Ask a Hypothetical Question, Get a Valuable Answer?

Joseph Herriges
Catherine Kling
Christopher Azevedo
Prairie Pothole Region

Prairie Pothole Region of North America includes the states of Montana, North Dakota, South Dakota, Minnesota, Iowa, and the provinces of Alberta, Saskatchewan, and Manitoba.
Iowa River Corridor Project

Tama County
Area A
Area B
Area C
Area D

Poweshiek County
Benton County
Iowa County

Legend

- Project
- Road
- County Line
- Town
- River

Vicinity
IOWA

Toledo
Tama
Belle Plaine
Marengo
Chelsea/Plume
Amana

Iowa River
Sources of Data

1. Revealed Preference (RP): Behavior
e.g., hedonic price data, travel cost data

2. Stated Preference (SP): Statements
e.g., contingent valuation, contingent behavior
Previous Work

- Pooling (RP, SP data of same form)
  - Adamowicz, Louviere, and Williams (1994)
  - Layman, Boyce, and Criddle (1996)
  - Englin and Cameron (1996)

- Combining (RP, SP data of different form)
  - Cameron (1992)
  - Larson (1990)
  - Loomis (1997)
  - Huang, Haab, and Whitehead (1997)
  - McConnell, Weninger, and Strand (1999)
  - Cameron, Poe, Ethier, Schulze (1999)
Previous Work (continued)

- Pooling or Combining (2 pieces of SP data)
  - Niklitschek and Leon (1996)
  - Huang, Haab, and Whitehead (1997)
Reasons Cited for Combining or Pooling Data

• Increase Precision of Estimates

• Test Consistency Across RP and SP Data

• Impose discipline of market on SP data, while allowing SP data to “fill in” some information about preferences not captured by RP data (Cameron)
Alternative Interpretations of Consistency Tests (rejection)

1. RP Lovers: View these as validity tests of SP against RP (MWS)

2. SP Lovers: View these as validity tests of RP(?)
   Basis: Randall, mis-measured RP data then biased price coefficient

3. Agnostics: Jointly estimate and constrain parameters to be alike to take advantage of strengths of both
Our Interests

• Value Wetland Use in Iowa using RP and SP Data Jointly

• Test for consistency of RP and SP generally

• Test specific hypotheses concerning sources of bias
Our Interests (more)

• Investigate these issues with two different forms of SP data
  – Dichotomous data: “yes/no” answers
  – Continuous data: “how many?” answers

• Reconsider Interpretation of bias and consistency tests
Model of RP Data

- Standard Demand Function

\[ x^{RP} = f(p^{RP}, y^{RP}; \beta^{RP}) + \sigma^{RP} \varepsilon^{RP}, \]

- Tobit: Correction for Censoring

\[ f(x \mid x > 0) = h(\varepsilon) = \frac{(1/\sigma)\phi[(x - f(p,y))/\sigma]}{\Phi[f(p,y)/\sigma]}, \]
Model of SP Continuous Data

• Standard Demand Model Again

\[ x^{SP} = f(p^{SP}, y^{SP}; \beta^{SP}) + \sigma^{SP} \varepsilon^{SP}, \]

• Tobit: Correction for Censoring

• Consumer Surplus at Current Use

\[ \hat{c} = g(p, y; \hat{\beta}, \hat{\sigma}). \]
Model of SP Discrete Data

• Only “yes” or “no” response from SP data

• Model Probability of “yes” as:

\[
\Pr(\text{wtp} = \text{"yes"}) = \Pr\{f(p^{SP}, y^{SP}; \beta^{SP}) + \sigma^{SP} \varepsilon^{SP} > 0\}
\]

\[
= \Pr\{\varepsilon^{SP} > -f(p^{SP}, y^{SP}; \beta^{SP}) / \sigma^{SP}\}
\]

\[
= \Phi[f(p^{SP}, y^{SP}; \beta^{SP}) / \sigma^{SP}].
\]

• \( \Pr(\text{wtp} = \text{"no"}) = 1 - \Pr(\text{yes}) \)
Joint Estimation of SP and RP Data

- Simple sum of RP and SP likelihoods if independent errors, but data from same individuals so correlation likely

- Why not identical errors from same individuals?

- Sources of RP error
  - Recall concerning # of visits
  - Errors in optimization
  - Random Preferences
  - Omitted variables

- Sources of SP error (previous three plus)
  - Inaccurate comprehension of survey wtp question
  - Phone vs. mail vs. in person survey
  - Inaccurate comprehension of other survey details (payment vehicle, time table for payment, etc.)
Joint Estimation of SP and RP Data (Correlated)
More stuff to do:

1. Modeling --- flexible forms, extend model to multiple sites

2. Bayesian view of combining data: weight different sources of data differently depending on ones priors?

3. Kerry’s ideas
More stuff to do:

1. Flexible forms,

2. Extend model to multiple sites,

3. Bayesian view of combining data: weight different sources of data differently depending on ones priors?
Consistency Tests

1. Parameter values are identical across RP and SP data: test equality of all coefficients

2. Parameter values are identical but errors have different variances (heteroskedasticity): test equality of all coefficients except variances
### RP and SP\textsuperscript{C} Joint Models

<table>
<thead>
<tr>
<th></th>
<th>Independent</th>
<th>Correlated</th>
<th>Fully Consistent</th>
<th>Heteroskedasticity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant\textsuperscript{RP}</td>
<td>20.65 (8.21)</td>
<td>19.12 (7.98)</td>
<td>14.83 (7.92)</td>
<td>16.11 (7.92)</td>
</tr>
<tr>
<td>$k\textsuperscript{constant}$ \textsuperscript{SP}</td>
<td>0.50 (-2.54)**</td>
<td>0.61 (-2.45)**</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Price\textsuperscript{RP}</td>
<td>-0.82 (-8.82)</td>
<td>-0.76 (-9.11)</td>
<td>-0.55 (-14.87)</td>
<td>-0.61 (-11.22)</td>
</tr>
<tr>
<td>$k\textsuperscript{price}$ \textsuperscript{SP}</td>
<td>0.58 (-3.54)**</td>
<td>0.60 (-3.97)**</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Income\textsuperscript{RP}</td>
<td>0.14 (3.84)</td>
<td>0.13 (3.65)</td>
<td>0.11 (3.24)</td>
<td>0.12 (3.41)</td>
</tr>
<tr>
<td>$k\textsuperscript{income}$ \textsuperscript{SP}</td>
<td>0.75 (-0.61)**</td>
<td>0.64 (-1.20)**</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>$\sigma\textsuperscript{RP}$</td>
<td>13.28 (18.42)</td>
<td>13.43 (18.13)</td>
<td>13.64 (19.77)</td>
<td>13.19 (18.56)</td>
</tr>
<tr>
<td>$k\textsuperscript{\sigma}$ \textsuperscript{SP}</td>
<td>1.12 (1.02)**</td>
<td>1.01 (0.11)**</td>
<td>1.00</td>
<td>1.15 (1.53)**</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.63 (13.27)</td>
<td>0.64 (13.48)</td>
<td>0.64 (13.41)</td>
<td></td>
</tr>
<tr>
<td>$-\log L$</td>
<td>1180.80</td>
<td>1136.92</td>
<td>1142.91</td>
<td>1141.55</td>
</tr>
<tr>
<td>CS\textsuperscript{RP}</td>
<td>114.14</td>
<td>122.35</td>
<td>169.45</td>
<td>153.27</td>
</tr>
<tr>
<td>CS\textsuperscript{SP}</td>
<td>197.93</td>
<td>203.92</td>
<td></td>
<td></td>
</tr>
<tr>
<td>p-values (Likelihood ratio tests)</td>
<td>0.02 (reject @5%)</td>
<td>0.03 (reject)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The t-statistics are in parentheses below the coefficient estimates. Double asterisks indicate tests for difference from 1.00.
The t-statistics are in parentheses below the coefficient estimates. Double asterisks indicate tests for difference from 1.00.
Tests of Bias Stories

- Hypothesis: When respondents answer SP questions, they ignore their budget constraint: Test equivalence of all parameters except income coefficient (and variance)
  
  Premise: RP is accurate, test SP against it (RP Lovers)

- Hypothesis: Price term in RP data is mis-measured: Test equivalence of all parameters except price coefficient (and variance)
  
  Premise: SP is accurate, test RP against it (SP Lovers)

- Hypothesis: Both of the above
  - Premise: Both potentially inaccurate (Agnostics)
The t-statistics are in parentheses below the coefficient estimates. Double asterisks indicate tests for difference from 1.00.
Iowa Wetlands Survey
Wetland Usage Information
Gathered by Zones
Respondents were asked:

- The number of trips made to each zone (traditional RP data)
- Would they still take any trips if cost of access were higher? (SP data, discrete)
- How their number of trips would change with increased costs (SP data, continuous)
Alternative Interpretation of Results

1. RP Lovers

• Results prove SP biased (RP/SPc reject consistency)

• RP/SPd fails to reject, but that is due to low information content; shows how insidious SP data can be! Can trick SP lovers into feeling confident when shouldn’t.

• Conclusion: Use RP data to estimate welfare measures, could jointly estimate to get efficiency gains, but probably not worth the trouble
Alternative Interpretation of Results

2. SP Lovers

• Results prove RP biased (RP/SPc reject consistency)

• RP/SPd fails to reject, but that is due to low information content

• Conclusion: Use SP data to estimate coefficients and compute welfare, could jointly estimate to get efficiency gains, but probably not worth the trouble
Alternative Interpretation of Results

2a. SP Lovers (less faithful sect)

- RP/SPd doesn’t reject consistency because easier to answer SPd, results indicate SPd may be more accurate

- Conclude: Combine RP/SPd to do welfare estimates

- More research is needed to understand how we should write SP questions for most reliable information.
Alternative Interpretation of Results

3. Agnostics

- These results support the idea that there are problems with both kinds of data: SP and RP, the fact that they are inconsistent indicates that there is bias in one or both

- Conclude: Combine RP/SPc to do welfare estimates, use all information available and throw it in the likelihood function, hope that whatever bias is present in one method is countered by the other and get some efficiency gains.

- More research is needed on all of these issues!
Table II: Summary Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>8.23</td>
<td>10.91</td>
</tr>
<tr>
<td>Percent of X positive</td>
<td>66.55%</td>
<td></td>
</tr>
<tr>
<td>X&lt;sup&gt;new&lt;/sup&gt;</td>
<td>2.68</td>
<td>6.27</td>
</tr>
<tr>
<td>Percent of X&lt;sup&gt;new&lt;/sup&gt; Positive</td>
<td>27.34%</td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>27.79</td>
<td>12.30</td>
</tr>
<tr>
<td>P&lt;sup&gt;new&lt;/sup&gt;</td>
<td>54.24</td>
<td>18.17</td>
</tr>
<tr>
<td>Y</td>
<td>40826</td>
<td>25.32</td>
</tr>
</tbody>
</table>
Table I: Bid distribution for the Stated Preference Data

<table>
<thead>
<tr>
<th>Bid</th>
<th>Number of Surveys</th>
</tr>
</thead>
<tbody>
<tr>
<td>$5</td>
<td>1200</td>
</tr>
<tr>
<td>$10</td>
<td>1200</td>
</tr>
<tr>
<td>$15</td>
<td>1200</td>
</tr>
<tr>
<td>$20</td>
<td>600</td>
</tr>
<tr>
<td>$30</td>
<td>600</td>
</tr>
<tr>
<td>$40</td>
<td>600</td>
</tr>
<tr>
<td>$50</td>
<td>600</td>
</tr>
</tbody>
</table>
Table III: Coefficient and Consumer Surplus Estimates: Independent Models

<table>
<thead>
<tr>
<th></th>
<th>RP</th>
<th>SP(^D)</th>
<th>SP(^C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>20.65</td>
<td>15.87</td>
<td>10.39</td>
</tr>
<tr>
<td></td>
<td>(8.21)</td>
<td>(3.56)</td>
<td>(2.70)</td>
</tr>
<tr>
<td>Price</td>
<td>-0.82</td>
<td>-0.59</td>
<td>-0.47</td>
</tr>
<tr>
<td></td>
<td>(-8.82)</td>
<td>(-6.73)</td>
<td>(-5.72)</td>
</tr>
<tr>
<td>Income</td>
<td>0.14</td>
<td>0.13</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>(3.84)</td>
<td>(2.31)</td>
<td>(.08)</td>
</tr>
<tr>
<td>(\sigma)</td>
<td>13.28</td>
<td>14.92</td>
<td>14.92</td>
</tr>
<tr>
<td></td>
<td>(18.42)</td>
<td>(not estimated)</td>
<td>(10.80)</td>
</tr>
<tr>
<td>-log(L)</td>
<td>791.35</td>
<td>134.90</td>
<td>389.45</td>
</tr>
<tr>
<td>CS</td>
<td>113.94</td>
<td></td>
<td>197.93</td>
</tr>
</tbody>
</table>

The t-statistics are in parentheses below the coefficient estimates.