FIXED EFFECTS PANEL DATA ANALYSIS OF GASOLINE PRICES, FARE, SERVICE SUPPLY, AND SERVICE FREQUENCY ON TRANSIT RIDERSHIP IN TEN U.S. URBANIZED AREAS

the 54th Annual Conference of the Association of Collegiate School of Planning (ACSP) in Philadelphia, Pennsylvania
November 2nd, 2014

Hiroyuki Iseki, Ph.D.
Assistant Professor
Urban Studies and Planning Program (URSP)

Rubaba Ali
Ph.D. Student
Department of Agricultural and Resource Economics

National Center for Smart Growth, Research and Education
University of Maryland, College Park
Introduction

• Substantial fluctuation in gasoline prices between 1999 and 2011.
• In July, 2008,
  – $4 per gallon
  – 3 to 5% increase from 2007
• 10.7 billion transit trips
• A decline of driving by 56 billion VMT (1.9%) in 2007-08
• In April, 2011, $4.
• Q. What affects transit ridership and to what degree?
  – with a focus on the effects of gasoline prices
• Introduction
• Literature review on the determinants of transit ridership
• Data, data sources, and methodology
• Analysis results
• Concluding remarks
Literature Review

• A large number of studies on the subject
• Substantial variation in terms of data and methods:
  – location of study
  – geographic scale used as a unit of analysis
  – mode of transit: single mode vs. multiple modes (aggregate)
  – trip characteristic (purpose, distance, etc.)
  – simple elasticity calculation of “before & after”
  – regression studies
    • Cross-sectional analysis
    • Time-series analysis
    • Panel-data analysis (Mattson 2008; Blanchard 2009)
• Transferability and generalizability
Data and Methodology

- **Dependent Variables:** Total urbanized area (UA) ridership
- **Mode:** 1) bus, 2) commuter rail (CR), 3) light rail (LR), 4) heavy rail (HR), 5) the aggregate
- **Data Sources:**
  - Transit operating and financial data: National Transit Database
  - Monthly gasoline prices: weekly data from the U.S. Energy Information Administration
  - Demographics: U.S. Census 2000 and American Community Survey (ACS) 1-year estimates between 2005 and 2011
  - Economic variables: Bureau of Labor Statistics
- **Unit of analysis:** month-UA
- **Variable Construction:** mostly Log-Log form with the percentage of HHs without cars and employment rate (%) as exceptions
- **Panel data:** Max. 120 months (2002-2010) * 10 UAs
Data and Methodology

- Fixed-effects Panel Data Analysis:
  - simultaneously takes into account temporal and cross-sectional variation to obtain more robust, generalizable results (Greene, 2012)

- Control for factors that potentially affect transit ridership
  - **Internal factors**: 1) average fare, 2) service supply (VRH), and 3) service frequency
  - **External factors**: factors outside the control of transit agencies—4) gasoline prices, 5) regional economy, 6) demographic changes, and 7) changes in highway infrastructure
Simultaneity and Endogeneity

- Potential endogeneity bias on estimated coefficients that could arise from the simultaneity between transit supply and consumption (bidirectional causality):
  1. The level of transit service supplied directly influences the consumption of transit trips.
  2. The level of transit service consumption (ridership) can affect the supply of transit service, as transit agencies adjust it within financial constraints in response to the ridership level.

- Addressing this problem using the instrumental variables (IV) method
Data and Methodology

- **Baseline Specification Model**
  \[
  R_{it}^M = \alpha_0 + y_{it}^M \alpha + \delta P_{it} + X_{it}' \beta + \mu_i + \eta_t + \kappa_y + \epsilon_{it}^M
  \]  

- **Instrumental Variables Model**
  \[
  y_{it}^M = \alpha_1 + Z_{it}' \gamma + \delta_1 P_{it} + X_{it}' \theta + \tilde{y}_{it}' \pi + \mu_i + \eta_t + \kappa_y + u_{it}^M
  \]
  \[
  R_{it}^M = \alpha_0 + \hat{y}_{it}' \alpha + \delta P_{it} + X_{it}' \beta + \tilde{y}_{it}' \omega + \mu_i + \eta_t + \kappa_y + \epsilon_{it}^M
  \]

- \(i\): urbanized area (UA), \(t\): time; \(M\): transit mode; \(P_{it}\): the price of gasoline
- \(R_{it}^M\): transit ridership in urbanized area (UA)
- \(X_{it}'\): a vector of external influential factors
- \(y_{it}^M\): a vector of internal influential factors in Eq. (1)
- \(y_{it}^M\): the supply of transit service measured as VRH in Eq. (2)
- \(\hat{y}_{it}'\): the predicted value of transit supply from the first stage.
- \(\tilde{y}_{it}'\): the vector of variables of transit fare and service frequency
- \(Z_{it}'\): the vector of excluded instruments that affect only transit supply but do not affect ridership
- \(\mu_i, \eta_t, \kappa_y\): fixed effects for UAs, monthly/yearly dummy variables
- \(\epsilon_{it}, u_{it}\): stochastic error terms
• **First Stage**: Predict Service Supply using an array of independent variables. *Instrumental variables* are:
  - total number of employees
  - total fleet: the total seating & standing capacity of transit vehicles
  - total funds for transit agencies in each UA in a particular year, combining local, state and federal funds
  - Two conditions as valid instruments.
    1. “Instrument relevance”
    2. “Instrument exogeneity”

• **Second Stage**: Predict Service Consumption *(ridership)*, using variables, including a variable to predict service supply estimated in the first stage
Data:
Distribution of Observations across Urbanized Areas

- New York City: account for approximately 40% of the nation’s fixed-route transit trips and comprises a significantly different transit environment, compared to other regions.
- “with/without NY” made difference for heavy rail (HR).

<table>
<thead>
<tr>
<th>Urbanized Area</th>
<th>Proportion of Observations from Each UAZ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>… / Mode</td>
</tr>
<tr>
<td>Boston, MA</td>
<td></td>
</tr>
<tr>
<td>Chicago, IL</td>
<td></td>
</tr>
<tr>
<td>Cleveland, OH</td>
<td></td>
</tr>
<tr>
<td>Denver, CO</td>
<td></td>
</tr>
<tr>
<td>Houston, TX</td>
<td></td>
</tr>
<tr>
<td>Los Angeles, CA</td>
<td></td>
</tr>
<tr>
<td>Miami, FL</td>
<td></td>
</tr>
<tr>
<td>New York, NY</td>
<td></td>
</tr>
<tr>
<td>San Francisco, CA</td>
<td></td>
</tr>
<tr>
<td>Seattle, WA</td>
<td></td>
</tr>
<tr>
<td>Total Num. of Observations</td>
<td></td>
</tr>
</tbody>
</table>
Analysis Results:
The First Stage of the Instrumental Variables Model

- F-stat/P-value rejects the hypothesis that the coefficients of the three IVs are jointly zero. (relevance)
  - Total number of employees (-) for bus and HR w/o NY
  - Total fleet (+) for all modes but CR
  - Total fund available (+)

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>(1) Bus</th>
<th>(2) CR</th>
<th>(3) LR</th>
<th>(4-1) HR w/o NY</th>
<th>(4-2) HR w/ NY</th>
<th>(5) Transit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log of total number of employees (full time+part-time/2)</td>
<td>-0.0776***</td>
<td>-0.0354</td>
<td>0.0639</td>
<td>-0.445***</td>
<td>-0.321</td>
<td>-0.00508</td>
</tr>
<tr>
<td>(0.0191)</td>
<td>(0.194)</td>
<td>(0.0451)</td>
<td>(0.110)</td>
<td>(0.4295)</td>
<td>(0.0317)</td>
<td></td>
</tr>
<tr>
<td>Log of total fleet (seating+standing capacity)</td>
<td>0.552***</td>
<td>-0.0565***</td>
<td>0.443***</td>
<td>3.050***</td>
<td>5.933***</td>
<td>0.499***</td>
</tr>
<tr>
<td>(0.0296)</td>
<td>(0.0154)</td>
<td>(0.0159)</td>
<td>(0.417)</td>
<td>(1.659)</td>
<td>(0.0530)</td>
<td></td>
</tr>
<tr>
<td>Log of total fund available to transit agencies</td>
<td>0.0547***</td>
<td>-0.0286</td>
<td>0.0939***</td>
<td>0.174***</td>
<td>0.728***</td>
<td>0.0352**</td>
</tr>
<tr>
<td>(0.00988)</td>
<td>(0.0538)</td>
<td>(0.0243)</td>
<td>(0.0313)</td>
<td>(0.124)</td>
<td>(0.0162)</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>1,126</td>
<td>777</td>
<td>840</td>
<td>669</td>
<td>789</td>
<td>1,132</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.529</td>
<td>0.581</td>
<td>0.829</td>
<td>0.444</td>
<td>0.348</td>
<td>0.300</td>
</tr>
<tr>
<td>Number of Urbanized Areas</td>
<td>10</td>
<td>7</td>
<td>9</td>
<td>6</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>F-stat</td>
<td>123.56</td>
<td>4.93</td>
<td>278.67</td>
<td>47.82</td>
<td>20.27</td>
<td>31.34</td>
</tr>
<tr>
<td>P-value</td>
<td>0.000</td>
<td>0.0022</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Dummies for months, years, and UA fixed effects included
Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1
Analysis Results: the Baseline Specification (BS) and Instrumental Variables (IV) Models

- Service supply (VRH)
  - Positive
  - BL’s < IV’s
  - 0.3-1.14: in line with those found in (Litman 2004; Taylor et al. 2009, Chen, Varley, & Chen 2010)

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>(1) Bus</th>
<th></th>
<th>(2) CR</th>
<th></th>
<th>(3) LR</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BL</td>
<td>IV</td>
<td>BL</td>
<td>IV</td>
<td>BL</td>
<td>IV</td>
</tr>
<tr>
<td>Log of vehicle revenue hours</td>
<td>0.263***</td>
<td>0.407***</td>
<td>0.283***</td>
<td>0.0625</td>
<td>0.786***</td>
<td>0.901***</td>
</tr>
<tr>
<td></td>
<td>(0.0258)</td>
<td>(0.0519)</td>
<td>(0.0197)</td>
<td>(0.152)</td>
<td>(0.0221)</td>
<td>(0.0314)</td>
</tr>
<tr>
<td>(4-1) HR w/o NY</td>
<td></td>
<td></td>
<td>(4-2) HR w/ NY</td>
<td></td>
<td>(5) Transit</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BL</td>
<td>IV</td>
<td>BL</td>
<td>IV</td>
<td>BL</td>
<td>IV</td>
</tr>
<tr>
<td></td>
<td>0.305***</td>
<td>0.110</td>
<td>0.00449</td>
<td>0.0139</td>
<td>0.166***</td>
<td>0.577***</td>
</tr>
<tr>
<td></td>
<td>(0.0366)</td>
<td>(0.0922)</td>
<td>(0.00815)</td>
<td>(0.0303)</td>
<td>(0.0157)</td>
<td>(0.0712)</td>
</tr>
</tbody>
</table>

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1
Analysis Results: the Baseline Specification (BS) and Instrumental Variables (IV) Models

- Very similar results between the two models
- Fare: negative (-)  Gasoline prices: positive (+)

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>(1) Bus BL</th>
<th></th>
<th>(1) Bus IV</th>
<th></th>
<th>(2) CR BL</th>
<th></th>
<th>(2) CR IV</th>
<th></th>
<th>(3) LR BL</th>
<th></th>
<th>(3) LR IV</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Log of fare</td>
<td>-0.230***</td>
<td>(0.0207)</td>
<td>-0.220***</td>
<td>(0.0212)</td>
<td>-0.372***</td>
<td>(0.0616)</td>
<td>-0.444***</td>
<td>(0.0940)</td>
<td>-0.124***</td>
<td>(0.0348)</td>
<td>-0.141***</td>
<td>(0.0355)</td>
</tr>
<tr>
<td>Log of monthly gasoline price</td>
<td>0.0611***</td>
<td>(0.0219)</td>
<td>0.0617***</td>
<td>(0.0223)</td>
<td>0.0547</td>
<td>(0.0387)</td>
<td>0.0640</td>
<td>(0.0428)</td>
<td>0.0330</td>
<td>(0.0507)</td>
<td>0.0253</td>
<td>(0.0516)</td>
</tr>
<tr>
<td>(4-1) HR w/o NY</td>
<td>-0.284***</td>
<td>(0.0324)</td>
<td>-0.263***</td>
<td>(0.0343)</td>
<td>-0.210***</td>
<td>(0.028)</td>
<td>-0.206***</td>
<td>(0.0304)</td>
<td>-0.340***</td>
<td>(0.0240)</td>
<td>-0.319***</td>
<td>(0.0308)</td>
</tr>
<tr>
<td>(4-2) HR w/ NY</td>
<td>-0.0217</td>
<td>(0.0436)</td>
<td>-0.028</td>
<td>(0.0446)</td>
<td>-0.0276</td>
<td>(0.0400)</td>
<td>-0.0263</td>
<td>(0.0403)</td>
<td>0.0494**</td>
<td>(0.0205)</td>
<td>0.0573**</td>
<td>(0.0262)</td>
</tr>
</tbody>
</table>

- Gas prices: **rail** modes shows statistically insignificant results
- Zero-car households (captive & choice)
- CR: Low sensitivity of commute trips

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1
Analysis Results: the Baseline Specification (BS) and Instrumental Variables (IV) Models

- Service frequency
  - Positive
  - Only for (1) bus and (3) LR
  - Little variance in this variable for rail modes, especially for (2) CR

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>(1) Bus</th>
<th>(2) CR</th>
<th>(3) LR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log of frequency of service</td>
<td>0.115***</td>
<td>0.0772***</td>
<td>0.0998***</td>
</tr>
<tr>
<td></td>
<td>(0.0167)</td>
<td>(0.0207)</td>
<td>(0.0273)</td>
</tr>
<tr>
<td>(4-1) HR w/o NY</td>
<td></td>
<td></td>
<td>0.0130</td>
</tr>
<tr>
<td>(4-2) HR w/ NY</td>
<td></td>
<td></td>
<td>(0.0324)</td>
</tr>
</tbody>
</table>

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1
Relative Roles: Changes in *Within R*-squared using the Baseline Specification

- Explanatory power of each variable or a set of variables in concert

<table>
<thead>
<tr>
<th></th>
<th>Bus</th>
<th></th>
<th></th>
<th>Transit</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R^2</td>
<td>Change in R^2</td>
<td>% change</td>
<td>R^2</td>
<td>Change in R^2</td>
<td>% change</td>
</tr>
<tr>
<td>(1) All variables in the Baseline Specification</td>
<td>0.518</td>
<td>-</td>
<td>-</td>
<td>0.530</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Internal Factors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2) (1) Fare</td>
<td>0.463</td>
<td>-0.055</td>
<td>11%</td>
<td>0.444</td>
<td>-0.086</td>
<td>16%</td>
</tr>
<tr>
<td>(3) (1) Service Frequency</td>
<td>0.497</td>
<td>-0.021</td>
<td>4%</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(4) (1) Service Supply (VRH)</td>
<td>0.472</td>
<td>-0.046</td>
<td>9%</td>
<td>0.482</td>
<td>-0.048</td>
<td>9%</td>
</tr>
<tr>
<td>(5) (1) Fare - Service Frequency</td>
<td>0.447</td>
<td>-0.071</td>
<td>14%</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(6) (1) Fare - VRH</td>
<td>0.407</td>
<td>-0.111</td>
<td>21%</td>
<td>0.391</td>
<td>-0.138</td>
<td>26%</td>
</tr>
<tr>
<td>(7) (1) Service Frequency - VRH</td>
<td>0.406</td>
<td>-0.112</td>
<td>22%</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(8) (1) Fare - Service Frequency - VRH</td>
<td>0.348</td>
<td>-0.170</td>
<td>33%</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>External Factors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(9) (1) Gasoline Prices</td>
<td>0.514</td>
<td>-0.003</td>
<td>1%</td>
<td>0.527</td>
<td>-0.002</td>
<td>0%</td>
</tr>
<tr>
<td>(10) (1) Gasoline Prices - Other external variables</td>
<td>0.460</td>
<td>-0.058</td>
<td>11%</td>
<td>0.457</td>
<td>-0.073</td>
<td>14%</td>
</tr>
<tr>
<td>Neither Internal nor External Factors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(11) Only monthly, yearly, and urbanized area dummies</td>
<td>0.249</td>
<td>-0.268</td>
<td>52%</td>
<td>0.301</td>
<td>-0.228</td>
<td>43%</td>
</tr>
</tbody>
</table>
Changes in Ridership

• **Bus:**
  1. A ridership increase by 21% by changing average fare from the 95th percentile ($1.26) to the 5th ($0.54)
  2. A ridership increase by 42% by “1”
     + [A change in service frequency from the 40th percentile (937) to the 60th (1,108)]
     + [A change in VRH from the 40th percentile (322,072) to the 60th (354,072)]

• **Aggregate transit ridership:**
  3. A ridership increase by 38% by a fare change from $1.40 to $0.54
  4. A ridership increase by 56% when “3” is combined with the change in VRH from the 40th percentile (393,887) to the 60th (798,228).
Summary of Findings

• The *instrumental variables* method did not affect most of the estimated coefficients except VRH.
  – BL’s < IV’s for bus, the aggregate, and LR
• The magnitude of fare elasticity > gasoline prices
  – Direct elasticity vs cross elasticity
• The elasticities to service frequency and gasoline prices are significant for bus and the aggregate but not for rails.
  – More detailed studies: service/catchment area
• Changes in three internal factors in concert could increase ridership substantially.
Concluding Remarks

• A sudden increase in ridership due to gasoline price hike can have substantial negative impacts on transit systems in urbanized areas.

• A comprehensive understanding of elasticity of ridership for different transit modes is important in long term to guide transit agencies for:
  – pricing strategies,
  – capacity management, and
  – supply of transit services.
Acknowledgement

• This study has been supported by the funding from Mineta Transportation Institute at San Jose State University.
  – Project Number 1106: Net Effects of Gas Price Changes on Transit Ridership in US Urban Areas

• Graduate Assistants at San Jose State University: Julie Huang and Aliza Paz
Hiro Iseki
hiseki@umd.edu

National Center for Smart Growth Research & Education
University of Maryland, College Park

web site: http://www.smartgrowth.umd.edu/