10-3-2014

Using "Big Data" for Transportation Analysis: A Case Study of the LA Metro Expo Line

Mohja L. Rhoads
South Bay Cities Council of Governments

Let us know how access to this document benefits you.

Follow this and additional works at: http://pdxscholar.library.pdx.edu/trec_seminar

Part of the Transportation Commons, and the Urban Studies and Planning Commons

Recommended Citation
http://pdxscholar.library.pdx.edu/trec_seminar/44

This Book is brought to you for free and open access. It has been accepted for inclusion in TREC Friday Seminar Series by an authorized administrator of PDXScholar. For more information, please contact pdxscholar@pdx.edu.
Using 'Big Data' for Transportation Analysis

A Case Study of the LA Metro Expo Line

Genevieve Giuliano
Sandip Chakrabarti
Mohja Rhoads

Sol Price School of Public Policy
University of Southern California
Presentation Outline

1. “Big” data, new opportunities

2. Case study application: The Expo (Phase 1) LRT
   - Impacts of LRT – what we know, and what we don’t
   - Research context, methods, data
   - Results on transit ridership impacts
   - Results on system performance impacts

3. Conclusions
“Big” Data, New Opportunities

Overview of the ADMS Research Project

Section 1
“Big” Data, New Opportunities

“Big Data”
- ITS + ICT = highly disaggregate data with respect to both time and space
- Examples: GPS trace data, roadway sensor data, accelerometers

New Opportunities
- Simulation model calibration
- Real-time transportation system management
- Travel behavior analysis
- Transport system analysis
ADMS: A database containing transportation system data from the LA Metro Region

Database structure created by USC Integrated Media Systems Center
- How to capture a large data stream (40Mb/sec)
- How to design a constantly growing database that can be queried efficiently

Database applications created by METRANS Transportation Center
The Value of ADMS

- **Multiple Modes**
  
  *Archive of historical highway, arterial and public transit system performance data*

- **Multiple Devices**
  
  *Roadway sensors, remote cameras, transit AVL and APC, etc.*

- **Multiple Agencies**
  
  *Caltrans, Metro, CHP, LADOT, etc.*

- **Multiple Data Types**
  
  *Traffic speeds/volumes/occupancies, incidents, transit supply and performance, etc.*
<table>
<thead>
<tr>
<th>Caltrans District 7, 8 and 12</th>
<th>Data Type</th>
<th>Data Attributes</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Friedway Detector Inventory and Real-Time Data</td>
<td>Routes</td>
<td>Varied: once per 30 second, minute, day; twice per year</td>
<td></td>
</tr>
<tr>
<td>Arterial Detector Inventory and Real Time Data</td>
<td>Cross Streets</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CMS Inventory and Real-Time Data</td>
<td>Directions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCTV Inventory and Real-Time Data</td>
<td>Occupancy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Event Data</td>
<td>Volumes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Freeway Travel Times</td>
<td>Speeds</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ramp Metering Inventory and Real-Time Data</td>
<td>Geo-locations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LADOT</td>
<td>Status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arterial Detector Inventory and Real-time Data</td>
<td>JPEG URLs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metro Bus</td>
<td>Start Time</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metro Rail</td>
<td>Clear Time</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHP</td>
<td>Vehicle and Route Data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Long Beach Transit</td>
<td>Event Data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Foothill Transit</td>
<td>Vehicle and Route Data</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
20 data feeds
- 2,000 Highway Sensors
- 4,700 Arterial Sensors
- Incident Reports
- CCTV Video Feeds
- Transit Vehicle GPS
- Transit Ridership Counts
- Highway Advisory Signs

Data Volume
- 7Mb per minute
- 3.7 Tb per year

ADMS is continuously expanding
The Value of ADMS

- Decade
- Year
- Month
- Minute
- Second

US Census Data
- HPMS Selected locations
- Aggregate over time

- Transit Ridership

ADMS
- X,Y Coord.
- Block
- Tract
- City
- County

Aggregate over space

Gasoline sales
Research Purpose

- To date, few applications in urban planning

- Our purpose: show how these new data sources can inform urban planning
  - Analyze impacts of capital investments, policy interventions
  - Make better investment, policy decisions
Case Study: The Expo (Phase 1) LRT

Section 2
Background, Context, Methods
Purpose of LRT investments
- Provide more and better public transport service
- Attract more transit patronage
- Reduce congestion, air pollution
- Promote more compact, sustainable urban form

What we know
- LRT may or may not generate more transit patronage
- LRT may or may not influence urban form

What we don’t know
- LRT impact on congestion, transportation system performance
Expo Phase 1 Route Alignment
The Expo Line was intended to:

- Improve access and mobility of residents and employees
- Provide an additional transit alternative through the corridor
- Increase transit mode share and alleviate congestion

I-10 (West) freeway operates at a level of service “F” for over three hours during each peak period with traffic volumes over 300,000 vehicles per day, and the Expo Line is key to congestion reduction

(Final EIS/EIR Report of the Mid-City/Exposition LRT Project, 2005)
Study Design

- Research questions:
  - Has the Expo Line generated more transit use?
  - Has it improved mobility and reliability of travel across its service corridor?

- Quasi-experimental design
  - Experimental/control corridor comparisons
  - Before/after comparisons
The Test Corridor – Experimental and Control Segments

Legend:
- Green: Highway Sensors
- Blue: Arterial Sensors
- Expo Line
- Red: Control Corridor(s)
- Green: Experimental Corridor
Data
Time periods for preliminary analyses:
  - “Before” period (pre-Expo): Nov 1, 2011 to Jan 31, 2012 (3 months)
  - “After” period (post-Expo): Nov 1, 2012 to Jan 31, 2013 (3 months)

Transit (bus and rail) data from Metro:
  - GIS data for lines and stops
  - Planned service/operations
  - Patronage (boardings and alightings by stop, trips by route)
  - System performance measures (e.g. on-time performance)

Sensor data:
  - Highway (I-10) sensors: 74 (Total 16 million+ records used for analysis)
  - Arterial sensors: 1066 (Total 180 million+ records used for analysis)
# Sensor Data Cleaning

<table>
<thead>
<tr>
<th>Details</th>
<th>Notes</th>
<th>Highway Sensors (Pre-Expo period)</th>
<th>Highway Sensors (Post-Expo period)</th>
<th>Arterial Sensors (Pre-Expo period)</th>
<th>Arterial Sensors (Post-Expo period)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of sensors in study area</td>
<td>N.A.</td>
<td>74</td>
<td>74</td>
<td>1066</td>
<td>1066</td>
</tr>
<tr>
<td>No. of records possible (ideal)</td>
<td>N.A.</td>
<td>19607040</td>
<td>19607040</td>
<td>282447360</td>
<td>282447360</td>
</tr>
<tr>
<td>No. of records actually available (from functional sensors)</td>
<td>N.A.</td>
<td>8449076</td>
<td>8273351</td>
<td>85922317</td>
<td>96040219</td>
</tr>
<tr>
<td>Records available after performing data cleaning</td>
<td>Filter for &quot;bad days&quot; and &quot;bad sensors&quot;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Filter for Speed</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Filter for Volume</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Filter for Occupancy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Additional filters for Speed and Occupancy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Store all records at this stage</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Filter for improbable Volume</td>
<td>8408345</td>
<td>8127632</td>
<td>82334911</td>
<td>90821083</td>
</tr>
<tr>
<td>Data available w.r.t to ideal scenario (as %)</td>
<td>N.A.</td>
<td>42.88</td>
<td>41.45</td>
<td>29.15</td>
<td>32.16</td>
</tr>
<tr>
<td>Data lost due to cleaning (as %)</td>
<td>N.A.</td>
<td>0.48</td>
<td>1.76</td>
<td>4.18</td>
<td>5.43</td>
</tr>
</tbody>
</table>
... and Aggregation

Unit: Each 15-minute time interval over a given period

<table>
<thead>
<tr>
<th>Details</th>
<th>Highway Sensors (Pre-Expo period)</th>
<th>Highway Sensors (Post-Expo period)</th>
<th>Arterial Sensors (Pre-Expo period)</th>
<th>Arterial Sensors (Post-Expo period)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of sensors in study area</td>
<td>74</td>
<td>74</td>
<td>1066</td>
<td>1066</td>
</tr>
<tr>
<td>No. of aggregated records possible (ideal)</td>
<td>653568</td>
<td>653568</td>
<td>9414912</td>
<td>9414912</td>
</tr>
<tr>
<td>No. of aggregated records actually available</td>
<td>421548</td>
<td>394853</td>
<td>7032939</td>
<td>8040335</td>
</tr>
<tr>
<td>% of missing data</td>
<td>35.50</td>
<td>39.59</td>
<td>25.30</td>
<td>14.60</td>
</tr>
</tbody>
</table>
Impacts of Expo Line on Corridor-level Transit Use

Has the Metro Expo Line (and associated transit service changes) had a significant impact on transit ridership/use within the line’s service corridor?
Research Framework

- **Analyses Performed**

1. Change in bus and rail boardings (average daily total for weekdays) at all stops and stations – before/after, experimental/control areas

2. Change in average weekday ridership of E-W Metro transit lines (connecting West LA with Downtown through the test corridor) across screenlines – before/after, experimental/control screenlines

3. Change in weekday peak-period (AM and PM) person throughput by Metro bus and rail across screenlines – before/after, experimental/control screenlines

4. Transfers at bus stops near Expo stations, and change in ridership of connecting (“feeder”) bus lines

*We have accounted for all Metro transit service changes during both shakeups*

*Note: All major service changes concentrated within the experimental area*
Transit Lines through the Corridor, and Screenlines

Rapid Bus
Local Non-CBD
Local CBD
Lim Exp
Com Circ
Expo Line
Control Corridor
Exp Corridor
Change in bus and rail boardings *(average daily total for weekdays)*, all routes, all stops and stations – before and after

<table>
<thead>
<tr>
<th>Mode</th>
<th>Experimental Area</th>
<th>Control Area (North)</th>
<th>Control Area (West)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bus</td>
<td>-5%</td>
<td>-6%</td>
<td>3%</td>
</tr>
<tr>
<td>Rail</td>
<td>NA</td>
<td>1%</td>
<td>NA</td>
</tr>
<tr>
<td>Total</td>
<td>6%</td>
<td>-4%</td>
<td>3%</td>
</tr>
</tbody>
</table>

**Observation:** Suggests positive impact on transit use
Change in **total weekday ridership** (patronage) of E-W transit lines **across screenlines** – before and after

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Experimental Area</th>
<th>Control Area (North)</th>
<th>Control Area (West)</th>
<th>All Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change in weekday RVH</td>
<td>4.3%</td>
<td>-1.0%</td>
<td>0.8%</td>
<td>0.1%</td>
</tr>
<tr>
<td>Change in patronage</td>
<td>7.2%</td>
<td>-3.3%</td>
<td>-1.4%</td>
<td>-2.2%</td>
</tr>
</tbody>
</table>

*Note: Red/Purple lines not included since they do not traverse across the Control (North) screenline*

Observation: Suggests positive impact on transit use; but RVH increased too
Change in **weekday peak-period (AM and PM)** person throughput by Metro bus and rail across screenlines, E&W directions combined – before/after

<table>
<thead>
<tr>
<th>Experimental Area</th>
<th>Control Area (North)</th>
<th>Control Area (West)</th>
<th>All Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>13.9%</td>
<td>-7.8%</td>
<td>-2.6%</td>
<td>-6.4%</td>
</tr>
</tbody>
</table>

*Note: Red/Purple lines not included since they do not traverse across the Control (North) screenline*

**Observation:** Suggests positive impact on transit use
Observation: Significant increase in transfers
Are increased transfers associated with higher ridership on the connecting ("feeder") bus lines?

<table>
<thead>
<tr>
<th>Type of Line</th>
<th>Pre-Expo Average Weekday Ridership</th>
<th>Post-Expo Average Weekday Ridership</th>
<th>Increase</th>
</tr>
</thead>
<tbody>
<tr>
<td>All feeder lines</td>
<td>161,079</td>
<td>156,531</td>
<td>-2.8%</td>
</tr>
<tr>
<td>Direct station connectors</td>
<td>140,627</td>
<td>142,582</td>
<td>1.4%</td>
</tr>
</tbody>
</table>

Note: Connections to Expo stations within experimental area are considered only

Observation: Some Expo riders are drawn from existing (bus) transit users.
The Expo LRT line seems to have had a positive impact on transit use within its service area.

Results indicate significant latent demand for high quality and reliable transit travel.

There is some evidence suggesting transit mode substitution (bus to LRT); mode shifts from auto to transit, although probable, may not be large enough to have any significant impact on corridor-level traffic.
Impacts of Expo Line on System Performance

Has the Metro Expo Line (and associated transit service changes) had a significant impact on freeway and arterial system performance?
Three different comparisons:
- Impacts on freeway (I-10 West)
- Impacts on Venice Blvd. (test arterial)
- Impacts on all major east-west arterials

Two system performance measures:
- Corridor Speed
- Travel Time Reliability

Peak periods:
- AM peak (7 am to 10 am)
- PM peak (4 pm to 7 pm)
DID Regression

The model for a generic member of any group is of the form:

\[ y = \beta_0 + \beta_1 G + \beta_2 T + \beta_3 G.T + \varepsilon \]

Where:

- \( y \) is the outcome of interest
- \( G \) is a dummy variable capturing differences between the two groups ("1" for "Experimental")
- \( T \) is a dummy variable for the second time period ("1" for the "post-Expo" period)
- \( G.T \) is therefore a dummy variable that assumes the value "1" for observations in the "Experimental" group in the "post-Expo" time period

The difference-in-differences estimate is:

\[ \widehat{\beta}_3 = (\overline{y}_{\text{Experimental,post-Expo}} - \overline{y}_{\text{Experimental,pre-Expo}}) - (\overline{y}_{\text{Control,post-Expo}} - \overline{y}_{\text{Control,pre-Expo}}) \]
“Experimental” and “Control” Segments

I-10 Freeway

Legend
- Highway Sensors
- Arterial Sensors
- Expo Line
- Control Corridor(s)
- Experimental Corridor
## Impact on Speed (Weekdays)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Group</th>
<th>Mean Diff (post vs. pre)</th>
<th>Mean Diff-in-Diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>AM Peak East</td>
<td>Experimental</td>
<td>2.02*</td>
<td>3.38**</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>-0.70</td>
<td></td>
</tr>
<tr>
<td>AM Peak West</td>
<td>Experimental</td>
<td>0.34</td>
<td>-3.83**</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>4.78**</td>
<td></td>
</tr>
<tr>
<td>PM Peak East</td>
<td>Experimental</td>
<td>0.10</td>
<td>-0.33</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>0.29</td>
<td></td>
</tr>
<tr>
<td>PM Peak West</td>
<td>Experimental</td>
<td>-1.27*</td>
<td>-2.65**</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>1.50**</td>
<td></td>
</tr>
</tbody>
</table>

* p<0.05; **p<0.01
Figures are in mph

**Observation:** No apparent impact on I-10
Impact on Speed (Weekdays)

Difference-in-Differences Regression Models of Average Weekday Freeway Speed

<table>
<thead>
<tr>
<th>Explanatory Variable</th>
<th>AM Peak (East)</th>
<th>AM Peak (West)</th>
<th>PM Peak (East)</th>
<th>PM Peak (West)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time (post-Expo=1)</td>
<td>-0.64</td>
<td>4.86***</td>
<td>0.31</td>
<td>1.54</td>
</tr>
<tr>
<td>Group (Experimental=1)</td>
<td>-13.50***</td>
<td>-4.64***</td>
<td>1.79**</td>
<td>1.97</td>
</tr>
<tr>
<td>Time x Group (DID estimator)</td>
<td>2.74</td>
<td>-4.52**</td>
<td>-0.20</td>
<td>-2.74</td>
</tr>
<tr>
<td>Constant</td>
<td>55.44***</td>
<td>37.40***</td>
<td>30.39***</td>
<td>50.60***</td>
</tr>
</tbody>
</table>

| N | 48 | 48 | 48 | 48 |
| R-square | 0.52 | 0.53 | 0.17 | 0.03 |

* p<0.05; **p<0.01; ***p<0.001
Figures are in mph

Observation: No apparent impact on I-10 Freeway
### Impact on Reliability (per mile buffer time; weekdays)

**I-10 Freeway**

<table>
<thead>
<tr>
<th>Peak Period</th>
<th>Direction</th>
<th>Mean Difference (post vs. pre)</th>
<th>Mean Diff-in-Diff</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Experimental Group</td>
<td>Control Group</td>
</tr>
<tr>
<td>AM Peak</td>
<td>East</td>
<td>0.12</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>West</td>
<td>0.89**</td>
<td>0.28**</td>
</tr>
<tr>
<td>PM Peak</td>
<td>East</td>
<td>-0.29*</td>
<td>-0.22*</td>
</tr>
<tr>
<td></td>
<td>West</td>
<td>0.10</td>
<td>0.26**</td>
</tr>
</tbody>
</table>

* p<0.05; **p<0.01

Figures are in min per mile

**Observation:** No apparent impact on I-10
“Experimental” and “Control” Segments Analyzed

- **Venice Blvd.**

- **Control Segment**
- **Experimental Segment**
### Impact on Speed (Venice Blvd. example; weekdays)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Group</th>
<th>Mean Diff (post vs. pre)</th>
<th>Mean Diff-in-Diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>AM Peak East</td>
<td>Experimental</td>
<td>0.46**</td>
<td>0.81**</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>-0.25</td>
<td></td>
</tr>
<tr>
<td>AM Peak West</td>
<td>Experimental</td>
<td>0.60**</td>
<td>8.60**</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>-7.78**</td>
<td></td>
</tr>
<tr>
<td>PM Peak East</td>
<td>Experimental</td>
<td>-0.36*</td>
<td>-1.06**</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>0.71**</td>
<td></td>
</tr>
<tr>
<td>PM Peak West</td>
<td>Experimental</td>
<td>1.00**</td>
<td>8.81**</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>-7.69**</td>
<td></td>
</tr>
</tbody>
</table>

* p<0.05; **p<0.01
Figures in mph

**Observation:** Suggests positive impact, but big changes in control segments suspect
Impact on Speed (Venice Blvd. example; weekdays)

Difference-in-Differences Regression Models of Average Weekday Speed (Venice)

<table>
<thead>
<tr>
<th>Explanatory Variable</th>
<th>AM Peak (East)</th>
<th>AM Peak (West)</th>
<th>PM Peak (East)</th>
<th>PM Peak (West)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time (post-Expo=1)</td>
<td>-0.21</td>
<td>-7.75***</td>
<td>0.73</td>
<td>-7.69***</td>
</tr>
<tr>
<td>Group (Experimental=1)</td>
<td>1.67</td>
<td>-4.74***</td>
<td>3.91***</td>
<td>1.67***</td>
</tr>
<tr>
<td>Time x Group (DID estimator)</td>
<td>0.69</td>
<td>8.35***</td>
<td>-1.08</td>
<td>8.69***</td>
</tr>
<tr>
<td>Constant</td>
<td>27.27***</td>
<td>30.43***</td>
<td>19.48***</td>
<td>25.99***</td>
</tr>
<tr>
<td>N</td>
<td>48</td>
<td>48</td>
<td>48</td>
<td>48</td>
</tr>
<tr>
<td>R-square</td>
<td>0.13</td>
<td>0.54</td>
<td>0.68</td>
<td>0.92</td>
</tr>
</tbody>
</table>

* p<0.05; **p<0.01; ***p<0.001
Figures are in mph

Observation: Suggests positive impact, but big changes in control segments suspect
## Impact on Reliability (per mile buffer time; weekdays)

<table>
<thead>
<tr>
<th>Peak Period</th>
<th>Direction</th>
<th>Mean Difference (post vs. pre)</th>
<th>Mean Diff-in-Diff</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Experimental Group</td>
<td>Control Group</td>
</tr>
<tr>
<td>AM Peak</td>
<td>East</td>
<td>-0.06</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>West</td>
<td>-0.16**</td>
<td>1.75**</td>
</tr>
<tr>
<td>PM Peak</td>
<td>East</td>
<td>0.04</td>
<td>-0.33</td>
</tr>
<tr>
<td></td>
<td>West</td>
<td>-0.06</td>
<td>1.35**</td>
</tr>
</tbody>
</table>

* p<0.05; **p<0.01
Figures are in min per mile

**Observation:** Suggests positive impact, but big changes in control segments suspect
<table>
<thead>
<tr>
<th>Arterial</th>
<th>AM Peak East</th>
<th>AM Peak West</th>
<th>PM Peak East</th>
<th>PM Peak West</th>
</tr>
</thead>
<tbody>
<tr>
<td>W 3rd (control)</td>
<td>-0.81**</td>
<td>-0.91**</td>
<td>-0.18</td>
<td>0.30*</td>
</tr>
<tr>
<td>Olympic</td>
<td>-1.04**</td>
<td>-0.48*</td>
<td>-0.91**</td>
<td>1.41**</td>
</tr>
<tr>
<td>Pico</td>
<td>0.78**</td>
<td>-0.19</td>
<td>2.19**</td>
<td>-0.13</td>
</tr>
<tr>
<td>Venice</td>
<td>0.46**</td>
<td>0.60**</td>
<td>-0.36*</td>
<td>1.00**</td>
</tr>
<tr>
<td>Washington</td>
<td>1.50**</td>
<td>-0.56**</td>
<td>1.15**</td>
<td>0.77**</td>
</tr>
<tr>
<td><strong>EXPO LINE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jefferson</td>
<td>1.14**</td>
<td>1.06**</td>
<td>1.18**</td>
<td>3.15**</td>
</tr>
<tr>
<td>ML King</td>
<td>3.47**</td>
<td>1.53**</td>
<td>2.33**</td>
<td>0.99**</td>
</tr>
</tbody>
</table>

* p<0.05; **p<0.01
Figures in mph

Observation: Suggests positive impact closer to Expo Line, but no data for control segments
Conclusions on Expo Line Impacts

- **Transit ridership**
  - Evidence for positive impact – net increase in transit use, but service increased too
  - Shift from bus to rail, but also new trips/riders – some evidence for latent demand for high quality transit travel

- **System performance**
  - No impact on 1-10 – very large traffic volumes swamp any possible effect of Expo Line
  - Some evidence for improved performance along arterials nearest the Expo Line
    - Can’t be attributed to less bus service
    - Supports attracting at least some riders from private vehicles
We have demonstrated how “big” transportation system data can be used for analyzing impacts (transit use + system performance) of a new infrastructure investment.

Data limitations exist, especially for arterials – we need:
- better instrumentation (more sensors + more working sensors)
- better inter-agency coordination (institutional + technical + operational)

Our story is incomplete – for example, we cannot measure/compare person throughput across the corridor (no automobile occupancy data)
However, we have developed a framework, identified performance indicators, and generated baseline measures for monitoring Expo impacts on the multi-modal I-10 (W) corridor over time…
Conclusions on using “Big” data
Conclusions on using “big” data

- Has potential to improve monitoring and analysis of major projects
  - Highly detailed – across space and time
  - Reduces cost of performance monitoring – we could do the same analysis for 6 months, or whole year, or another 3 month panel, or…….

- Big data is only as good as what is generated at the source
  - Problems of missing data, unreliable data
  - Not all critical or interesting data are collected

- Using it effectively requires programming and database skills
Acknowledgements

LA Metro, for funding the ADMS research project

Cyrus Shahabi, USC IMSC
Ugur Demiryurek, USC IMSC
Udit Agrawal, USC Vitebi
THANK YOU