An Algorithm to Measure Daily Bus Passenger Miles Using Electronic Farebox Data

Alex Lu, Alla Reddy

Operations Planning
New York City Transit Authority

Presented at the 90th Annual Meeting of the Transportation Research Board
Washington D.C. (2011)

Purpose and Need

- **Implement 100% electronic data reporting**
  - Monthly “safety module”
  - Eliminates surveying, data entry, manual checking
  - More consistent & accurate
- **Algorithm requirements**
  - Zero manual intervention
  - Fast: running time of a few minutes per day of data
  - Rely on schedules and AFC data (no GPS/AVL/APC)

Photo: **Adam E. Moreira**

NYCT's MetroCard AFC Data

- **"Trip" file**
  - partial trip records
  - no timestamps for cash transactions
  - evaders/half fares inconsistently logged
  - can't be automatically matched to schedule
- **"Transaction" file**
  - MetroCards only, no cash data
  - no trip information
  - no geographic information (not 'location-stamped')
  - transaction time rounded to nearest six minutes
  - headsign route and directions not always correct

73 bytes per record × about 8,000,000 bus and subway records per weekday = approximately 550 MB per weekday (3am to 2.59am next day)

Hypothetical card with bus-only records shown:

```
....x....1....x....2....x....3....x....4....x....5....x....6....x....7.
2653058017 20080416  55400 157   027 F02569  1      R482      0     362
2653058017 20080416  63000 157   027 F0027F  1      R480      0     494
2653058017 20080416  73600 157   027 F01E70  2      R494      0     153
2653058017 20080416 160000 157   027 F01E72  2      R494      0     152
2653058017 20080416 161800 157   027 F00214  1      R480      0     494
2653058017 20080416 163600 157   027 F00129  1      R480      0     495
2653058017 20080416 184800 157   027 F020B0  3      R515      0     645
```

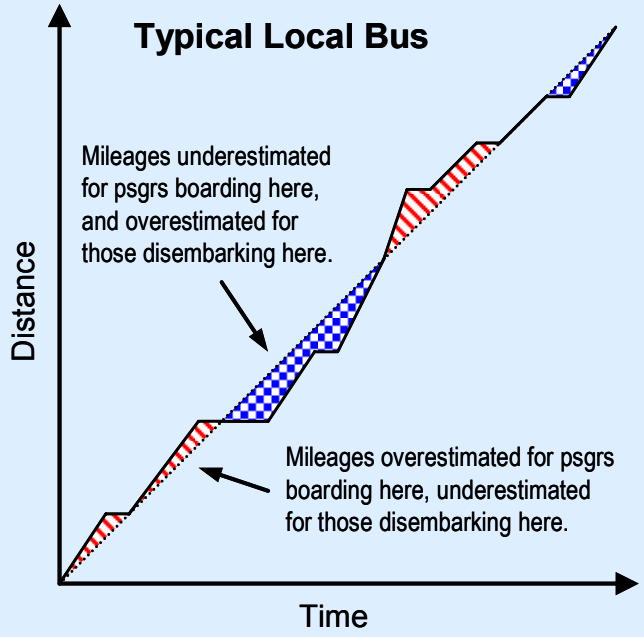
Fare Media Class (Unlimited, Pay-per-Ride)
Transcation Code (Value Deduct, Transfer)
Transaction Date and Time (Nearest Six Minutes)
MetroCard ID (Serial Number)

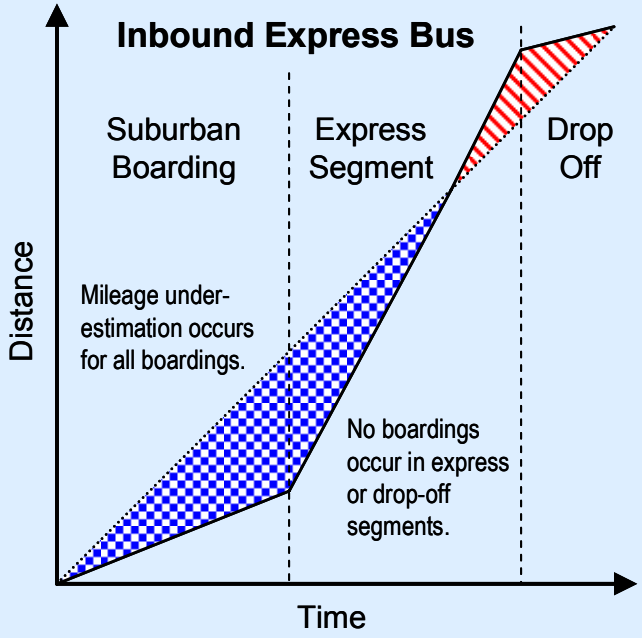
Bus Number in Hex (Subtract 0xF00000)
Carrier (1 = TA, 2 = OA, 3 = MTABC)
Value Deducted (0 for Unlimited Pass)
Location Code (Bus Route and Direction)

Assumptions

- **Localizations anchored relative to trip origin**
  - no en-route put-ins
- **Constant average speed (origin to destination)**
- **Symmetric daily activity**
  - Navick & Furth (2002)
- **Unbalanced ridership doesn't affect pattern**
- **End-of-trip enforced after maximum trip time**
  - Service continues in opposing direction
- **Uniform non-AFC rider correction factors**

Constant Average Speeds

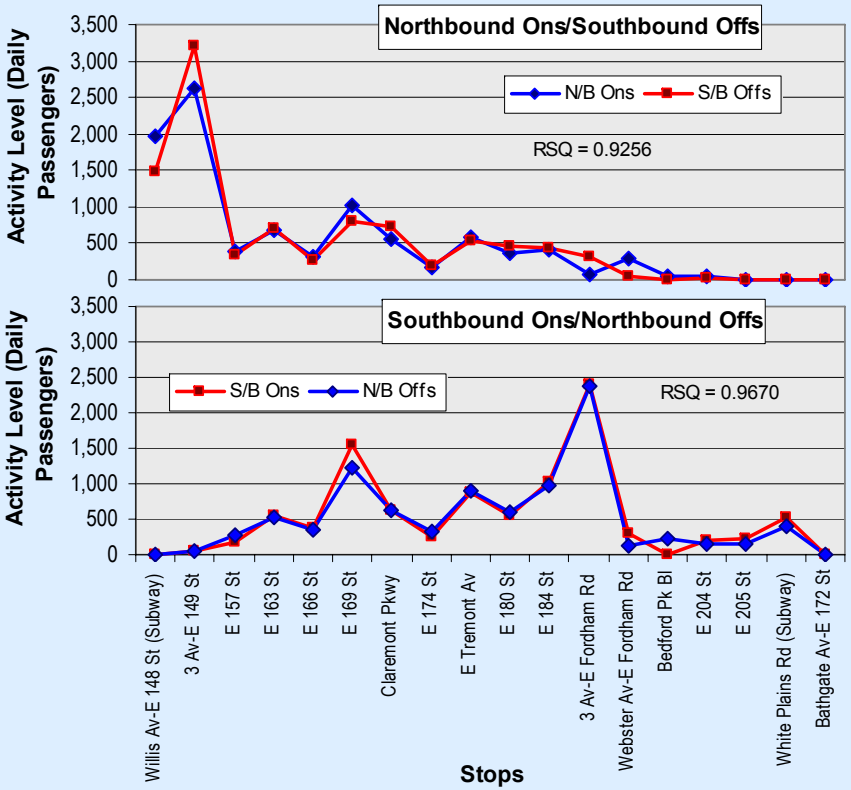


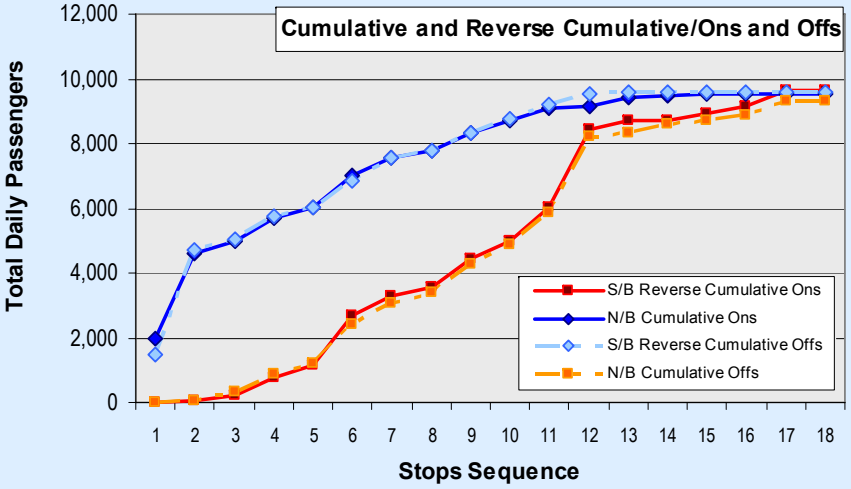


- **Local Bus**
  - Uniform speed profiles
  - Equally likely over- or under- estimation
- **Express Bus**
  - Many boardings in suburban segment
  - Speeds over-estimated
  - Mileages under-estimated
- **Future Work**
  - Use average speeds between timepoints

Symmetric Daily Activity Patterns

- **Conservation of Passengers**
  - Those disembarking here, reboard here later
- **Equal & Opposite Passenger Activities**
  - Return passengers reboard an opposing number
  - Minimal and balanced triangulation
- **High Correlation (BX55):**
  - Northbound Ons versus Southbound Offs
  - R-squared > 0.90





Symmetry Analysis Results

Route	Type	Boarding Direction	R-squared	Boarding Direction	R-squared
BX12	Crosstown-feeder hybrid	EB	0.992	WB	0.996
BX29	Double-ended feeder	EB	0.983	WB	0.981
BX55	Limited stop trunk	NB	0.997	SB	0.999
B31	Low density feeder	NB	0.992	SB	0.994
B42	Subway extension	NB	0.978	SB	0.985
B51	Bridge route	NB	0.833	SB	0.907
M15	High volume trunk route	NB	0.987	SB	0.993
M18	Branch shuttle	NB	0.924	SB	0.978
Q46	Suburban trunk	EB	0.972	WB	0.999
Q74	Campus feeder/circulator	NB	0.904	SB	0.962
Q79	Suburban crosstown	NB	0.982	SB	0.981
Q83	Suburban feeder	EB	0.995	WB	0.996
S4090	Suburban trunk	EB	0.994	WB	0.997
S52	Suburban feeder	NB	0.973	SB	0.986
S60	Neighbourhood circulator	NB	0.698	SB	0.833
S79	Suburban trunk	NB	0.999	SB	0.986

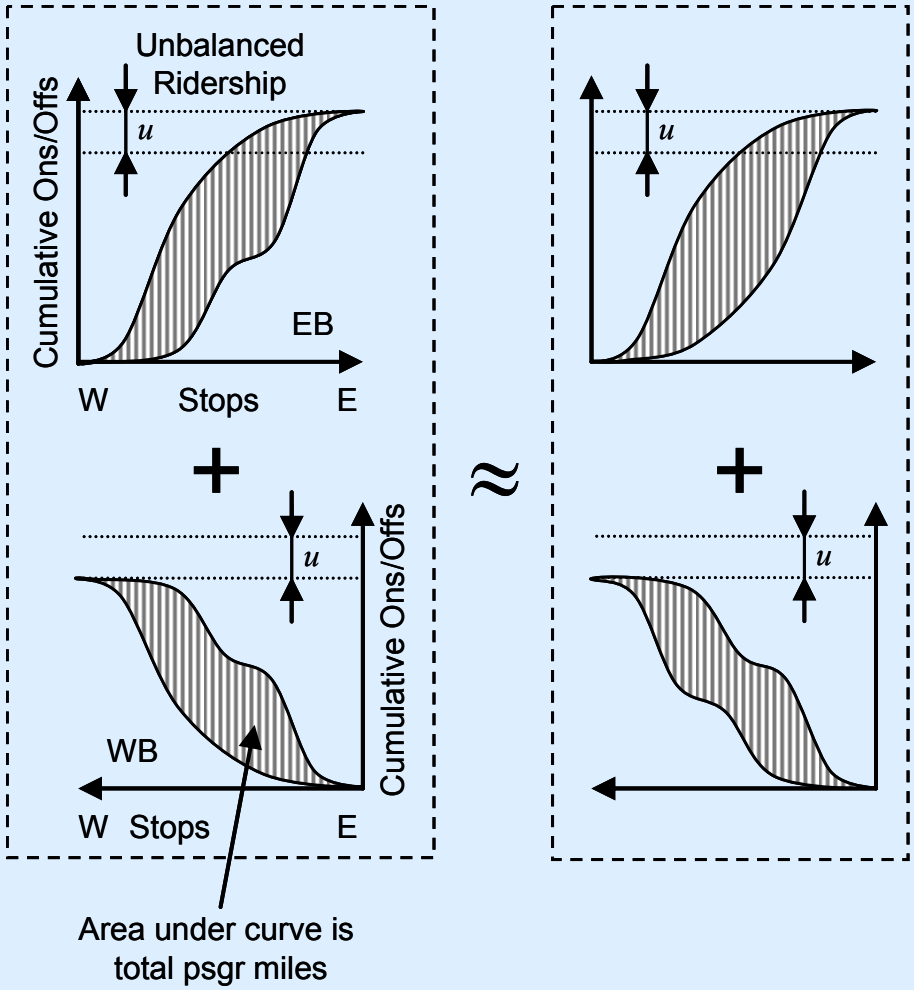
- **Supports earlier results**
  - Pittsburgh light rail*
  - Los Angeles bus (5 routes)**
  - New York City bus (sixteen routes)
- **Two notable exceptions:**
  - **S60 Grymes Hill** (neighbourhood circulator)
  - **B51 Manhattan Bridge**
  - Both lightly used routes eliminated in MTA's *2010 Service Reduction Plan*
  - Both operate highly asymmetrical route paths

* Furth, Peter G. Innovative Sampling Plans for Estimating Transit Passenger Kilometers. In *Transportation Research Record 1618*, TRB, Washington D.C., 1998, pp. 87–95.
** Navick, David S. and P.G. Furth. Estimating Passenger Miles, Origin-Destination Patterns, and Loads with Location-Stamped Farebox Data. In *Transportation Research Record 1799*, Paper No. 02-2466, TRB, Washington D.C., 2002, pp. 107–113.

Asymmetrical 'Loopy' Routes

- **S60 moves 210 daily psgrs mostly 'up the hill'**
  - Fewer downhill riders
  - Dogbone-shaped path
  - More stops and longer path uphill than downhill
- **B51 drives 900 daily psgrs 'over the bridge'**
  - All stops Manhattan bound; limited stop Brooklyn bound
  - Baseball-cap shaped path
- **Substantially violate symmetry assumptions**

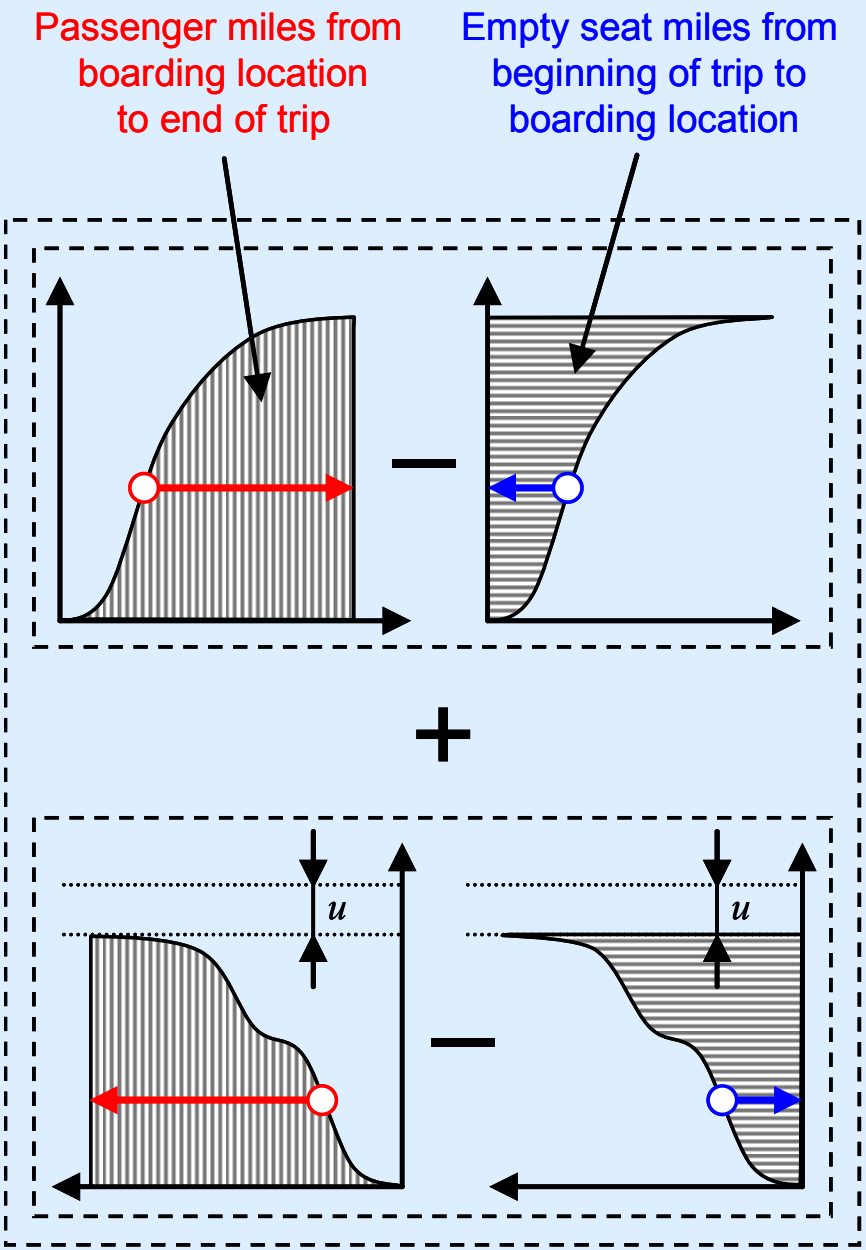
Unbalanced Route Assumption



- **"Unbalanced" routes**
  - Significantly different daily inbound/outbound ridership
- **Reasons:**
  - Afternoon/night travel pattern and modal bias differences
  - Carpool/drop-off
  - Triangulation due to afternoon activities
  - Express bus competition
- **Algorithm tolerant of small route imbalances**
  - Integration errors are small
  - Correlations remains high: M15, B42, B31, Q46, Q83

Direct Numerical Integration

- **Simple algebraic transformations**
  - Greatly speeds up algorithm execution
  - Sequential one-pass AFC file evaluation
  - 600 MB of daily data analyzed in about 3 minutes



All Correction Factors

Service	Bus Passengers	Description	Factor Req'd?	Factor Value[1]
Standard Bus	In EU65 MetroCard Transaction File	Base "raw" passenger boarding data from the MetroCard AFC system.	No	—
	Cash Passengers	Passengers not using electronic fare media.	Yes	15.4%[2]
	Non-Farebox Passengers	Passengers not interacting with farebox due to broken farebox, fare evasion, paper tickets, flash passes, etc.	Yes	12.1%[3]
	Farebox Data Transmission Errors	Passengers paying fare normally but data not in EU65 file due to farebox data transmission malfunction.	Yes	5.4%[4]
	Total Adjustment Factor (Standard Bus)			**32.9%**
Select Bus Service	Revenue Passenger Data from Select Bus Service Fare Validation Machines	Number of receipts issued by the Proof-of-Payment wayside fare collection machines (Cash and MetroCard) on the BX12 Select Bus Service.	No	—
	Non-Receipt Passengers	Passengers without POP receipts due to fare evasion, paper tickets, and flash passes, etc.	Yes	14.9%[3]
	Fare Validation Machine Data Transmission Errors	Passengers paying fare normally, but not recorded because of fare payment machine malfunction.	Yes	2.0%[3]
	Total Adjustment Factor (Select Bus)			**16.9%**

Notes: 1. Correction factor values are expressed as a percentage of raw AFC counts and are current as of December, 2008; **2.** Updated monthly; **3.** Updated as required; **4.** Updated every two months.

Non-Farebox Ridership Survey

Category	Passengers Observed	Percentage
Valid MetroCard Fare	19,630	85.4%
Valid Split Fare	294	1.3%
Invalid MetroCard Fare	449	2.0%
Invalid Split Fare	130	0.6%
Subtotal AFC Counted Passengers	**20,503**	**89.2%**
Front Door Non-Paying Passenger	644	2.8%
Rear Door Non-Paying Passenger	110	0.5%
Child Over 44" Travelling Without Fare	413	1.8%
Paper Ticket	47	0.2%
Child Under 44" Travelling Without Fare	557	2.4%
Flash Pass, Uniform, or Official Badge	268	1.2%
Wheelchair Travelling Without Fare	44	0.2%
Seamless Transfers[1]	31	0.1%
Select Bus Receipt on BX12 Local[2]	27	0.1%
% of Trips with Farebox Malfunction	1.40%	
Broken Farebox Psgr Boardings[3]	336	1.5%
Subtotal Unaccountable Passengers	**2,477**	**10.8%**
Total Passengers	22,980	100.0%

Notes

1. Transfers between L Train and **B42** Bus to Canarsie Beach at Canarsie Subway station occurs within fare control. Transferring passengers are not required to swipe upon boarding, thus no AFC record is generated.

2. BX12 Select Bus utilizes a proof-of-payment (POP) fare collection system. POP receipts are not valid on board BX12 local buses, but are occasionally accepted by drivers.

3. Estimated passenger boardings during farebox malfunction based on 1.4% evenly distributed.

Survey Method: Trained surveyors seated near bus front door rode 421 trips using a 24/7 stratified random sample, and classified each boarding into 13 categories during May 13~November 29, 2008. Lost data (memory failure) separately estimated. Non-farebox passenger expansion factor is 2,477 ÷ 20,503 = **12.1%**.

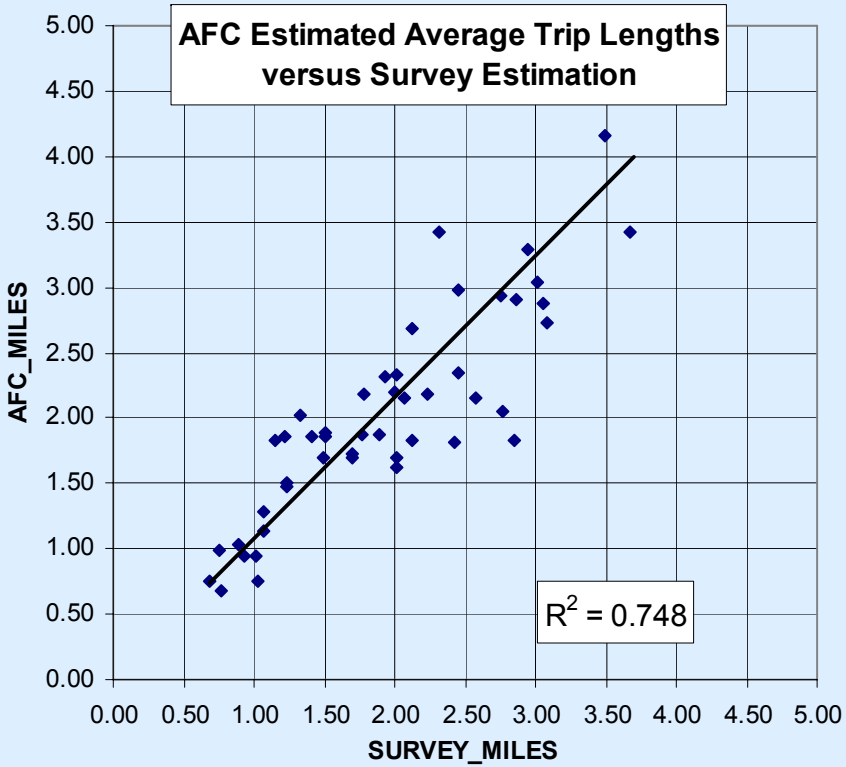
Non-Receipt Boardings on BX12 +**select**bus service

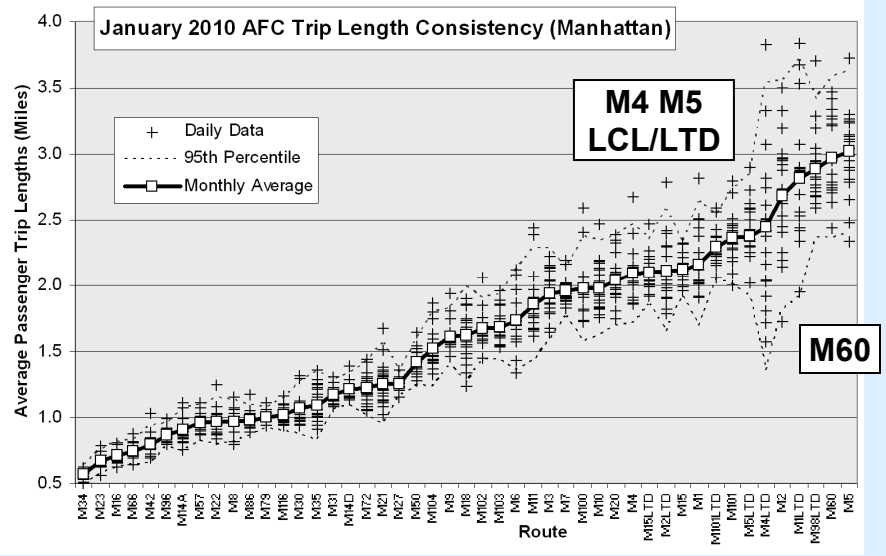
Category	Total Count	Rate	Notes
Fares Paid – MetroCard Validators	1,847		4. POP receipts are occasionally redeemed on BX12 local buses.
Fares Paid – Coin Fare Collectors	41		
Paid Passengers Boarding Local Service (Leakage)[4]	-7		
Passenger Registrations Observed	**1,881**	—	5. Includes Children under 44", and passengers with UniTickets (commuter railroad universal fare media, accepted only on feeder buses).
Front Door Entries	1,383		
Rear Door Entries	895		
Passenger Boardings Observed	**2,278**	—	
Boardings minus Registrations	397	17.4% ±2%	
Exempt (non-Receipt) Adjustment[5]	—	4.4%	
Rate of Unaccountable Boardings	—	**13.0% ±2%**	

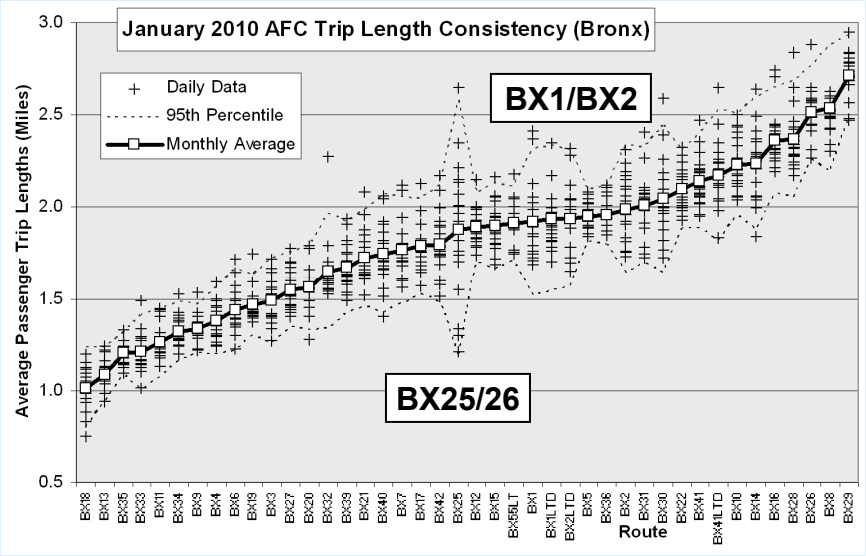
Survey Notes: Payment study conducted May~June 2009. 3.4% of non-receipt boardings were by Children Under 44" (requiring no receipts when accompanied by paying adults); about 1% used railroad universal farecard "UniTickets" (estimated from sales & usage data). "Unaccountable" boardings account for 17.4% – 4.4% = 13.0% of ridership. Non-receipt adjustment expansion factor is 13.0% ÷ (1 – 13.0%) = **14.9%.**

Parallel Testing

- **Route avg. trip lengths agree with ridechecks**
  - $R^2 = 0.748$ and Slope ~ 1:1
- **Trip lengths internally consistent across days**
  - Symmetry violations cause scattering in few routes







FTA Approval

- **Comparisons with 700 trips annual sample for 2007~09**
- **Provided list of AFC data advantages:**
  - Eliminates paper, data entry, attendance issues, “human” elements
  - 100% data, no sampling requires, eliminates sample fluctuations
- **Provided detailed documentation of analysis process**

PARALLEL TEST

Surveyor's (Sample) versus MetroCard AFC Data (100%)

Passenger Miles & Unlinked Trips -- NTD Bus (MB): 2007 - 2009

	2007		2008		2009	
	Surveyor	AFC Data	Surveyor	AFC Data	Surveyor	AFC Data
REVENUE RIDERSHIP	738,039,531		746,977,406		726,433,247	
% Change (from previous year)			1.21%		-2.75%	
UNLINKED TRIPS	862,630,526	863,838,154	902,640,956	868,638,444	906,529,603	842,865,961
% Change - MetroCard AFC to Surveyor Data		0.14%		-3.77%		-7.02%
% Change (from previous year)			4.64%	0.56%	0.43%	-2.97%
PASSENGER MILES	1,812,108,125	1,887,689,579	1,861,302,947	1,892,676,473	1,865,303,339	1,838,901,551
% Change - MetroCard AFC to Surveyor Data		4.17%		1.69%		-1.42%
% Change (from previous year)			2.71%	0.27%	0.21%	-2.84%
TRIP LENGTH (miles)	2.10	2.19	2.06	2.18	2.06	2.18
% Change - MetroCard AFC to Surveyor Data		4.03%		5.67%		6.03%
% Change (from previous year)			-1.84%	-0.29%	-0.21%	0.13%

Acknowledgements

- **Federal Transit Administration**
  - Gary Delorme
- **Savantage (National Transit Database)**
  - Jane Bailey
- **NYCT Colleagues in**
  - System Data & Research
  - Automated Fare Collection
  - Office of Management & Budget
- **TRB AP010 Committee's Anonymous Reviewers**