How Much Does the Method Influence the Answer?
A Ceteris Paribus Empirical Comparison of Approaches for Assessing the Impact of Self-selection on Travel Behavior

Patricia L. Mokhtarian, Georgia Tech
patmokh@gatech.edu

David van Herick, UC Davis
dmvanherick@ucdavis.edu

World Symposium on Transport and Land Use Research
June 24-27, 2014, Delft, the Netherlands

Overview

- Motivation
  - Introduction to the residential self-selection (RSS) problem
  - Key empirical findings to date
- Research questions of interest
- Empirical context
- Methodology(ies)
- Initial results
- Summary and discussion
Overview

- Motivation
  - Introduction to the residential self-selection (RSS) problem
  - Key empirical findings to date
- Research questions of interest
- Empirical context
- Methodology(ies)
- Initial results
- Summary and discussion
Motivation: Land use as VKT-reduction policy tool

- Land use (LU) policies are receiving increased attention, as a way to (potentially) reduce
  - vehicle-km traveled (VKT), & thus
  - congestion
  - air pollution
  - energy consumption
  - obesity
  - greenhouse gas (GHG) emissions

Land use clearly matters to TB

- Many studies have compared travel behavior (TB) of residents of “urban” or “traditional” versus “suburban” neighborhoods, and found that *urban dwellers walk more and drive less than suburban dwellers*, supporting the rationale for more compact urban forms.
What’s the problem?

- Are the observed TB differences because of
  - a true independent influence of the built environment (BE)?

  or because
  - people who like walking (or, who prefer to drive) choose to live in neighborhoods supportive of that desire (AT)?

  or
  - some of both?

Kitamura et al. (1997)

What difference does it make?

- Suppose the effect of the BE on TB is primarily due to attitudinal predispositions (AT)

- Then if a “car-lover” lands in an urban neighborhood for other reasons (e.g. financial policy incentives), s/he may still drive like the typical suburban dweller

- If so, then policies promoting denser, more diverse land use patterns may not have the effect expected on the basis of studies that did not correct for self-selection
Let us stipulate…

- Public policies (zoning, mortgage interest deductions, etc.) have distorted the markets
- “New urban” housing is undersupplied
- Preferences are changing, at least to some extent
- There are excellent reasons (options for the mobility-limited, promoting physical activity, meeting consumer demand) other than a reduction of VKT for increasing new urban development

Levine et al. (2012)

Then what’s the big issue?

- Just the more narrowly-defined question: DO such LU policies produce the transportation benefit that constitutes one of their major selling points?
  - In California, specific targets for VKT reduction have been set for “Sustainable Community Strategies” to meet
    » E.g., for San Francisco, from 2005 baseline: 7% per capita GHG reduction by 2020, 15% by 2035
  - There are opportunity costs of being wrong about how effective these policies will be
    » Time, money, & political capital could have been spent on more useful policies
  - There are potential direct costs of increasing density
    » Less satisfaction, less privacy, less children’s play space/green space, congestion, tensions, contagion
Then what’s the big issue? (cont’d)

- Just the more narrowly-defined question: DO such LU policies produce the transportation benefit that constitutes one of their major selling points?
- Just the (probably naïve) belief that public policies should be promoted on the basis of actual benefits, not desired ones
- Thus, to evaluate the transportation effectiveness of (proposed) LU policies, it’s important to know the relative roles of BE and AT in influencing TB
- Self-selection arises in many other policy contexts as well

To illustrate...

BUILT ENVIRONMENT MATTERS: Among people with the same attitude, those living in traditional neighborhoods walked more often than suburban dwellers.

Handy *et al.* (2006)
ATTITUDE MATTERS: Among people living in the same type of neighborhood, those who consider having nearby shops to be very important walked (~4x) more often than those who don’t.

THE COMBINED EFFECT: Suburban dwellers who considered nearby stores important walked more often than traditional neighborhood residents who didn’t.
A definition

*Self-selection* exists when

people are not randomly-distributed into conditions (residential location in our case) relevant to an outcome of interest (TB in our case), but rather place themselves into the condition conducive to producing an outcome they desire

- And effects for them will differ (on average) from those for a randomly-selected person placed in the same condition

Nine approaches for addressing self-selection

1. direct questioning
2. statistical controls (SC)
3. instrumental variables models
4. propensity score models (PS)
5. sample selection models (SS)
6. joint discrete choice models
7. structural equations models
8. mutually-dependent discrete choice models
9. longitudinal designs

Cao et al. (2008, 2009); Mokhtarian & Cao (2008); many others
What does the prior empirical evidence show?

- Intriguing observation
  - The share of total BE influence on TB that’s “true” rather than due to RSS *varies widely across studies*

### Survey of literature

<table>
<thead>
<tr>
<th>BE, RSS Effects Quantified</th>
<th>BE, RSS Effects Not Quantified</th>
<th>No conclusion</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>100% RSS ≈ 50% BE &lt; RSS</td>
<td>BE &gt; RSS</td>
<td></td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>BE ≈ RSS</td>
<td></td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>BE &gt; RSS</td>
<td></td>
<td>13</td>
</tr>
<tr>
<td>Statistical Control</td>
<td></td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>Joint Discrete Choice</td>
<td></td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>Structural Eq. Modeling</td>
<td></td>
<td></td>
<td>59</td>
</tr>
<tr>
<td>Sample Selection</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Propensity Scores</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Although effects were not quantified, they were concluded to be approximately equal*
Although effects were not quantified, they were concluded to be approximately equal.

### Survey of literature

<table>
<thead>
<tr>
<th>BE, RSS Effects Quantified</th>
<th>BE, RSS Effects Not Quantified</th>
</tr>
</thead>
<tbody>
<tr>
<td>100% RSS = 50% 0%</td>
<td>BE &lt; RSS BE ≥ RSS BE &gt; RSS No conclusion Total</td>
</tr>
<tr>
<td>Statistical Control</td>
<td>0 0 0 2 1 0 8 12 23</td>
</tr>
<tr>
<td>Joint Discrete Choice</td>
<td>0 0 0 1 1 0 0 7 9</td>
</tr>
<tr>
<td>Structural Eq. Modeling</td>
<td>0 0 0 2 1 1 3 6 13</td>
</tr>
<tr>
<td>Sample Selection</td>
<td>0 0 1 2 1 0 0 2 6</td>
</tr>
<tr>
<td>Propensity Scores</td>
<td>0 1 1 5 1 0 0 0 8</td>
</tr>
<tr>
<td>Total</td>
<td>0 1 2 8 7 2 1 11 27 59</td>
</tr>
</tbody>
</table>

*Although effects were not quantified, they were concluded to be approximately equal.*

### Studies that quantify RSS effect (1)

<table>
<thead>
<tr>
<th>Study</th>
<th>TB variable</th>
<th>% of influence of BE due to RSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistical Control</td>
<td>Joh et al., 2012 # of walking trips</td>
<td>0%</td>
</tr>
<tr>
<td>Joint Discrete Choice</td>
<td>Larco et al., 2012 (#, %) of trips by walking, biking, driving</td>
<td>0%</td>
</tr>
<tr>
<td>Joint Discrete Choice</td>
<td>Salon, 2006 Walking level: none, some, a lot</td>
<td>33-50%</td>
</tr>
<tr>
<td>Joint Discrete Choice</td>
<td>Bhat and Guo, 2007 Auto ownership</td>
<td>0%</td>
</tr>
</tbody>
</table>
### Studies that quantify RSS effect (2)

<table>
<thead>
<tr>
<th>Study</th>
<th>TB variable</th>
<th>% of influence of BE due to RSS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Structural Eq. Models</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bhat and Eluru, 2009</td>
<td>Vehicle miles traveled</td>
<td>49%</td>
</tr>
<tr>
<td>Cao, 2009</td>
<td>Vehicle miles driven</td>
<td>24%</td>
</tr>
<tr>
<td>Zhou and Kockelman, 2008</td>
<td>Vehicle miles traveled</td>
<td>10-42%</td>
</tr>
<tr>
<td>Greenwald, 2003</td>
<td>Various substitution rates (by mode)</td>
<td>0%</td>
</tr>
<tr>
<td><strong>Sample Selection Models</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cao and Fan, 2012</td>
<td>Personal miles traveled</td>
<td>28%</td>
</tr>
<tr>
<td></td>
<td>Transit duration</td>
<td>49%</td>
</tr>
<tr>
<td></td>
<td>Driving duration</td>
<td>64%</td>
</tr>
<tr>
<td>Cao et al., 2010</td>
<td>Vehicle miles driven</td>
<td>15-24%</td>
</tr>
<tr>
<td>Cao, 2010</td>
<td>Strolling frequency</td>
<td>14%</td>
</tr>
<tr>
<td></td>
<td>Vehicle miles driven</td>
<td>22%</td>
</tr>
<tr>
<td></td>
<td>Walking to store frequency</td>
<td>39%</td>
</tr>
</tbody>
</table>
What are some sources of these (big) differences?

- **Genuine** variations in outcome across contexts (different locations, times, but also different dependent variables)

- **Artifactual** differences, such as in variable measurement (e.g. “pro-high-density attitude”)

- **Methodological** differences??

The research questions

- All else equal (ceteris paribus), *will different methodologies give different answers*?
  - I.e. markedly differing estimates of the % of the total BE impact on TB that is truly due to BE

- In our application,
  - Which method explains the *estimation* data best?
  - Which method predicts best on a *validation* (holdout) sample?
The approach

- Control for date, location, and variable definition, by *using the same data set* to…
- Compare the share of the total effect of the BE on TB that is “true BE”:
  \[
  \frac{\text{true effect of } BE}{\text{total effect of } BE \ (\text{incl. } BE \ & AT)}
  \]
  for 3 different approaches:
  - Statistical control (SC)
  - Propensity scores (PS)
  - Sample selection modeling (SS)

Empirical context

- Self-administered survey
- November 2003
- Movers and nonmovers randomly selected from 8 neighborhoods in Northern California (4 traditional, 4 suburb.)
- For this study, only commuting workers
  - \( N_{\text{calibration}} = 630 \)
  - \( N_{\text{validation}} = 274 \)

Handy et al. (2005)
**Empirical context** (cont’d)

Sac Traditional  Sac Suburban

**Empirical context – variables** (1)

- **Travel behavior (TB)** – *dependent variable*
  - Number of drive-alone commute trips per week

- **Neighborhood characteristics (BE)**
  - *Subjective perceptions* (factor analysis):
    - accessibility, physical activity options, safety, socializing, outdoor spaciousness, attractiveness
  - *Objective measures* (GIS): # of business types within specified distance from residence, distance to closest business for each type

TB = travel behavior  BE = built environment  GIS = geographic information systems
Empirical context – variables (2)

- **Residential preferences (AT)**
  - Parallel to perceived neighborhood characteristics

- **Travel attitudes (AT)**
  - (factor analysis): pro-walk/bike, pro-transit, travel liking, travel minimizing, safety of car, and car dependent

- **Socioeconomic (SE)**
  - Auto ownership, household structure, education, income, age, mobility limitation…

Methodological overview

- The typical model is
  \[ TB = f(BE, X) + \varepsilon \]

  - Standard techniques (regression, discrete choice) require that observed variables \((BE, X)\) be uncorrelated with unobserved ones \((\varepsilon)\)
  - Otherwise, the resulting *endogeneity bias* means that coefficients of \(BE\) and \(X\) will be *biased* and *inconsistent*
  - But if \(TB = f(BE(\text{AT}), X) + \varepsilon(\text{AT})\), then this requirement is violated
The 3 methods compared here

1. Statistical control

$$TB = f(BE, AT, X) + \xi$$

Removes AT from ε (unobserved) and makes it observed, reducing/eliminating the correlation of observed vars with unobserved ones.

2. Propensity scores

- **Regression**
  $$PS = \text{Pr}[RC \mid Y]$$  
  or  
  $$PS = \text{Pr}[RC \mid Y, AT]$$

- **Matching**
  Match cases with same PS but different RC to simulate random experiment
  $$TB = f(BE, X, PS) + \eta$$

- **Stratification**
  Divide into strata based on PS and compare mean $TB_U - TB_S$ for each stratum

RC = residential choice (U, urb. or S, suburb.)
2. Propensity scores

Matching
Match cases with same PS but different RC to simulate random experiment

\[ PS = \Pr [RC | Y] \]
\[ PS = \Pr [RC | Y, AT] \]

Stratification
Divide into strata based on PS and compare mean \( TB_U - TB_S \) for each stratum

\[ PS = \Pr [RC | Y] \]
\[ PS = \Pr [RC | Y, AT] \]
3. Sample selection*

\[ RC^* = \Phi(\beta'X) / \Phi(\beta'X) \]

where \( \Phi(\beta'X) \) is the observed if \( RC^* \geq 0 \) and \( \Phi(\beta'X) [/1-\Phi(\beta'X)] \) is observed if \( RC^* < 0 \)

The 3 methods compared here

* AKA “endogenous switching” or “mover-stayer model”

More details about SS

- Zhou and Kockelman (2008) classified 1,903 households in the 1998-1999 Austin Travel Survey into two groups: CBD and urban residents, and rural and suburban residents. They chose rural and suburban residents as a treatment group and the others as a control group. Using a sample selection model, they first modeled the prior residential choice (pseudo-R\(^2\) was 0.07) and then inserted a derived lambda (which is the inverse Mills ratio (IMR), i.e. \( \phi(\beta'X)/\Phi(\beta'X) \), for the treatment group and \( -\phi(\beta'X)/[1-\Phi(\beta'X)] \) for the control group) into the two equations for VMT of the treatment and control groups. They calculated and compared the average treatment effect (ATE: the average increase in VMT of moving a randomly-selected person from an urban neighborhood to a suburban one, or the true influence of the built environment) and the effect of treatment on the treated (TT: the average increase in VMT of having moved a randomly-selected suburban resident from an urban neighborhood to a suburban one, or the total influence of the built environment) (Heckman et al., 2001).
Goodness-of-fit measures

- \( R^2 = \frac{\text{regression sum–of–squares}}{\text{total sum–of–squares}} \) (prop. of var. expl.)
  \[
  \frac{\sum_i (\hat{y}_i - \bar{y})^2}{\sum_i (y_i - \bar{y})^2},
  \]
  where \( \hat{y}_i \) is the predicted value and \( y_i \) is the observed value of the dependent variable (TB) for case \( i \), and \( \bar{y} \) is the sample mean of TB

- \( \text{RMSD} = \sqrt{\frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{N}} \) (root mean squared deviation)

- \% correctly classified = \% of cases for which \( \hat{y}_i \) is within 0.5 (days per week) of \( y_i \)

The key substantive question

- Of the total apparent influence of the BE on TB, what proportion is due to self-selection, and what proportion due to the separate influence of the BE itself?
  
  \( \text{Tot BE infl} = \text{true BE infl} + AT \text{ (or RSS) infl} \)

- We’re interested in
  
  \[
  \frac{\text{true BE infl}}{\text{true BE infl} + AT \text{ infl}} \quad \text{or} \quad \frac{AT \text{ infl}}{\text{true BE infl} + AT \text{ infl}}
  \]
How do we operationalize this?

Two logical metrics:

1. **Incremental % variance (in TB) explained**
   - R²-based
   - There’s a long history in regression of using “decomposition of variance” to assess the contributions of specific (blocks of) variables

2. **Marginal contributions to TB itself**
   - \[
     \frac{\text{ave. treatment effect (ATE)}}{\text{effect of treatment on the treated (TT)}} = \frac{\text{true effect (of BE)}}{\text{biased effect (incl. BE & AT)}}
   \]
   - Conventional metric in the “program evaluation” lit
   - Natural to focus directly on the “effect size” of the outcome of interest (TB)

---

Metrics for assessing (true BE contrib / total BE contrib)

1. **Incremental % variance explained**
   - **Hierarchical assumption:**
   - **Statistical control (SC):**
     \[
     \frac{\text{incremental contrib of BE}}{\text{inclem. contrib of BE + AT}} = R^2_{\text{full}}
     \]
   - **Propensity score (sample selection):** same as SC, except
     - Model yielding \( R^2_{SE} \) has BE & AT removed from both PS (selection) model and outcome model(s)
     - Model yielding \( R^2_{SE+AT} \) has BE removed from both PS (selection) model and outcome model(s)
1. Incremental % variance explained

- Hierarchical assumption:

- Statistical control (SC):
  \[ \text{incremental contrib of } BE = \frac{R_{full}^2 - R_{SE+AT}^2}{R_{full}^2 - R_{SE}^2} \]

- Propensity score (sample selection): same as SC, except
  - Model yielding \(R_{SE}^2\) has BE & AT removed from both PS (selection) model and outcome model(s)
  - Model yielding \(R_{SE+AT}^2\) has BE removed from both PS (selection) model and outcome model(s)

R² for the SS model

- \(R^2 = \frac{\sum_i (\hat{y}_i - \bar{y})^2}{\sum_i (y_i - \bar{y})^2} = [\text{corr}(\hat{y}_i, y_i)]^2\), where

\[
\hat{y}_i = \begin{cases} 
TB_{Ui} & \text{if } RC_i = 1 \text{ (urban location)} \\
TB_{Si} & \text{if } RC_i = 0 \text{ (suburban location)} 
\end{cases}
\]
2. Contributions to TB directly

- **Statistical control (SC) – option 1:**
  \[
  \beta_{BE} \text{ for model with } AT, SE, & BE \\
  \beta_{BE} \text{ for model with only } SE & BE
  \]
  Each \( \beta \) measures the impact on TB of a one-unit change in BE:
  - “total effect of BE” in the denominator, and
  - “true effect of BE” in the numerator
  But what if there is more than one BE variable?

- **Statistical control (SC) – option 2:**
  \[
  \beta_{RC} \text{ for model with } AT, SE & RC \\
  \beta_{RC} \text{ for model with only } SE & RC
  \]
  Each \( \beta \) measures the impact on TB of changing RC from 0 to 1

- **Propensity score (PS, matching or stratification):**
  \[
  \frac{\text{ave. } (TB_U - TB_S \text{ after matching})}{\text{ave. } TB_U - \text{ave. } TB_S \text{ without matching}}
  \]

- **Sample selection (SS):**
  - Complicated formula! for Ave. Treatment Effect / Treatment effect on the Treated (ATE/TT)

**Metrics for assessing (true BE contrib / total BE contrib)**

Heckman et al. (2001)
2. Contributions to TB directly

**Sample selection (SS):**
- Treatment effect (TE):
  \[ TE = \beta_1 x_1 + \rho_1 \sigma_x \left( \frac{\phi(\alpha' w_i)}{\Phi(\alpha' w_i)} \right) - \beta_0 x_0 - \rho_0 \sigma_0 \left( -\frac{\phi(\alpha' w_i)}{\Phi(-\alpha' w_i)} \right) \]
- This assumes you have data on both states for an individual!
- Instead, look at an “average” treatment effect (ATE):
  \[ ATE = \beta_1 x - \beta_0 x = (\beta_1 - \beta_0) x \]
- “Treatment effect on the treated” (TT):
  \[ TT = E[Y_1 | z = 1] - E[Y_0 | z = 1] \]
  \[ = (\beta_1 - \beta_0) x + \left[ (\rho_1 \sigma_x) - (\rho_0 \sigma_0) \right] \left( \frac{\phi(\alpha' w_i)}{\Phi(\alpha' w_i)} \right) \]

**The approach**

For each method:
- Evaluate the *goodness of fit* (R², RMSE, %CC) and the *substantive answer* (“share truly due to BE”)
- for both a *calibration* sample and a holdout *validation* sample
- We should prefer the *answer given by the method that best fits the validation sample*
## Results

### Calibration (N ~ 630)

<table>
<thead>
<tr>
<th></th>
<th>Statist. control</th>
<th>Propensity score</th>
<th>Sample selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$</td>
<td>0.10</td>
<td>0.10</td>
<td>0.14</td>
</tr>
<tr>
<td>RMSD/sd(y)</td>
<td>0.95</td>
<td>0.95</td>
<td>0.94</td>
</tr>
<tr>
<td>% corr. class.*</td>
<td>60.8</td>
<td>59.7</td>
<td>64.4</td>
</tr>
<tr>
<td>% BE impact ($R^2$)</td>
<td>11.8</td>
<td>8.0 (regression)</td>
<td>9.1</td>
</tr>
<tr>
<td>% BE imp (ATE/TT)</td>
<td>61.1</td>
<td>58.7 (stratific.)</td>
<td>72.7</td>
</tr>
</tbody>
</table>

### Validation (applying calibration model parameters) (N ~ 274)

<table>
<thead>
<tr>
<th></th>
<th>Statist. control</th>
<th>Propensity score</th>
<th>Sample selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$</td>
<td>0.08</td>
<td>0.07</td>
<td>0.10</td>
</tr>
<tr>
<td>RMSD/sd(y)</td>
<td>0.96</td>
<td>0.97</td>
<td>0.97</td>
</tr>
<tr>
<td>% corr. class.*</td>
<td>63.1</td>
<td>51.3</td>
<td>55.5</td>
</tr>
<tr>
<td>% BE impact ($R^2$)</td>
<td>-13.7**</td>
<td>-7.0**</td>
<td>-11.8**</td>
</tr>
<tr>
<td>% BE imp (ATE/TT)</td>
<td>61.1</td>
<td>59.5 (stratific.)</td>
<td>75.0</td>
</tr>
</tbody>
</table>

---

*RMSD = root mean squared deviation between y and $\hat{y}$*

*RMSD = root mean squared deviation between y and $\hat{y}$*

* 5-7 days combined into single category  ** Model better without BE variables than with them
Summary & discussion (1)

- (Attitudinal) self-selection is an issue in nearly every choice we study
  - E.g., mode choice, vehicle type, telecommuting
- It’s dangerous to project future effects (especially of a policy) from those of (natural) early adopters
  - Later adopters may have different circumstances (including attitudes), and may adopt less voluntarily, or for different reasons
- We should be more aware of this issue, and of ways to deal with it

van Wee (2009)

Summary & discussion (2)

- However, the three methods we compared using the same dataset…
  - 
  - Statistical control (SC)
  - Propensity score (PS)
  - Sample selection (SS)
- … had similar fits on the calibration sample
  - SS had a slight edge, but
- a different method (SC) was (markedly) better on % correctly classified for the validation sample (63% SC, 56% SS)
Further, the two methods for assessing the “true BE” share of “total BE” gave radically different substantive answers:

- 8-12% (cal.) and 0% (val.) for the $R^2$-based answers
- 59-73% (cal.) and 60-75% (val.) for the effect-size (ATE/TT)-based answers

- Is ATE/TT attributing too much to “total BE”? Does everything get thrown in there – measurement error, reporting error, idiosyncratic factors – as well as AT?
- Is $R^2$ attributing too little to BE? Depends on how well BE is observed

Even the method with the most desirable answer (SS) indicates that RSS discounts the total BE impact by ~25%:

- Coincidentally (?), that is the approximate unweighted average of the RSS impacts for the ~15 studies shown previously

But if we prefer the best-fitting model on the validation sample (SC), the discount deepens to 32%
WHAT IS THE EFFECT SIZE? I.e. the ATE itself?
Mokhtarian, Patricia Lyon; 21-6-2014
Summary & discussion (5)

- In estimation, SS has an “unfair advantage”, in that all coefficients in the TB model are allowed to differ by RC
  - Can investigate doing the same for the SC & PS methods, and compare to SS
- In validation, maybe the simplicity of the SC method makes it more robust/transferable?
- At least in this sample, PS never seems to be best…

Summary & discussion (6)

- Further research is needed to
  - Compare the results with and without attitudes
  - Compare the results when the set of final explanatory variables is held constant across method
  - Compare additional methods
  - See if the patterns observed here are consistent across empirical contexts
  - Analyze the reasons for the difference in results
- We look forward to seeing additional studies along these lines! (But please let us publish this one first… 😊)
Selected references (1)


Selected references (2)

Selected references (3)


Questions?

patmokh@gatech.edu

Slide borrowed from David Ory
Endogeneity bias

- Can take the form of omitted variables bias (OVB):
  
  \[ TB = f_1(BE(AT), X) + \epsilon(AT) \]
  
or
  \[ TB = f_1(BE, X) + \epsilon(AT(BE)) \]

Potential forms of omitted variables bias (OVB)

- OVB: AT antecedent
  
  \[ TB = f_1(BE(AT), X) + \epsilon(AT) \]
  
  AT = attitudes
  BE = built environment
  TB = travel behavior

- OVB: AT intervening
  
  \[ TB = f_1(BE, X) + \epsilon(AT(BE)) \]
Endogeneity bias

- Can take the form of omitted variables bias (OVB):
  \[ TB = f_1(BE(AT), X) + \varepsilon(AT) \]
  or \[ TB = f_1(BE, X) + \varepsilon(AT(BE)) \]

- Or simultaneity bias (SB):
  \[ TB = f_1(BE, X,Y) + \varepsilon_1 \]
  \[ BE = f_2(TB, X,Z) + \varepsilon_2 \]

Potential forms of endogeneity bias
**Relationship between self-selection and misestimation**

Low Accessibility | High Accessibility
---|---
Random $\mu_1$ | $\mu_2$
All Matched $\mu_1'$ | $\mu_2'$
All Mismatched $\mu_1''$ | $\mu_2''$

$\mu_1$, $\mu_1'$, and $\mu_1''$ are the observed means of a walking behavior measure for people living in low-accessibility neighborhoods;

$\mu_2$, $\mu_2'$, and $\mu_2''$ are the observed means of a walking behavior measure for people living in high-accessibility neighborhoods (the "treatment");

$\text{ATE} = \mu_2 - \mu_1$.

$\text{Diff}_1 = \mu_2' - \mu_1'$.

$\text{Diff}_2 = \mu_2'' - \mu_1''$.

Adapted from Cao (2010)