Big Data in Housing: An Overview of Zillow Microdata and its Potential for National Accounts

Marina Gindelsky & Scott Wentland
Office of the Chief Economist

BEA Advisory Committee
May 18, 2018

Disclaimer: The results and opinions are those of the authors and do not reflect the position of Zillow Group.

Acknowledgements: We would like to thank members of our “project share” across BEA directorates (Wade Petty, Ethan Schein, Mitchell Shabani, and Sarah Osborne) who have helped us clean, validate, and improve the functionality of the Zillow data.
• **Introduce a Big Data set** we have recently acquired at BEA from **Zillow**
• **Compare it to Census’s** more traditional survey-based national housing data (ACS)
• **Discuss potential applications** and its path forward as a resource to enhance the accounts
  – Potential to improve or supplement Census survey data as well
• Zillow has agreed to give BEA its “ZTRAX Data Set”
  – Large microdata that underlies its website
    • NOT its “Zestimates” (yet), but the data underlying these
    • NOT its MLS or scraped data (yet)
  – Comes from multiple sources (e.g., public, administrative tax records)
    • Why do we need Zillow, then?
    • 3000+ counties and municipalities report the data differently and assembling/organizing this data into a single national data set is a Herculean effort
• **Assessment data**
  – From local *tax assessors’* public data
    • 200 million parcels in over 3,100 counties (Zillow)
  – **Characteristics of individual properties** (bedrooms, bathrooms, sq. ft., year built, etc.)

• **Transactions data**
  – 374 million detailed public records across over 2,750 counties (Zillow)
  – **Characteristics of transactions** (sale price, dates, mortgage information, foreclosure status, etc.)
  – **Observed prices**, not imputed or surveyed
Zillow’s Website – Example

Key
Orange: included in both ACS & Zillow; Blue: only in Zillow
Key

Orange: included in both ACS & Zillow; Blue: only in Zillow
Zillow’s Website – Example

Key
Orange: included in both ACS & Zillow; Blue: only in Zillow; Purple: inferred from Zillow
Zillow’s Website – Example

Key
Orange: included in both ACS & Zillow; Blue: only in Zillow
Previously on Zillow.....

For Washington State

<table>
<thead>
<tr>
<th>ACS</th>
<th>Zillow</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015</td>
<td>2013-2016</td>
</tr>
<tr>
<td>1% Sample = 31,000 households (assumed 3.1 million in full survey)</td>
<td>3.25 million parcels</td>
</tr>
</tbody>
</table>
Zillow-ACS Coverage Nationwide

<table>
<thead>
<tr>
<th>Public-use ACS</th>
<th>Zillow</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015</td>
<td>2013-2016</td>
</tr>
<tr>
<td>1% Sample = 1,375,481 households</td>
<td>~150 million parcels</td>
</tr>
<tr>
<td>(assumed ~150 mill in full survey)</td>
<td></td>
</tr>
<tr>
<td>478 unique counties (includes</td>
<td>2673 unique counties</td>
</tr>
<tr>
<td>pooled ones)</td>
<td></td>
</tr>
<tr>
<td>Medians</td>
<td>ACS</td>
</tr>
<tr>
<td>-----------------</td>
<td>------</td>
</tr>
<tr>
<td># Bedrooms</td>
<td>3</td>
</tr>
<tr>
<td>Year Built</td>
<td>1975</td>
</tr>
<tr>
<td># Rooms</td>
<td>6</td>
</tr>
<tr>
<td>Tax Amount</td>
<td>2150</td>
</tr>
<tr>
<td>Acres &gt;10</td>
<td>0</td>
</tr>
</tbody>
</table>

Omitted States: KS, MT, UT, VT (no 2015 assessment records)
## Ratio of Zillow to ACS by MSA (2015)

<table>
<thead>
<tr>
<th>City</th>
<th># Bedrooms</th>
<th>Year Built</th>
<th>Tax Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atlanta</td>
<td>1.00</td>
<td>1.00</td>
<td>0.90</td>
</tr>
<tr>
<td>Boston</td>
<td>1.00</td>
<td>1.00</td>
<td>0.95</td>
</tr>
<tr>
<td>Chicago</td>
<td>1.00</td>
<td>1.01</td>
<td>0.84</td>
</tr>
<tr>
<td>Detroit</td>
<td>1.00</td>
<td>1.01</td>
<td>0.40</td>
</tr>
<tr>
<td>New York</td>
<td>1.00</td>
<td>1.00</td>
<td>1.05</td>
</tr>
<tr>
<td>Washington, D.C.</td>
<td>1.00</td>
<td>1.00</td>
<td>1.02</td>
</tr>
<tr>
<td>National</td>
<td>1.00</td>
<td>1.00</td>
<td>0.87</td>
</tr>
</tbody>
</table>

Omitted States: KS, MT, UT, VT (no 2015 assessment records)
## Zillow Stats (ASMT) for 2015

<table>
<thead>
<tr>
<th></th>
<th>Bathrooms</th>
<th>Basement</th>
<th>Garage (Cars)</th>
<th>Stories</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean</strong></td>
<td>1.45</td>
<td>0.193</td>
<td>1.91</td>
<td>1.45</td>
</tr>
<tr>
<td><strong>Median</strong></td>
<td>1.5</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td><strong>Std. Dev</strong></td>
<td>1.17</td>
<td>0.395</td>
<td>8.197</td>
<td>1.45</td>
</tr>
<tr>
<td><strong>Min.</strong></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Max</strong></td>
<td>21</td>
<td>1</td>
<td>5204</td>
<td>105</td>
</tr>
</tbody>
</table>

Omitted States: KS, MT, UT, VT (no 2015 assessment records)
*Note: National numbers do not include non-disclosure states: ID, IN, KS, MS, MO, MT, NM, ND, SD, TX, UT, WY
ACS Home Value vs. Zillow Home Sale 2015

*Note: National numbers do not include non-disclosure states: ID, IN, KS, MS, MO, MT, NM, ND, SD, TX, UT, WY
Applications to the Accounts

• GDP – PCE & “space rent”
  – Improve rental equivalence & user-cost methods
    • Aten has previously discussed this in more detail
    • Complementary approaches – using different data
Applications to the Accounts

• An important issue for space rent: the distribution of owner-occupied homes is different from rentals
  – Zillow potential: value data for renters by merging this data with Census → linked data
  • Non-parametric approaches: use recent transactions of rental properties to approximate value (with some adjustment based on how recent the transaction was)
  • Parametric/regression approaches: use detailed transaction and property characteristic data to estimate values hedonically
User-Cost Approach?

\[ U_{it} = P_{it} (i_{it} + \gamma_{it} + \tau_{it} - E[\pi_{it}]) \]

• Where
  – \( P \) is the price of an individual home
  – \( i \) is the nominal interest rate for a given home
  – \( \gamma \) is the maintenance and repair for an individual home (note: this may be a separate line item)
  – \( \tau \) is the property taxes and insurance paid
  – \( E[\pi] \) is expected appreciation (revaluation) for a given home over the year constructed via Zillow data
User-Cost Approach – Micro Data

• Estimate the **user costs using individual-level microdata**
  – Leveraging extensive **price** (transactions) data from Zillow
  – **Property taxes & mortgage** info. from Zillow
    • **Validate and expand using Census survey data**
  – Estimate expected appreciation from the microdata
    • Test these against other automated valuation models (AVMs) and price indices
  – More on this in Gindelsky, Moulton, Wentland @ NBER-CRIW 2018 & 2019
Other Applications

• Residential **fixed assets**
• Land value
  – Integrated macroeconomic accounts
  – Nonproduced, nonfinancial fixed asset
    • Land is currently not incorporated into the balance sheet for fixed assets and inventories, but could be.
Additional questions/comments?

Marina Gindelsky
Marina.Gindelsky@bea.gov

Scott Wentland
Scott.Wentland@bea.gov