	South	$I_{\{d=0\}}$
	1	North	$I_{\{d=1\}}$
Station	1	Target Field Station (TFS)	$I_{\{station=1\}}$
	2	Warehouse District Station (Warehouse)	$I_{\{station=2\}}$
	3	Nicollet Mall Station (Nicollet)	$I_{\{station=3\}}$
	4	Government Plaza Station(Gvt Plaza)	$I_{\{station=4\}}$
	5	Metrodome Station (Metrodome)	$I_{\{station=5\}}$
	6	Cedar-Riverside Station (Cedar Riv)	$I_{\{station=6\}}$
	7	Franklin Avenue Station (Franklin)	$I_{\{station=7\}}$
	8	Lake Street Midtown Station(Lake Street)	$I_{\{station=8\}}$
	9	38th Street Station (38th)	$I_{\{station=9\}}$
	10	46th Street Station (46th)	$I_{\{station=10\}}$
	11	50th Street Station (50th)	$I_{\{station=11\}}$
	12	VA Medical Center Station (VA)	$I_{\{station=12\}}$
	13	Fort Snelling Station (Ft Snelling)	$I_{\{station=13\}}$
	14	Lindbergh Station (Lindbergh)	$I_{\{station=14\}}$
	15	Humphrey Station (Humphrey)	$I_{\{station=15\}}$
	16	American Blvd Station (Am Blvd)	$I_{\{station=16\}}$
	17	Bloomington Central Station (BCS)	$I_{\{station=17\}}$
	18	28th Avenue Station (28th Ave)	$I_{\{station=18\}}$
	19	Mall of America Station (MOA)	$I_{\{station=19\}}$

Four models are considered:

Base Model (M_0): $\log(p) = \beta_0$.

Model 1 (M_1): $\log(p) = \sum_{i=1}^{3} \beta_i \, I_{\{stratum=i\}} + \beta_0$.

Model 2 (M_2): $\log(p) = \sum_{i=1}^{3} \beta_i \, I_{\{stratum=i\}} + \sum_{i=1}^{3} \gamma_i \, I_{\{stratum=i\}} \times I_{\{d=1\}} + \gamma_0 I_{\{d=1\}} + \beta_0$.

Model 3 (M_3): $\log(p) = \sum_{i=2}^{19} \beta_i \, I_{\{station=i\}} + \beta_0$.

In Model 1, Stratum 0 (Night) is chosen as the base. In Model 2, the base is Stratum 0 (Night) and Direction 0 (South). In Model 3, the base station is Station 1 (Target field station). That is, the regression analysis finds coefficients that help us calculate the effect of different covariates relative to the base case in each model.

The regression analysis is comparative, i.e. it allows us to measure whether factors included in a particular model are significant. For example, if model M_1 is significantly different from model M_0, then we will conclude that evasion rate varies by stratum. Similarly, if model M_3 is significantly different from M_0, then evasion rate varies by station, and if M_2 is significantly different from M_1, then evasion rate also varies by direction. If a particular factor is found to not affect evasion rates, then it is removed from further consideration. The statistical significance of differences is tested by comparing the difference of *deviances* with $\chi^2_{(\theta)}$ distribution, where θ is the difference of two models' degrees of freedom. If the *p*-value is less than 0.05, then we conclude that the two models are significantly different.

2) Compute mean evasion rate and corresponding confidence intervals.

We need the following notation to explain our methodology. Let

w_i = adjusted weight for group i,

$\hat{p}_i$ = an estimate of evasion rate for group i,

x_i = total number of evaders checked for group i, and

n_i = total number of riders checked for group i.

Within each group, we assume that the number of evaders follows a binomial distribution. The evasion rate for group i is estimated by $\hat{p}_i = x_i/n_i$, and the variance of $\hat{p}_i$ is estimated by $Var(\hat{p}_i) = \hat{p}_i \times (1 - \hat{p}_i)/n_i$. For group i , the adjusted weight w_i is calculated as the corresponding ridership proportion of the daily average ridership.

The overall average evasion rate equals $\hat{p} = \sum_i w_i \, \hat{p}_i$, which is the weighted average over the estimation of evasion rate for each group. The variance of $\hat{p}$ is estimated as $Var(\hat{p}) = \sum_i w_i^2 \, Var(\hat{p}_i)$. The 95% confidence interval of $\hat{p}$ is calculated as $\hat{p} \pm 1.96 \times \sqrt{Var(\hat{p})}$. Recall that when the sample size is large, the critical point for 95% confidence interval approximately equals $\Phi^{-1}(0.975) = 1.96$. That is the 95% confidence interval of $\hat{p}$ is

$$\hat{p} \pm 1.96 \times \sqrt{\sum_i w_i^2 \times \frac{\hat{p}_i(1-\hat{p}_i)}{n_i}}.$$

3) Future number of riders that need to be checked

Based on the confidence interval analysis of evasion rate in Step 2, we also provide a recommendation for the inspection workload, i.e., the number of riders that need to be checked in the future to achieve the required precision of the calculated confidence interval. We assume that relative police effort will remain invariant, i.e. the proportion of number of riders checked for each stratum and direction will not change. This assumption is based on the feedback we received from transit police leadership.

We calculate the required effort index to achieve a precision level of α. The half-length of the 95% confidence interval, as a fraction of mean evasion rate, equals

$$\left(\frac{1.96}{\hat{p}}\right)\sqrt{Var(\hat{p})} = \left(\frac{1.96}{\hat{p}}\right)\sqrt{\sum_i w_i^2 \times \frac{\hat{p}_i(1-\hat{p}_i)}{n_i}}.$$

Let $\hat{\mu}$ denote an arbitrary effort index. This index will change the number of patrons inspected in each group i from n_i to $\hat{\mu} \times n_i$. Therefore, $\hat{\mu}$ is a number such that

$$\left(\frac{1.96}{\hat{p}}\right)\sqrt{\sum_i w_i^2 \times \frac{\hat{p}_i(1-\hat{p}_i)}{\hat{\mu} \times n_i}} \leq \alpha$$

$$\Leftrightarrow \hat{\mu} \geq \left(\frac{1.96}{\alpha\hat{p}}\right)^2 \left(\sum_i w_i^2 \times \frac{\hat{p}_i(1-\hat{p}_i)}{n_i}\right).$$

That is, the minimum effort index equals

$$\mu = \left(\frac{1.96}{\alpha\hat{p}}\right)^2 \left(\sum_i w_i^2 \times \frac{\hat{p}_i(1-\hat{p}_i)}{n_i}\right).$$

Chapter 3 Results

3.1 Ridership Estimation

3.1.1 Weekday average ridership estimation.

Table 3.1 Weekday Ridership

	Mean	SD	95% Confidence Interval	
Weekday (TRR*[1])	29,925	1,336	29,532	30,318
Weekday (Cubic)	16,183	1,973	15,939	16,428
Weekday (TRR + Special Events)	31,515			
Weekday (MT)	31,079			

TRR estimated ridership (29,925) is much larger than the Cubic ridership (16,183). The difference may consist of noncompliant riders and valid transfers from buses. The adjusted ridership is close to Metro Transit's estimation, which is shown in the last row of Table 3.2. In this row and in other similar tables MT stands for Metro Transit.

3.1.2 Total ridership estimation

Table 3.2 Total Ridership

	Total Ridership
Weekday (TRR + Special Events)	7,973,258
Weekday (MT)	7,924,172

3.1.3 Daily ridership analysis:

1) Weekday ridership by hour

[1] TRR* includes ridership on night shuttle that runs between the airport stations, which is 588/day.

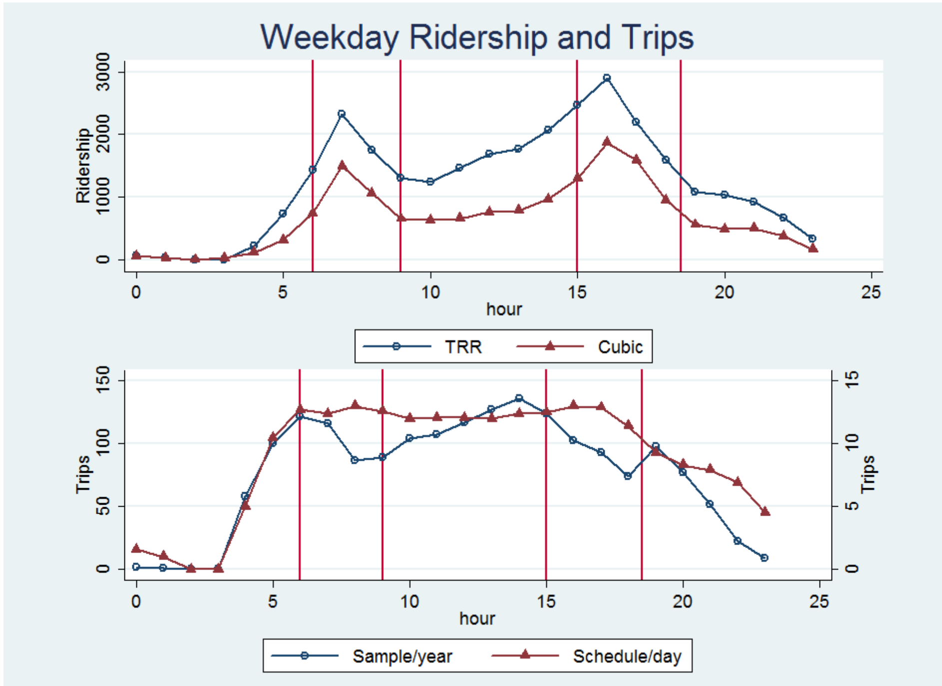


Figure 3.1 Weekday Ridership and Trips

Figure 3.1 shows the average ridership by hour for a regular weekday, estimated from TRR and Cubic data. Cubic and TRR data ridership counts by hour are highly correlated. Adjusted Cubic average ridership may be a good estimate for the actual ridership.

The number of trips counted during each hour in 2012 and the number of trips scheduled during that hour do not match. Ridership within a stratum changes by the hour. Therefore, taking an average of counts per trip within a stratum and multiplying the number of trips in that stratum can lead to inaccurate estimates. Counts averaged by hour will lead to more accurate estimation.

2) TRR ridership by station

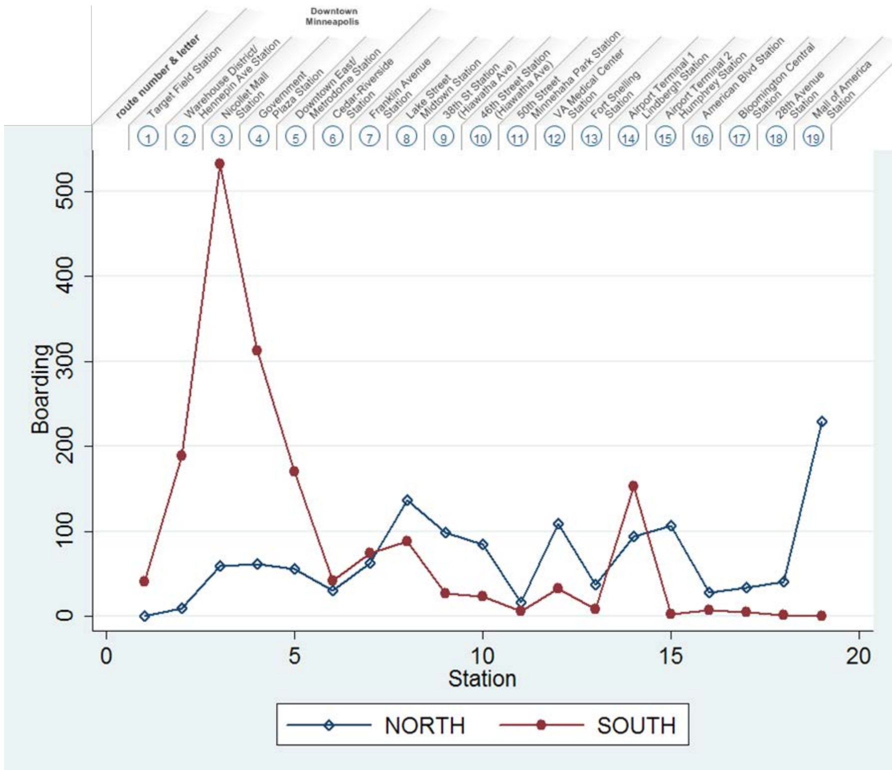


Figure 3.2 Hourly Average # of Passenger Boarding by Station

Figure 3.2 shows the average weekday ridership between 4 pm and 5 pm for each station. Ridership changes by station. In Figure 3.2, the number of passengers boarding at Station 3 (Nicollet Mall) is much larger than at other stations.

3) Cubic ridership analysis: the findings are listed below.

   a. Cubic ridership has seasonal patterns. Therefore, it may be useful in service planning if patterns are stable across years.

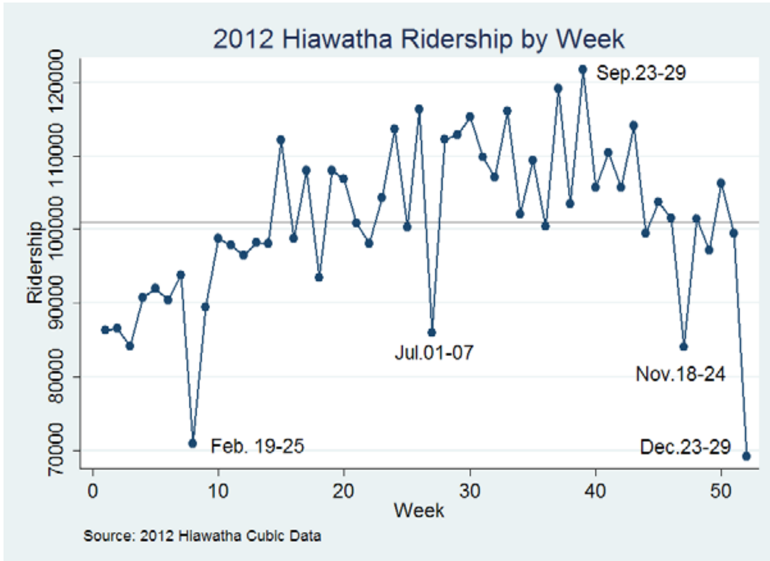
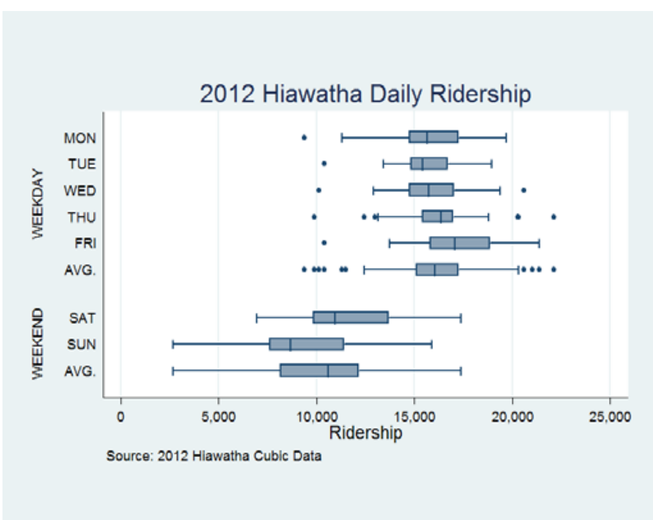


Figure 3.3 Cubic Ridership by Week

Figure 3.3 shows the total ridership for each week in 2012. The ridership tends to be smaller in winter and larger in summer.

b. The daily ridership for weekdays and weekends are different. However, ridership is fairly stable across weekdays.

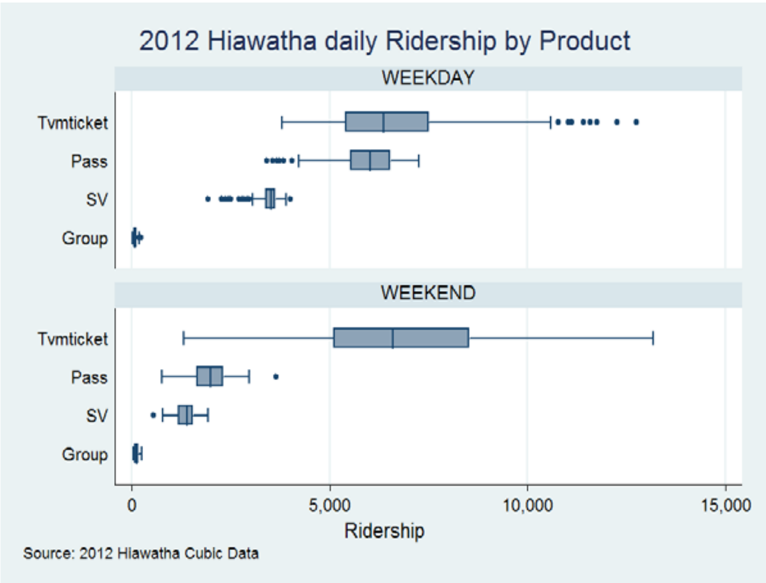


Obs.	Mean	Std. Err	95% Conf. Intl.	
MON(49)	15,829	296.5	15,233	16,425
TUE(51)	15,782	227.3	15,325	16,238
WED(51)	15,974	251.4	15,469	16,479
THU(51)	16,236	278.5	15,677	16,795
FRI(51)	17,082	297.6	16,484	17,680
Weekday (253)	16,183	124.1	15,939	16,428
SAT(52)	11,646	343.5	10,956	12,335
SUN(61)	9,404	369.9	8,664	10,144
Weekend (113)	10,436	274.7	9,891	10,980

Figure 3.4 Cubic Daily Ridership by Day of Week

Figure 3.4 shows the box plot, mean of the daily ridership and the corresponding 95% confidence interval for each day of week in 2012. It shows that weekdays and weekends are different. Monday to Friday are similar.

c. TVM tickets are the most used product.

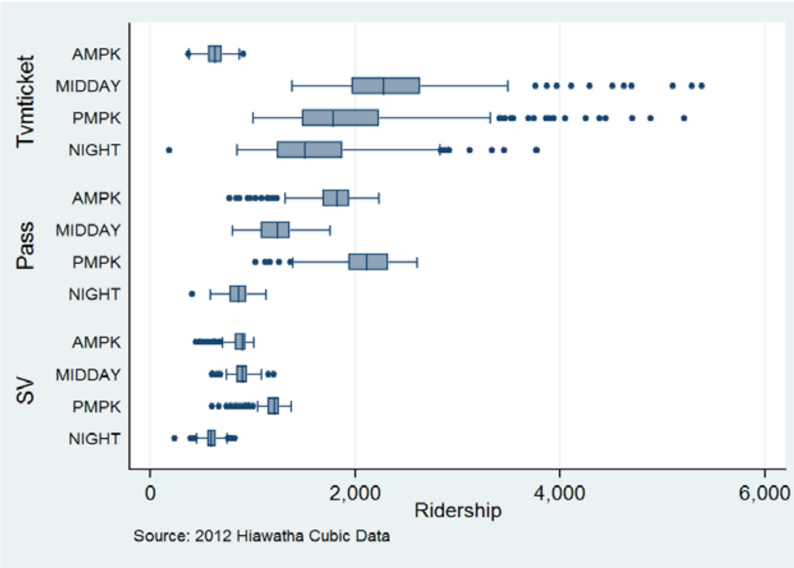


Obs.	Mean	Std. Err	95% Conf. Intl.	
WEEKDAY (253)				
TVM	6637	108.9	6422	6851
Pass	5981	47.6	5887	6074
SV	3471	17.4	3437	3505
Group	95	2.4	90	100
WEEKEND (113)				
TVM	6940	228.2	6487	7392
Pass	1993	47.1	1899	2086
SV	1379	24.8	1330	1428
Group	124	4.4	116	133

Figure 3.5 Cubic Ridership by Product

Figure 3.5 shows the box plot, mean and 95% confidence interval of the daily ridership for each product type in 2012. It shows that TVM tickets are the most used product and mean TVM ticket ridership seems to be similar for weekdays and weekends. The ridership of Passes is close to that of TVM ticket on weekdays, but about 1/3 of the TVM ticket ridership on weekends. SV cards are generally used less than Passes and TVM tickets.

d. TVM tickets are used more after AMPK and on weekends. Passes are more often used during AMPK and PMPK.



Obs.	Mean	Std. Err	95% Conf. Intl.	
TVM ticket				
AMPK	633	5	622	645
MIDDAY	2362	39	2284	2440
PMPK	2004	47	1911	2097
NIGHT	1638	34	1572	1705
PASS				
AMPK	1807	16	1775	1839
MIDDAY	1217	12	1194	1240
PMPK	2098	18	2062	2134
NIGHT	859	7	845	872
SV				
AMPK	871	6	859	883
MIDDAY	900	5	890	910
PMPK	1189	7	1175	1203
NIGHT	605	5	596	615

Figure 3.6 Cubic Ridership by Stratum and Product

Figure 3.6 shows the box plot, mean and 95% confidence interval of the daily ridership for each stratum and product type in 2012. TVM tickets are used less during AMPK. Passes are used the most during AMPK and PMPK.

3.2 Fare Compliance Estimation

3.2.1 Regression results for weekdays from PAL data:

1) Compare M_0 with M_1

Table 3.3 Regression Results for Strata: Model 0

Evasion	Coef.	Std. Err.	z	P>z	[95% Conf. Interval]	
_cons	-5.159712	.0408779	-126.22	0.000	-5.239831	-5.079593

In this table and others that report results of regression analyses, _cons represents the constant term, i.e. β_0. In tables with multiple regressors, the rows corresponding to statistically significant terms are shown in **bold font**.

Table 3.4 Regression Results for Strata: Model 1

Evasion	Coef.	Std. Err.	z	P>z	[95% Conf. Interval]	
Stratum						
1 AMPK	**-.707902**	**.140435**	**-5.04**	**0.000**	**-.9831495**	**-.4326545**
2 MIDDAY	-.1772812	.0977089	-1.81	0.070	-.3687871	.0142247
3 PMPK	**-.3911248**	**.1280504**	**-3.05**	**0.002**	**-.642099**	**-.1401507**
_cons	-4.912194	.07529	-65.24	0.000	-5.05976	-4.764628

Deviance analysis with these two models shows that p-value = 8.785e-07, which means that these two models are significantly different, i.e., evasion rate varies by stratum.

Table 3.3 shows that the overall average evasion rate is about $exp(-5.159712) = 0.0057$. Table 3.4 shows that Stratum 1 (AMPK) and Stratum 3 (PMPK) are significantly different from the base stratum (Night). The coefficients are negative, which means that the evasion rates for AMPK and PMPK are significantly smaller than that for Night stratum. The risk ratio of Stratum 1 (AMPK) is $exp(-.7079) = 0.493$, and the risk ratio of Stratum 3 (PMPK) is $exp(-0.3911) = 0.6763$. So the evasion rate for AMPK is about 50% of that for Night, and that the evasion rate for PMPK is about 68% of that for Night.

2) Compare M_0 with M_3

Table 3.5 Regression Results for Station: Model 0

Evasion	Coef.	Std. Err.	z	P>z	[95% Conf. Interval]	
_cons	-5.196002	.0445524	-116.63	0.000	-5.283323	-5.108681

Table 3.6 Regression Results for Station: Model 3

Evasion	Coef.	Std. Err.	z	P>z	[95% Conf. Interval]	
Station						
2 (Warehouse)	.4335253	.418853	1.04	0.301	-.3874116	1.254462
3 (Nicollet)	.7596316	.467536	1.62	0.104	-.156722	1.675985
4 (Gvt Plaza)	**1.047059**	**.4563273**	**2.29**	**0.022**	**.152674**	**1.941444**
5 (Metrodome)	**1.093967**	**.4199853**	**2.60**	**0.009**	**.2708109**	**1.917123**
6 (Cedar Riv)	-.3520757	.4747737	-0.74	0.458	-1.282615	.5784636
7 (Franklin)	.6526641	.4017355	1.62	0.104	-.134723	1.440051
8 (Lake Street)	.1339619	.397626	0.34	0.736	-.6453708	.9132946
9 (38th)	.1174255	.3989725	0.29	0.769	-.6645462	.8993972
10 (46th)	-.09105	.4383228	-0.21	0.835	-.9501469	.768047
11 (50th)	.6613848	.4103925	1.61	0.107	-.1429697	1.465739
12 (VA)	.0408178	.4330991	0.09	0.925	-.8080409	.8896765
13 (Ft Snelling)	**.9693516**	**.4103057**	**2.36**	**0.018**	**.1651671**	**1.773536**
14 (Lindbergh)	-.3896143	.516697	-0.75	0.451	-1.402322	.6230932

15 (Humphrey)	.4357411	.4232245	1.03	0.303	-.3937636	1.265246
16 (Am Blvd)	**1.161811**	**.5020352**	**2.31**	**0.021**	**.1778403**	**2.145782**
17 (BCS)	-.5079438	1.067678	-0.48	0.634	-2.600553	1.584666
18 (28th Ave)	.1385245	.6254529	0.22	0.825	-1.087341	1.36439
19 (MOA)	.3805071	.4008956	0.95	0.343	-.4052339	1.166248
_cons	-5.553513	.3772317	-14.72	0.000	-6.292874	-4.814153

Deviance analysis with these two models shows that *p*-value = 7.72e-09, which means that these two models are significantly different, i.e., evasion rate varies by station. Specifically, evasion rates at stations 4 (Gvt Plaza), 5 (Metrodome), 13(Ft Snelling) and 16(Am Blvd) are significantly different from those of other stations in the dataset. The risk ratios (exponential of the coefficients) of these four stations are 2.85, 2.99, 2.64 and 3.20 respectively, which means the evasion rates at these four stations are about three times of that at Target Field station.

3) Compare M_1 with M_2

Table 3.7 Regression Results for Direction: Model 1

Evasion	Coef.	Std. Err.	z	P>z	[95% Conf. Interval]	
Stratum						
1 AMPK	-.8981918	.1654413	-5.43	0.000	-1.222451	-.5739329
2 MIDDAY	-.2000269	.1079718	-1.85	0.064	-.4116478	.011594
3 PMPK	-.3323041	.1363014	-2.44	0.015	-.5994499	-.0651583
_cons	-4.932404	.0801745	-61.52	0.000	-5.089543	-4.775265

Table 3.8 Regression Results for Direction: Model 2

Evasion	Coef.	Std. Err.	z	P>z	[95% Conf. Interval]	
Stratum						
1 AMPK	-.8792594	.2280377	-3.86	0.000	-1.326205	-.4323138
2 MIDDAY	-.4683027	.1528422	-3.06	0.002	-.7678679	-.1687375
3 PMPK	-.3787696	.18006	-2.10	0.035	-.7316807	-.0258586
Direction=1 (North)	-.0521892	.1611111	-0.32	0.746	-.3679611	.2635826
Stratum & Direction						
1(AM) 1(N)	-.0350339	.331452	-0.11	0.916	-.684668	.6146001
2(MID) 1(N)	**.5530956**	**.2169991**	**2.55**	**0.011**	**.1277853**	**.978406**
3(PM) 1(N)	.1077144	.275804	0.39	0.696	-.4328514	.6482803
_cons	-4.908502	.1082459	-45.35	0.000	-5.120661	-4.696344

In Table 3.7 and Table 3.8, we run regression on the data with information on both direction and stratum (85674 riders) and the base case is Stratum 0 (Night) and Direction 0 (South). Deviance analysis with these two models shows that *p*-value = 0.016, which means that these two models are significantly different, i.e., evasion rate also varies by travel direction. In particular, Direction 1 (North) and Stratum 2 (Midday) interact and their combined effect serves to increase evasion rate.

The coefficients show that the evasion rate for Direction 0 (South) and Stratum 2 (Midday) equals $\exp(-.4683027) = 0.626$ times of the evasion rate for the base case Direction 0 (South) and Stratum 0 (Night), but the evasion rate for Direction 1 (North) and Stratum 2 (Midday) equals $\exp(-.4683027 - .0521892 + .5530956) = 1.03$ times the evasion rate for the base case. This means that the evasion rate for Direction 1 (North) and Stratum 2 (Midday) is about 1.65 times of the evasion rate for Direction 0 (South) and Stratum 2 (Midday).

We perform next a similar analysis on PAL data for weekend & holiday, and MPV data. The results are summarized in Table 3.9.

Table 3.9 Model Comparison

		M_0 vs. M_1		M_1 vs. M_2		M_0 vs. M_3	
		# of riders checked	*p*-value	# of riders checked	*p*-value	# of riders checked	*p*-value
Weekday	PAL	103598	8.78E-06	85674	0.016	90455	7.72E-09
	MPV	17472	1.64E-24				
Weekend & Holiday	PAL	37978	0.017	31883	4.80E-04	33700	1.44E-10
	MPV	3388	0.090				

In the data sets we received from Metro Transit, parts of the information for some of the observations are missing (such as direction and station). In the model comparisons, we can only use the data with full information, so in Table 3.9 the number of riders checked varies for different model comparisons. The numbers of riders checked are sufficiently large that the comparison results are statistically significant. For the MPV data, we do not have the information about direction and station at the time of checking. Therefore, MPV analysis compares weekday and weekend evasion rates across all strata. The *p*-values are obtained from the deviance analysis. We find that only the *p*-value for comparing Model 0 and Model 1 for weekend & holiday MPV data is not smaller than 0.05, which means that evasion rate for cardholders varies by stratum for weekdays, but not for weekends. In Table 3.9, we also summarize evasion rate comparisons for PAL data, which previously confirmed that evasion rate varied by stratum, direction and station for both weekdays and weekends.

3.2.2 Mean evasion rate and corresponding confidence intervals.

In TRR and Cubic data, we do not have information on the nearest station at the time when fare compliance was checked. Therefore, when computing the average evasion rate, we have to ignore possible differences in evasion rate across different stations. To estimate the overall evasion rate, we group PAL data by stratum and stratum & direction, the weight for each group is estimated by the relative volume of ridership from TRR data.

We also estimate the evasion rate for cardholders by using MPV data. However, we do not have access to the ridership for all cardholders, since TRR data do not separate paper ticket holders from cardholders. Similarly, Cubic data do not include noncompliant riders and those who transfer from a bus after buying paper tickets. We use three ways to obtain the weight for each stratum. The first method utilizes ridership of all passengers in the Cubic data. The second weight is calculated based only on cardholders in Cubic data. Finally, in the third method, we run linear regression on TRR ridership with cardholders' and paper-ticket holders' ridership in Cubic data for each stratum, and utilize the ridership of cardholders multiplied by corresponding regression coefficient as the weight.

The average evasion rates and 95% confidence intervals are summarized in Table 3.10. The relative precision is the half-length of the 95% confidence interval divided by the mean evasion rate, expressed as percent.

Table 3.10 Evasion Rate

Weekday					
	Weight	Mean	95% Confidence Interval		Relative Precision
PAL	TRR ridership by stratum & direction	0.546%	0.459%	0.597%	9.3%
	TRR ridership by stratum	0.559%	0.513%	0.606%	8.3%
	Non-weight	0.574%	0.528%	0.620%	8.0%
MPV	Cubic ridership by stratum	4.36%	3.98%	4.75%	8.8%
	Cubic ridership of cardholder by stratum	4.13%	3.75%	4.51%	9.2%
	Estimated ridership of cardholder by stratum	4.58%	4.17%	4.98%	8.8%
	Non-weight	3.82%	3.54%	4.11%	7.3%
Weekend & Holiday					
	Weight	Mean	95% Confidence Interval		
PAL	Cubic ridership by stratum	0.705%	0.620%	0.790%	12.1%
	Non-weight	0.700%	0.617%	0.784%	12.0%
MPV	Cubic ridership by stratum	5.52%	4.67%	6.36%	19.2%
	Non-weight	5.82%	5.05%	6.66%	13.1%

These analyses show that evasion rate for weekends is slightly larger than that for weekdays. The evasion rate for cardholders is much larger than the overall evasion rate.

For the PAL data, the weighted average evasion rate is quite close to the non-weighted average, which suggests that the number of riders inspected is approximately proportional to the overall ridership. However, the weighted evasion rate for cardholders is different from (slightly larger than) the non-weighted average for weekdays, which means that simply taking average over all evasion may underestimate the evasion rate for cardholders.

In Table 3.11, we summarize the respective percentage of citations and warnings in evasions from PAL data, and the respective percentage of card expirations and time expirations in cardholder evasions from MPV data. The 95% confidence intervals are calculated by the same method of the evasion calculation.

Table 3.11 Components of Evasion Rate

Weekday						
	Weight	Evasion	Components	Mean	95% Confidence Interval	
PAL	TRR ridership by stratum & direction	0.55%	Citations	0.34%	0.30%	0.38%
			Warnings	0.21%	0.18%	0.24%
	TRR ridership by stratum	0.56%	Citations	0.34%	0.30%	0.38%
			Warnings	0.22%	0.19%	0.25%
	Non-weight	0.57%	Citations	0.35%	0.31%	0.38%
			Warnings	0.23%	0.20%	0.26%
MPV	Cubic ridership by stratum	4.36%	Time Exp.	4.00%	3.61%	4.35%
			Card Exp.	0.39%	0.28%	0.50%
	Cubic ridership of cardholder by stratum	4.13%	Time Exp.	3.77%	3.40%	4.14%
			Card Exp.	0.36%	0.25%	0.47%
	Estimated ridership of cardholder by stratum	4.57%	Time Exp.	4.17%	3.78%	4.56%
			Card Exp.	0.40%	0.29%	0.52%
	Non-weight	3.82%	Time Exp.	3.45%	3.19%	3.73%
			Card Exp.	0.37%	0.28%	0.47%
Weekend & Holiday						
	Weight	Evasion	Components	Mean	95% Confidence Interval	
PAL	Cubic ridership by stratum	0.70%	Citations	0.44%	0.37%	0.50%
			Warnings	0.27%	0.22%	0.32%
	Non-weight	0.70%	Citations	0.43%	0.36%	0.50%
			Warnings	0.27%	0.22%	0.32%
MPV	Cubic ridership by stratum	5.52%	Time Exp.	4.97%	4.11%	5.70%
			Card Exp.	0.61%	0.33%	0.90%
	Non-weight	5.82%	Time Exp.	5.20%	4.47%	6.00%
			Card Exp.	0.62%	0.38%	0.95%

Chapter 4 Summary of Findings

1. The total adjusted ridership estimate (7,973,258) is slightly larger than Metro Transit's estimate (7,924,172).
2. Ridership varies by travel direction, station, stratum, and day-of-week.
3. The overall evasion rate is about 0.55% for weekdays and 0.7% for weekends, which is quite low.
4. The evasion rate for cardholders is much larger than the overall evasion rate. Our analysis shows that compared to overall evasion rate of 0.55% to 0.7%, the comparable evasion rate among cardholders is in the range 4.36% to 5.52%.
5. The evasion rate varies by stratum, direction and station.
6. To obtain a required confidence interval in the future, the total number of riders and effort index are shown in Table 4.1 as a function of acceptable tolerance.

Table 4.1 Recommended Sample Size and Effort Index

Weekday			
# of Riders Checked from PAL data	Tolerance α	Recommended Sample Size	Effort Index μ
103,598	1%	7,135,692	68.88
	5%	285,428	2.76
	10%	71,357	0.69
Weekend & Holiday			
# of Riders Checked from PAL data	Tolerance α	Recommended Sample Size	Effort Index μ
37,978	1%	5,542,003	145.93
	5%	221,680	5.84
	10%	55,420	1.46

Table 4.1 shows that from 01/01/2012 to 03/31/2012, police officers inspected about 103,598 passengers on weekdays and 37,978 passengers on weekends and holidays. To obtain a confidence interval with α relative precision, the total number of passengers inspected needs to be at least μ times the original sample size. Put differently, the recommended sample sizes are obtained by multiplying the number of riders checked in PAL data by the effort index μ.

Chapter 5 Concluding Remarks & Recommendations

We find that evasion rate estimates from PAL and MPV data are quite different. Below, we provide possible explanations for these differences.

1) Estimates from PAL data are "actual" citations and warnings as percent of all passengers checked. Estimates from MPV data are based on a rule, which is applied to all data. The former is subject to officers' discretion, but the latter is not. For example, if the passenger last tagged at 2 hours and 35 minutes earlier than the checking time, our rule when applied to MPV data will deem that as evasion, but the police officer upon observing that may not issue either a citation or a warning.

2) It is possible that evasion is more prevalent among magnetic card holders. That is, passengers who carry paper tickets and transfers from buses are in fact more likely to carry valid proof of paid fare.

3) It is possible that some police officers do not document warnings in their logs.

4) Another possibility is that police officers make mistakes in recalling how many patrons they checked, relative to the number of citations or warnings they issued. Because these numbers are based on officers' recollections, they could be inaccurate.

Our analysis also shows that the current inspection protocol is more than adequate to achieve a 10% relative precision (95% CI being no more than 10% of the mean evasion rate) on weekdays (current relative precision is about 9%), but that the current effort is inadequate for weekends for the 10% precision target (current relative precision for weekends is about 12%). Metro Transit police can reduce their weekday effort to 70% of current values, without changing proportion of shifts in each time window and direction/station inspection strategy. Concomitantly, they need to increase their weekend effort by a factor of approximately 1.5 to achieve a similar 10% estimation accuracy. Our analysis also shows that going to a 5% precision will be prohibitively expensive, requiring almost tripling of inspection effort on weekdays and more than quintupling inspection effort on weekends. Based on these observations, we make the following recommendations.

1. Metro Transit could use Cubic data to estimate ridership with appropriate inflation factors to account for riders who transfer from buses and those who are non-compliant.

2. Metro Transit could substantially reduce the frequency of manual (TRR) counts, which would be used largely to verify that the relationships between TRR and Cubic data remain intact.

3. Metro Transit police could reduce weekday inspection frequency by about 30% without affecting non-compliance estimates. That said, police presence and fare-compliance checks serve multiple purposes. It also serves as a deterrent and improves passenger and employee safety. Consideration of such concerns is beyond the scope of this project.

4. In order to realize a similar 10% estimation precision for weekends, Metro Transit police needs to increase its inspection frequency on weekends by approximately 50% relative to the current practice.

5. Researchers recommend further investigation of the calculated differences in non-compliance rates from PAL and MPV data. One area of improvement is better documentation of citations and warnings' data.

6. Researchers recommend Metro Transit repeat analysis similar to that presented in this report each year with new data set. The introduction of new fare types, new validator devices and new rail lines will change patterns observed in this report.

7. Two areas for future research emerge from this report. First, transit agencies will benefit from an integrated inspection strategy design that simultaneously improves non-compliance estimation and reduces non-compliance rate. Second, it will help transit agencies to conduct a detailed analysis of how patrons use different types of fare products (e.g. different types of passes), which will result in an improved design of fare products that promote transit usage and improve revenues.

Appendix A
Glossary of Terms

HLR	Hiawatha Light Rail Line
MPV	Mobile Phone Validator
MT	Metro Transit
PAL	Patrol Activity Log
PoP	Proof of Payment
RSV	Rail Validator
TRR	The Ridership and Revenue Group
TVM	Ticket Vendor Machine

Appendix B
Slides from May 31, 2013 TAP Meeting (Task 1 report)

The slides are used for explaining results for Task 1, describing detailed analysis of different types of data collected by Metro Transit.

Statistical Analysis of Fare Compliance

Task 1 Report

May 31, 2013

1

Task 1

Description

- Detailed analysis of the different types of data collected by Metro Transit

Deliverables

- A power point presentation containing analysis of data.

Key question

- Is the data sufficient for developing an overall strategy for non-compliance estimation?

2

2012 Cubic, TRR, and MPV data

RESULTS OF DATA ANALYSIS

3

Estimated 2012 Daily Ridership

Cubic + MPV Data

- Weekday
  20,229
- Saturday
  15,838
- Sunday
  12,836

TRR Data + Special Events

- Weekday
  31,515
- Saturday (MT)
  25,561
- Sunday (MT)
  20,151

The differences in estimates are likely caused by unaccounted riders – e.g. transfers from buses.

4

Weekday Ridership and Trips

5

Big Picture Take Aways

- Cubic and TRR data ridership counts by hour are highly correlated
- The number of trips counted during each hour in 2012 and the number of trips during that hour do not match
- Ridership within a Strata changes by the hour
- Taking an average of counts/trip within a Strata and multiplying the number of trips in that Strata can lead to inaccurate estimates
- Counts should be averaged by the hour

6

Sampling Strategy

- Sampling strategy needs to consider variability in hourly counts
- More variable ridership would generally mean more counts in that hour
- If the ratios between hourly counts of Cubic and TRR data are stable across years, these ratios can be used to estimate total ridership with fewer manual counts

7

2012 Data

TRR DATA ANALYSIS

8

Daily Ridership Calculations

1
- Group TRR data by direction and hour.

2
- Calculate average ridership per trip for every hour and direction.
- Count scheduled trips for every hour, south and north.

3
- Multiply average ridership per trip by trips for every hour and direction.
- Sum ridership over hour and direction.

Weekday Ridership and Trips (Copy of Slide #5)

TRR Daily Ridership

	Mean	SD	95% Confidence Interval	
Weekday (TRR*)	29,925	1,336	29,532	30,318
Weekday (Cubic)	16,183	1,973	15,939	16,428
Weekday (MT)	31,079			

The differences in our estimates from TRR data are likely caused by Metro Transit averaging across all trips within a Strata, rather than by the hour.

TRR* includes ridership on night shuttle that runs between the airport stations, which is 588/day.

SD = Standard Deviation in all slides.

Why are TRR and Cubic Counts so Different?

Conjectures:

- Noncompliance – may not be estimated well from MPV data
- Transfer from bus with TVM ticket (both valid and invalid) – neither Cubic nor MPV data captures these rides
- Incomplete data
- Other reasons

Takeaways - TRR Data Analysis

- The estimated weekday ridership (29,925) is slightly less than MT estimate (31,079)
- TRR estimated ridership is much larger than the Cubic ridership (16,183).
- TRR and Cubic ridership follow the same hourly pattern.
- Adjusted cubic average ridership may be a good estimate for the actual ridership.
- The procedure used to calculate TRR estimates needs to be studied more carefully.

13

2012 Data

SPECIAL EVENTS DATA ANALYSIS

14

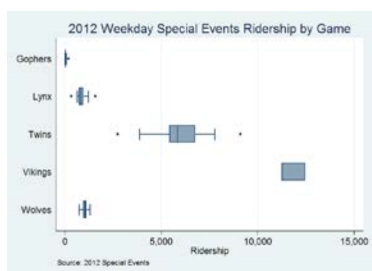
Total Ridership of Special Events

Day of Week	Ridership
Weekday	402,233
Saturday	113,499
Sunday & Holiday	194,891
Total (+ various)	714,623

Various: Miscellaneous Aquatenial 4000.

15

Weekday Special Events Ridership by Game



Obs	Mean	Std. Err.	95% Conf. Interval	
Gophers (11)	51	15.4	17	86
Lynx (16)	867	68.2	722	1012
Twins (54)	5920	153.5	5612	6228
Vikings (3)	11658	397.4	9949	13368
Wolves (31)	1069	23.7	1020	1117

16

TRR + Special Events Weekday Ridership

	Mean	95% Confidence Interval	
Weekday (TRR*)	29,925	29,532	30,318
Weekday (TRR + Special Events)	31,515		
		Total Ridership	
Weekday (TRR + Special Events)		7,973,258	
Weekday (MT)		7,924,172	

17

2012 Data

CUBIC DATA ANALYSIS

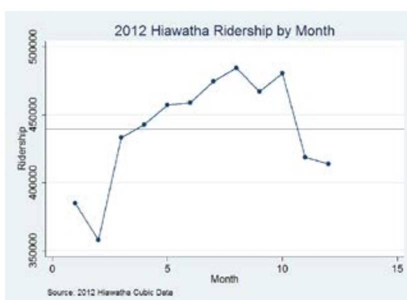
18

Many Possibilities

- Cubic data can be analyzed in a number of different ways depending on what statistic is of interest
- From ridership estimation viewpoint, it suffices to look at ridership by weekday strata and weekend
- We show data summaries by month, week, product, strata, and strata and product as examples

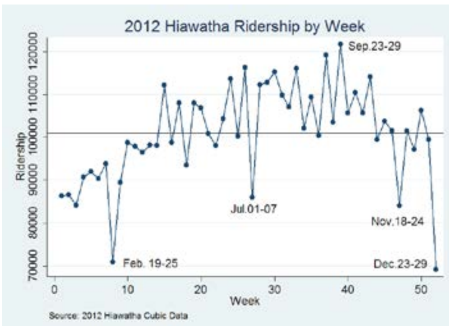
19

Cubic Ridership by Month

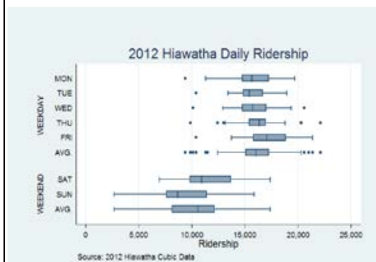


20

Cubic Ridership by Week



Cubic Daily Ridership by Day of Week

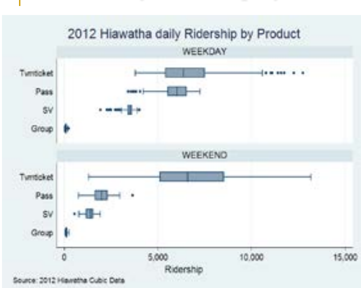


SUN: Sunday + Holiday

Obs	Mean	Std. Err.	95% Conf. Interval	
MON (49)	15829	296.5	15233	16425
TUE (51)	15782	227.3	15325	16238
WED (51)	15974	251.4	15469	16479
THU (51)	16236	278.5	15677	16795
FRI (51)	17082	297.6	16484	17680
Weekday (253)	16183	124.1	15939	16428
SAT (52)	11646	343.5	10956	12335
SUN (61)	9404	369.9	8664	10144
Weekend (113)	10436	274.7	9891	10980

Weekdays and Weekends are different. Monday to Thursday are similar. Friday is a little different from other weekdays.

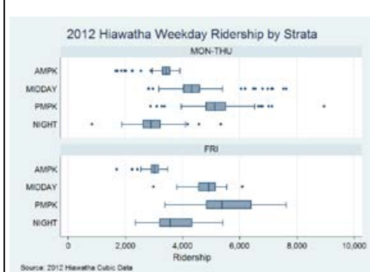
Cubic Daily Ridership by Product



Obs	Mean	Std. Err.	95% Conf. Interval	
WEEKDAY (253)				
TVM	6637	108.9	6422	6851
Pass	5981	47.6	5887	6074
SV	3471	17.4	3437	3505
Group	95	2.4	90	100
WEEKEND (113)				
TVM	6940	228.2	6487	7392
Pass	1993	47.1	1899	2086
SV	1379	24.8	1330	1428
Group	124	4.4	116	133

Mean TVM sales seem to be similar across all days of the week. TVM are greater relative to other categories on weekends. TVM and pass are similar on weekdays.

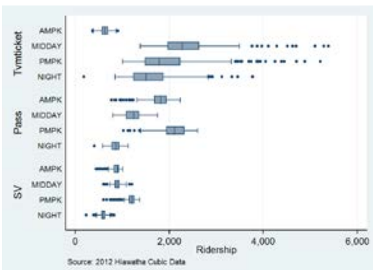
Cubic Weekday Ridership by Strata



Obs	Mean	Std. Err.	95% Conf. Interval	
MON-THU				
AMPK (202)	3385	24.9	3336	3435
MIDDAY (202)	4397	49.8	4299	4495
PMPK (202)	5224	52.8	5120	5328
NIGHT (202)	2950	36.5	2878	3022
FRI				
AMPK (51)	3015	41.8	2931	3099
MIDDAY (51)	4806	74.3	4657	4956
PMPK (51)	5558	133.6	5289	5826
NIGHT (51)	3704	98.3	3506	3901

Ridership is highest during PMPK period. MIDDAY ridership is greater than AMPK. This may be caused by MIDDAY being a 6h and AMPK being a 3h period.

Cubic Weekday Ridership by Strata and Product



*SV includes SV and Group tickets

Obs	Mean	Std. Err.	95% Conf. Interval	
TVM ticket				
AMPK	633	5	622	645
MIDDAY	2362	39	2284	2440
PMPK	2004	47	1911	2097
NIGHT	1638	34	1572	1705
PASS				
AMPK	1807	16	1775	1839
MIDDAY	1217	12	1194	1240
PMPK	2098	18	2062	2134
NIGHT	859	7	845	872
SV				
AMPK	871	6	859	883
MIDDAY	900	5	890	910
PMPK	1189	7	1175	1203
NIGHT	605	5	596	615

TVM used less during AMPK. SV cards are generally used less than other product types. Passes are used the most during AMPK and PMPK periods.

25

Takeaways - Cubic Data Analysis

- Cubic ridership has seasonal patterns – may be useful in service planning if patterns are stable across years
- The daily ridership for weekday and weekend are different
- However, ridership is fairly stable across weekdays
- TVM tickets are the most used product
- TVM tickets are used more after AMPK and on weekends
- Pass are more often used on AMPK and PMPK
- What is the true cost to MT of TVM vs. Pass vs. SV card use?
- Are there pricing strategies that could reduce net costs to MT?

26

2012 Data

MPV DATA ANALYSIS

27

Noncompliance (includes Evasion)

- Invalid Line:
  - Last use mode does not match with the current route.
- Time Expired:
  - Read time > Last use time + 2.5 hours
- Card Expired:
  - Read date > Pass expiration date
- Not Used:
  - Card is never used
- Data Errors (affect Evasion estimates):
  - We made certain assumptions
- Pass not supported is considered as valid

28

Evasion - assumptions

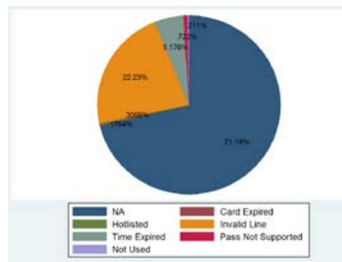
The following count as invalid fare cases:

- Card Expired
- Time Expired for SV and 10Ride card

The following count as valid fare cases:

- If Read time < Last used time, we assume that case as valid fare.

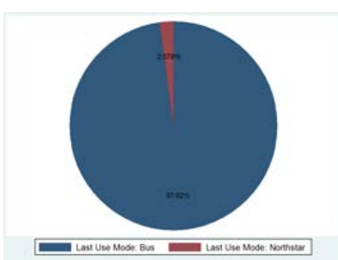
Invalid Reason



Invalid Reason	Freq.	Percent
NA	43,181	71.18
Card Expired	107	0.18
Hotlisted	186	0.31
Invalid Line	13,484	22.23
Time Expired	3,140	5.18
Pass Not Supported	438	0.72
Not Used	128	0.21
Total	60,664	100

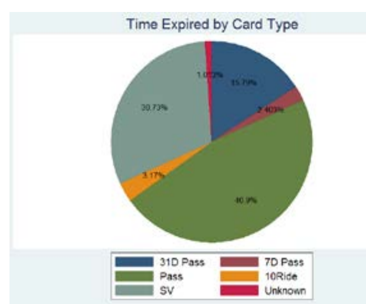
Source: MPV data for Hiawatha

Invalid Line by Last use mode



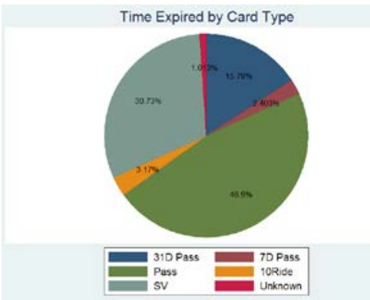
Last Use Mode	Freq.	Percent
Bus	13,292	97.92
Northstar	282	2.08

Time Expired by Card Type



Card Type	Freq.	Percent
31D Pass	1,091	15.79
7D Pass	166	2.4
Pass	3,240	46.9
10Ride	219	3.17
SV	2,123	30.73
Unknown	70	1.01
Total	6,909	100

Time Expired by Card Type



Card Type	Freq.	Percent
31D Pass	1,091	15.79
7D Pass	166	2.4
Pass	3,240	46.9
10Ride	219	3.17
SV	2,123	30.73
Unknown	70	1.01
Total	6,909	100

Noncompliance Proportion

Noncompliance	Mean	[95% Conf. Interval] (Binomial)	
Weekday (50994 obs.)	25%	24.6%	25.4%
Saturday (4459 obs.)	35.9%	34.5%	37.3%
Sunday/Holiday (5211 obs.)	36.4%	35.1%	37.7%
Overall	26.8%	26.4%	27.1%

Evasion	Mean	[95% Conf. Interval] (Binomial)	
Weekday (50994 obs.)	4.5%	4.3%	4.7%
Saturday (4459 obs.)	7.4%	6.6%	8.2%
Sunday (5211 obs.)	7.1%	6.4%	7.8%
Overall	4.9%	4.7%	5.1%

Noncompliance summary

- Invalid line (22%) > Time Expired (5%)> Card Expired (0.2%)
- Within Invalid line, transfer from bus dominates other reasons (98%).
- Within Time expired, Go-to-Pass (47%) and SV (31%) are major reasons.
- Noncompliance proportion is high on Weekend
- Total Noncompliance proportion: 27%
- Evasion proportion (5%-7%) is much smaller than Noncompliance.

2012 Data

CITATIONS DATA ANALYSIS

Daily Average Citations

Citations	Mean	[95% Conf. Interval] (Poisson)	
Weekday	5.4	5.1	5.7
Saturday	4.1	3.6	4.8
Sunday & Holiday	5.3	4.7	5.9
Overall	5.2	5.0	5.4

37

SUMMARY & NEXT STEPS

38

Summary Statistics

	Mean
Weekday (TRR)	29,925
Weekday (TRR + Special Events)	31,515
Weekday (Cubic)	16,183
Weekday (Cubic + MPV)	20,229
Weekday (MT)	31,079
Non Compliance Weekday	25%
Non Compliance Sat	36%
Non Compliance Sun/Holiday	37%
Non Compliance Overall	27%
Estimated Evasion	5%-7%
Estimated Citation Rate (over total ridership)	0.016%

39

Possible Strategies

1. Calculate hourly TRR and Cubic counts ratio, check if ratio is constant across years, use average TRR/Cubic to estimate ridership from Cubic data
2. Develop a TRR sampling plan based on variance of hourly counts
3. Is the difference between TRR and Cubic + MPV well understood? How many noncompliant riders does this include?
4. Better understand MPV data (errors)
5. Data make it difficult to tease out evasion within noncompliance
6. Could Mobile Validator Device record transfers?
7. Better understand citations data

40

THANK YOU.

41

Appendix C
Slides from September 16, 2013 TAP Meeting

The slides are used for discussing methods to analyze fare compliance and possible data collection with Transit police.

Statistical Analysis of Fare Compliance

September 16, 2013

1

Summary of Noncompliance Proportion

Noncompliance	Mean	[95% Conf. Interval] (Binomial)	
Weekday (50994 obs.)	25%	24.6%	25.4%
Saturday (4459 obs.)	35.9%	34.5%	37.3%
Sunday/Holiday (5211 obs.)	36.4%	35.1%	37.7%
Overall	26.8%	26.4%	27.1%

Evasion	Mean	[95% Conf. Interval] (Binomial)	
Weekday (50994 obs.)	4.5%	4.3%	4.7%
Saturday (4459 obs.)	7.4%	6.6%	8.2%
Sunday (5211 obs.)	7.1%	6.4%	7.8%
Overall	4.9%	4.7%	5.1%

Based on 2012 MPV data

2

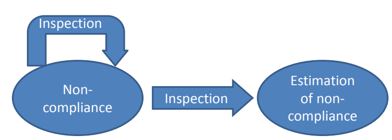
Project Goal

- Provide a better estimation of non-compliance.

3

However,

- Inspection is the only method to collect non-compliance data.
- Inspection pattern affects the true non-compliance performance.

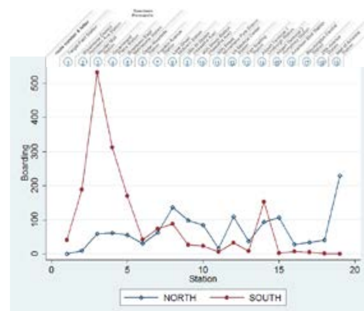


4

Our Approach

- Address both noncompliance estimation and reduction within the same strategy
- Use statistical methods to guide inspection, but leave substantial autonomy with MT Police and Revenue Operations Staff to identify implementable solutions
- Start simple, collect data, and refine models as more information becomes available

Hourly Average # of Passengers Boarding by Station



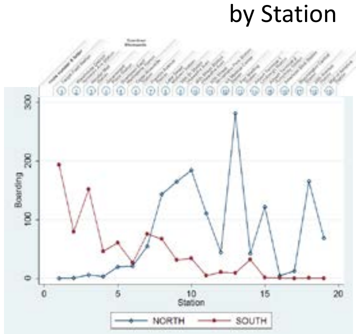
•Peak: Station 3 (Nicollet Mall), South.

•The No. of passengers boarding at Station 3 is much larger than at other stations. Inspecting at Station 3 may be a good strategy.

•To cover Station 6 to Station 18, round trip between 6 and 15 may be a good strategy.

Weekday 4 pm-5 pm, from 2012 TRR data

Hourly Average # of Passengers Boarding by Station



• The # boarding are more evenly distributed across stations, so it may be good to inspect in train.

•Mostly, whether to inspect in-train or on-station depends on how many passengers can be checked.

Weekday 7 am-8 am, from TRR data

Specifics

- Two deployment approaches
  - on-stations (peak volume, possibly multiple officers)
  - in-trains (low volume, possibly not covering the entire route)
- Identify high-value deployment windows, based on volumes at different time and stations.
- MT Police officers choose random shift schedules
- Collect more data, feedback from MT Police
- Refine strategies

On Station & In-train

Station	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
	6:07	6:09	6:07	6:05			6:08	6:05			6:09	6:07	6:05			6:07	6:06		
	6:12	6:09	6:11	6:13	6:12	6:10			6:13	6:11				6:13	6:10			6:14	6:11
	6:17	6:19	6:17	6:15	6:15	6:18	6:18	6:15			6:19	6:17	6:15			6:17	6:16		
	6:22	6:19	6:21	6:23	6:22	6:20	6:20	6:23	6:23	6:21				6:23	6:20			6:24	6:21
	6:27	6:29	6:27	6:25	6:25	6:28	6:28	6:25	6:25	6:27	6:29	6:27	6:25			6:27	6:26		
	6:32	6:29	6:31	6:33	6:32	6:30	6:30	6:33	6:33	6:31	6:29	6:31	6:33	6:33	6:30			6:34	6:31
	6:37	6:39	6:37	6:35	6:35	6:38	6:38	6:35	6:35	6:37	6:39	6:37	6:35	6:35	6:37	6:37	6:36		
	6:42	6:39	6:41	6:43	6:42	6:40	6:40	6:43	6:43	6:41	6:39	6:41	6:43	6:43	6:40	6:40	6:42	6:44	6:41
	6:47	6:49	6:47	6:45	6:45	6:48	6:48	6:45	6:45	6:47	6:49	6:47	6:45	6:45	6:47	6:47	6:46	6:44	6:48
	6:52	6:49	6:51	6:53	6:52	6:50	6:50	6:53	6:53	6:51	6:49	6:51	6:53	6:53	6:50	6:50	6:52	6:54	6:51
	6:57	6:59	6:57	6:55	6:55	6:58	6:58	6:55	6:55	6:57	6:59	6:57	6:55	6:55	6:57	6:57	6:56	6:54	6:58
	7:02	6:59	7:01	7:03	7:02	7:00	7:00	7:03	7:03	7:01	6:59	7:01	7:03	7:03	7:00	7:00	7:02	7:04	7:01
	7:07	7:09	7:07	7:05	7:05	7:08	7:08	7:05	7:05	7:07	7:09	7:07	7:05	7:05	7:07	7:07	7:06	7:04	7:08
	7:12	7:09	7:11	7:13	7:12	7:10	7:10	7:13	7:13	7:11	7:09	7:11	7:13	7:13	7:10	7:10	7:12	7:14	7:11
	7:17	7:19	7:17	7:15	7:15	7:18	7:18	7:15	7:15	7:17	7:19	7:17	7:15	7:15	7:17	7:17	7:16	7:14	7:18
	7:22	7:19	7:21	7:23	7:22	7:20	7:20	7:23	7:23	7:21	7:19	7:21	7:23	7:23	7:20	7:20	7:22	7:24	7:21
	7:27	7:29	7:27	7:25	7:25	7:28	7:28	7:25	7:25	7:27	7:29	7:27	7:25	7:25	7:27	7:27	7:26	7:24	7:28
	7:32	7:29	7:31	7:33	7:32	7:30	7:30	7:33	7:33	7:31	7:29	7:31	7:33	7:33	7:30	7:30	7:32	7:34	7:31
	7:37	7:39	7:37	7:35	7:35	7:38	7:38	7:35	7:35	7:37	7:39	7:37	7:35	7:35	7:37	7:37	7:36	7:34	7:38
	7:42	7:39	7:41	7:43	7:42	7:40	7:40	7:43	7:43	7:41	7:39	7:41	7:43	7:43	7:40	7:40	7:42	7:44	7:41
	7:47	7:49	7:47	7:45	7:45	7:48	7:48	7:45	7:45	7:47	7:49	7:47	7:45	7:45	7:47	7:47	7:46	7:44	7:48
			7:51	7:53	7:52	7:50	7:50	7:53	7:53	7:51	7:49	7:51	7:53	7:53	7:50	7:50	7:52	7:54	7:51
					7:55	7:58	7:58	7:55	7:55	7:57	7:59	7:57	7:55	7:55	7:57	7:57	7:56	7:54	7:58
							8:00	8:03	8:03	8:01	7:59	8:01	8:03	8:03	8:00	8:00	8:02		
									8:05	8:07	8:09	8:07	8:05	8:05	8:07			8:04	8:08
											8:09	8:11	8:13			8:10	8:12		

Grey: South (left to right)

Green: North (right to left)

E.g. On Station 13 from 6:05 to 6:55 then take round trip between Station 13 and Station 8.

9

Inspection Guide

Station	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
	6:07	6:09	6:07	6:05			6:08	6:05			6:09	6:07	6:05			6:07	6:06		
	6:12	6:09	6:11	6:13	6:12	6:10			6:13	6:11				6:13	6:10			6:14	6:11
	6:17	6:19	6:17	6:15	6:15	6:18	6:18	6:15			6:19	6:17	6:15			6:17	6:16		
	6:22	6:19	6:21	6:23	6:22	6:20	6:20	6:23	6:23	6:21				6:23	6:20			6:24	6:21
	6:27	6:29	6:27	6:25	6:25	6:28	6:28	6:25	6:25	6:27	6:29	6:27	6:25			6:27	6:26		
	6:32	6:29	6:31	6:33	6:32	6:30	6:30	6:33	6:33	6:31	6:29	6:31	6:33	6:33	6:30			6:34	6:31
	6:37	6:39	6:37	6:35	6:35	6:38	6:38	6:35	6:35	6:37	6:39	6:37	6:35	6:35	6:37	6:37	6:36		
	6:42	6:39	6:41	6:43	6:42	6:40	6:40	6:43	6:43	6:41	6:39	6:41	6:43	6:43	6:40	6:40	6:42	6:44	6:41
	6:47	6:49	6:47	6:45	6:45	6:48	6:48	6:45	6:45	6:47	6:49	6:47	6:45	6:45	6:47	6:47	6:46	6:44	6:48
	6:52	6:49	6:51	6:53	6:52	6:50	6:50	6:53	6:53	6:51	6:49	6:51	6:53	6:53	6:50	6:50	6:52	6:54	6:51
	6:57	6:59	6:57	6:55	6:55	6:58	6:58	6:55	6:55	6:57	6:59	6:57	6:55	6:55	6:57	6:57	6:56	6:54	6:58
	7:02	6:59	7:01	7:03	7:02	7:00	7:00	7:03	7:03	7:01	6:59	7:01	7:03	7:03	7:00	7:00	7:02	7:04	7:01
	7:07	7:09	7:07	7:05	7:05	7:08	7:08	7:05	7:05	7:07	7:09	7:07	7:05	7:05	7:07	7:07	7:06	7:04	7:08
	7:12	7:09	7:11	7:13	7:12	7:10	7:10	7:13	7:13	7:11	7:09	7:11	7:13	7:13	7:10	7:10	7:12	7:14	7:11
	7:17	7:19	7:17	7:15	7:15	7:18	7:18	7:15	7:15	7:17	7:19	7:17	7:15	7:15	7:17	7:17	7:16	7:14	7:18
	7:22	7:19	7:21	7:23	7:22	7:20	7:20	7:23	7:23	7:21	7:19	7:21	7:23	7:23	7:20	7:20	7:22	7:24	7:21
	7:27	7:29	7:27	7:25	7:25	7:28	7:28	7:25	7:25	7:27	7:29	7:27	7:25	7:25	7:27	7:27	7:26	7:24	7:28
	7:32	7:29	7:31	7:33	7:32	7:30	7:30	7:33	7:33	7:31	7:29	7:31	7:33	7:33	7:30	7:30	7:32	7:34	7:31
	7:37	7:39	7:37	7:35	7:35	7:38	7:38	7:35	7:35	7:37	7:39	7:37	7:35	7:35	7:37	7:37	7:36	7:34	7:38
	7:42	7:39	7:41	7:43	7:42	7:40	7:40	7:43	7:43	7:41	7:39	7:41	7:43	7:43	7:40	7:40	7:42	7:44	7:41
	7:47	7:49	7:47	7:45	7:45	7:48	7:48	7:45	7:45	7:47	7:49	7:47	7:45	7:45	7:47	7:47	7:46	7:44	7:48
			7:51	7:53	7:52	7:50	7:50	7:53	7:53	7:51	7:49	7:51	7:53	7:53	7:50	7:50	7:52	7:54	7:51
					7:55	7:58	7:58	7:55	7:55	7:57	7:59	7:57	7:55	7:55	7:57	7:57	7:56	7:54	7:58
							8:00	8:03	8:03	8:01	7:59	8:01	8:03	8:03	8:00	8:00	8:02		
									8:05	8:07	8:09	8:07	8:05	8:05	8:07			8:04	8:08
											8:09	8:11	8:13			8:10	8:12		

Grey: South (left to right)

Green: North (right to left)

10

Shift Schedules

- We will provide color-coded high-value on-station or in-train inspection blocks
- MT Police officers will use this information to develop many different 8-hr work templates
- On a given day, each police officer will pick his or her work template (or could be randomized)
- Scheduler will make sure that a variety of templates are used each day
- Work templates will be varied across officers and inspection days

11

Information & Data Collection

- Input: Average number of passengers checked per minute
- Data Collection
  - # of inspection
  - # of non-compliance
  - # of evasion
  - # of citation
  - Inspection time & location
  - (Currently no data about transfers with paper tickets or passengers without any ticket.)

12

Questions?

13

APPENDIX

14

Key Problem

- Noncompliance/ Evasion rate may depend on
  - Ticket type (Paper ticket or Go-to card)
    - No data about transfers with paper tickets or passengers without any ticket.
  - Time (by hour) and Location (by station)
    - Different ridership and traveller types (leisure, student, work, etc.)
  - Inspection (Inspection schedule & Citations)
    - Intense inspection may lead low noncompliance rate.

15

Task 2 Goal

- Deliverable:

  A power-point presentation containing an overall strategy for non-compliance estimation. This will be a first-pass analysis presented for feedback from Metro Transit and modification (as needed).
- June 1 through November 30, 2013

16

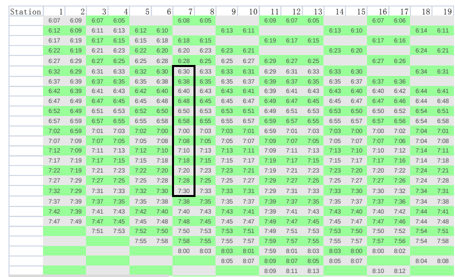
Hourly Average # of Passengers Boarding by Station

1
- Group TRR data by hour, station and direction.

2
- Calculate average # of passengers boarding per trip for every hour, station and direction.
- Count scheduled trips for every hour, south and north.

3
- Multiply average # of passenger boarding per trip by trips for every hour and direction.

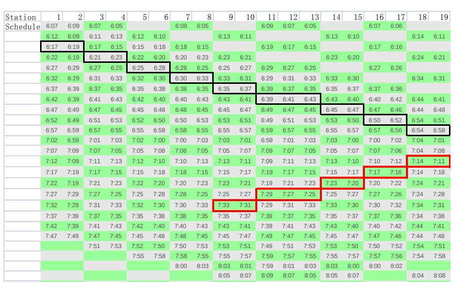
On Station Blocks



Grey: South (left to right)
Green: North (right to left)

On Station:
E.g. On Station 7 from 6:30 to 7:30.

In-train Blocks



Grey: South (left to right)
Green: North (right to left)

In-train:
E.g. 1. Take train from Station 1 to Station 19.
2. Take train from Station 19 to Station 9.

Assumptions

- Single line, fixed schedule
- Only MT Police officers can check compliance
- Fixed # of MT police
- Fixed shift constraints
  - E.g. any 8 hrs in a 12 hrs spread
- MT police are able to check compliance both on stations and in trains
- Average number of passengers checked/minute is constant and known

Resources

- Different MPV device / Software
- Someone to shadow police officers to track number checked (including paper transfers), and proportion of those found noncompliant
  - Base line data (at least 1 month)
  - After data (at least 1 month)
- Work with police officers to design shifts if desired

21

Station-Based Counting for a deployment window

- $X_{i,j}$: # of noncompliant checks at station i during hour j.
- $X_{i,j} = \min\{ N_{i,j}/60, C_s\} * t * \gamma_j$.
- $N_{i,j}$: # of passengers boarding on station i during hour j.
  - Obtain from Cubic data after inflating by a multiplier obtained from TRR data.
- C_s : # of passenger checked per minute.
  - Need to get this from MT police.
- γ_j : Non-compliance rate for hour j.
- t (min): time duration for this particular deployment window (time between arrivals of trains for station i)

22

In-train Counting for a deployment window

• Since we do not have much alighting information, we assume officers will only check new passengers that board between two station.

•$Y_{i,j}$: # of noncompliant passengers checked in train that starts from station i during hour j.
• $Y_{i,j} = \min\{N_{i,j} * t/60, C_T * \tau\} * \gamma_j$.

• $N_{i,j}$: # of passengers boarding on station i during hour j.
• C_T : # of passenger checked per minute.
• γ_j : Non-compliance rate for hour j.
•τ (min) : time duration for this particular deployment window (time spent from station i to the next station in hour j).
•t (min): time between arrivals of trains for station i

23