

TAXABLE PROPERTY VALUES: EXPLORING THE FEASIBILITY OF DATA COLLECTION METHODS

Brian Zamperini, Jennifer Charles, and Peter Schilling
U.S. Census Bureau*

INTRODUCTION

PROPERTY TAX IS A PRINCIPAL REVENUE SOURCE of local governments within the United States. According to the Census Bureau's 2008 State and Local Government Finance data, at the local government level, property taxes made up the largest source of revenue, amounting to \$397 billion or 72.3 percent of total local tax collection¹. Understanding the property tax base is a key component to understanding local government financing roles and intergovernmental relationships.

Currently no comprehensive data source exists that summarizes taxable property in the U.S., nor does a source exist to compare the property tax base and assessment quality of property values across states. The Census Bureau originally compiled data on taxable property values in 1850. From 1957 through 1992, these data were compiled every five years as the Taxable Property Values (TPV) survey, as part of the Census of Governments. The TPV survey consisted of two parts: *Assessed Valuations for Local General Property Taxation and Taxable Real Property Assessment/Sales Price Ratios*. Over the years, the survey encountered a number of cost, resource, and statistical quality issues, and, in 1982, the last comprehensive TPV survey was conducted. Due to resource issues, in 1987, only assessed valuations data for local general property taxation and use class information were published, eliminating the ratio study. By 1992, only statistics on the official values of properties were reported. Taxable property products were completely discontinued thereafter.

In 2007, the National Research Council Committee on National Statistics evaluated the programs of the Governments Division at the U.S. Census Bureau. The panel recommended that the Governments Division explore the possibility of collecting data on taxable property values.

Based on this recommendation and a general interest from data users through outreach operations and literature reviews on the subject, the Governments Division conducted a study to determine the feasibility of collecting and producing national statistics on taxable property values.² The feasibility study explored data collection options, which included data from jurisdiction websites, data provided by a third party vendor, and direct collection from local jurisdictions. Data collection options were evaluated by availability, resource requirements, data quality, and compliance with Census Bureau's statistical standards.

WEBSITE

The feasibility study evaluated the possibility of collecting TPV data directly from government jurisdiction websites. If available, using sales and assessment data already provided on government websites would decrease much of the respondent burden, survey costs, and other constraints often associated with the historical TPV survey.

A representative sample (101 assessment jurisdictions and 101 sales jurisdictions) was selected to determine whether website data were accessible and usable. Analysts attempted to identify a general website for each sampled jurisdiction. Appropriate jurisdiction websites were found for a total of 99 of the 101 units in the assessment sample, and 96 of the 101 units in the sales sample.

Jurisdiction websites were then examined for online property data for the 2009 calendar year. Jurisdiction websites that provided comprehensive sales lists or tax assessment rolls for all parcels in a jurisdiction were considered sufficient for a TPV survey. It was expected that where data were available it would be in one of two forms: a searchable database or a database with a full assessment roll or a full sales listing. In cases where data were found, most websites had searchable databases that required some form of parcel information, such as parcel ID or address, to be entered as search criteria. Of the 73 assessment jurisdictions with

Disclaimer: This report is released to inform interested parties of research and to encourage discussion of work in progress. The views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.

online data, 59 had searchable databases. Out of the 41 sales jurisdictions with online data, 34 had searchable databases; 11 of the 34 searchable sales databases allowed for a specified date range to be used for finding 2009 sales data. While searchable databases can be useful to search a handful of records, it is not an option for large-scale, comprehensive queries. Details about the online data can be found in table 1.

To be suitable for a TPV survey, data available on a jurisdiction website must have both sufficient content and a usable format. A website offered sufficient content for the TPV study if it included an assessment or sales date, a matching mechanism (such as assessor parcel number or secondary matching fields such as name and street address), designation of use class, and the sale or assessment value. Usable format websites could provide data in a Microsoft Excel or text file format. Only five of the original 101 assessment jurisdiction websites provided sufficient assessment data in a usable format. Similarly, only five of the original 101 sales jurisdictions in the sample had sufficient sales data in a usable format. For both types of websites, the content most often missing was the use class designation.

Jurisdiction websites must also be able to provide historical data, as well as current year data, for a TPV survey. Historical data are required because analysts anticipate publishing TPV data products one year behind current year (real time) data. Though five of the 101 assessment jurisdiction websites provided a comprehensive assessment list with sufficient content in a usable format, none of the five could provide sufficient historical data. Two of the 101 sales jurisdiction websites provided complete data (both current year and historical) in a usable format. Thus, of the original sample of

jurisdictions, none of the assessment jurisdiction websites and only two of the sales jurisdiction websites could meet all the requirements for a TPV survey.

The lack of jurisdiction websites that can provide sufficient data (both current and historical) in a usable format makes data collection via websites an unsuitable option for a TPV survey. At best, online data resources may be used to verify data obtained through some other means.

THIRD PARTY VENDOR DATA EVALUATION

A third party vendor’s ability to collect, package, and resell information and services could provide a more efficient means of data collection. Data vendors were evaluated for detailed data availability and geographic coverage. Required data variables included the following parcel level information: assessment value, market value, sales price (if the parcel was sold within the year under examination), parcel land use code, and matching parcel identifiers (such as the Assessor’s Parcel Number, or APN). For geographic coverage, only vendor data that included the majority of the assessing and recording jurisdictions in the U.S. were considered.

After conducting comprehensive market research and issuing an request for proposal (RFP), the Census Bureau made a one-time purchase of 2009 data from the selected vendor. The purchase consisted of two sets of data: one national database with assessment and assessment-related information, and a second national database for sales and sales-related information. The selected vendor provided data for over 140 million parcels of property within the U.S., organized by county. The vendor data files were populated using data provided by

Table 1
Website Study Results

	<i>Assessments</i>	<i>Sales</i>
Data Available Online	73	41
Data by Parcel Search	59*	34*
Full Database	14*	18*
Full Database / Sufficient Detail	9	6*
Data in Usable Format	5	5
Full Database/Sufficient Detail in a Usable Format (with Current and Historical Data)	0	2

* includes Washington, DC

jurisdiction offices. Jurisdiction offices that did not provide data to the vendor were not covered in vendor files.

The vendor provided assessment data for 3,049 counties (99.2 percent of what was expected) and sales data for 2,540 counties (98.5 percent of what was expected). In some cases where 2009 assessment data were unavailable other historical data were provided. In these cases, 2008 data were preferable, but the Census Bureau also received some data for 2006, 2007, and 2010. According to Census Bureau statistical standards, the prior year data were not a suitable substitute unless assessments have not been reported more recently. Assessment information was provided in 100 percent of counties in 44 states, including the District of Columbia. The vendor attributed lower assessment data coverage in the remaining six states to the geographic nature and low population of the areas. The lack of coverage in these states would require mail surveys or field visits for data collection to these areas.

For sales data, coverage of counties was 100 percent in 23 states, including the District of Columbia. Low coverage of sales data in some states is due to non-disclosure laws. Twelve states are considered non-disclosure states, or states that do not publicly report or disclose the sales prices of real estate transactions. The non-disclosure states are Alaska, Idaho, Kansas, Louisiana, Mississippi, Missouri (some counties), Montana, New Mexico, North Dakota, Texas, Utah, and Wyoming. The lack of sales data in these states would require a mail survey of buyers and sellers to be conducted in these areas, if we were to conduct a TPV survey.

Analysis of Assessment Files

The vendor data were evaluated state by state to better understand what information was available and to discover the underlying metadata structure. One of the primary objectives of this analysis was to understand what the assessed values actually represented. The assessment value in a vendor file could represent the market value of the parcel, a fractional market value of the parcel, or a taxable value after exemptions have been deducted.

The evaluation of the exemption versus fractional assessment began with a comparison of the assessed values in the vendor files to the market values in the vendor files. If the values were the same then it was assumed that no exemptions or fractional assessments were accounted for; if the

values were different, then it was probable that exemptions or fractional values accounted for the difference. Data values for individual parcels were verified using the parcel search function found on local government websites, if available. This verification process showed what data were being captured in the vendor data file and if any data were missing.

In most jurisdictions where properties are assessed using full market value or fractional assessments, the corresponding vendor files generally reflected the correct value. However, some jurisdictions that should have shown fractional assessments or full market value did not. Thus, all jurisdictions in each state still had to be inspected for exceptions.

Exemptions that reduce the assessment value of a parcel were rarely reflected in the vendor assessment data. Independently calculating exemption values would be an onerous task. The laws used to determine exemptions in the various jurisdictions are too vast and complex to allow the Census Bureau to determine which parcels of property should reflect which type of exemption.

When forming the metadata for the vendor files, each state had to be examined individually. No two states followed a similar pattern; similar patterns were found within a state, but not necessarily for all jurisdictions. Metadata must be defined for each vendor assessment file before that assessment data be used in a TPV survey. In addition, the metadata structure found in the 2009 vendor files may not stay consistent in future years. Similar research would be required for *all* forthcoming vendor assessment files before their metadata patterns could be discovered.

The analysis of the vendor's assessment files showed that data quality issues must be resolved before they could be used as a data source for a TPV survey. In addition, vendor sales files must also be edited and tested for completeness. The assessment values in the assessment data files must be defined in each vendor file on a jurisdiction-by-jurisdiction basis. Finally, supplemental data collection would be necessary for the areas that the vendor does not sufficiently cover.

DIRECT DATA COLLECTION FROM JURISDICTIONS

The feasibility study evaluated direct data collection from local jurisdictions as another option for a TPV survey. Though third party vendors have

potential as a mostly comprehensive data source, a TPV survey would also need to use direct data collection in some situations. In particular, direct data collection would be required for jurisdictions in non-disclosure states and jurisdictions not covered by another source.

This evaluation began with a request letter and instructions for submitting data that was sent to each assessment and deeds jurisdiction in two representative samples. These samples were not designed to give estimates of totals. They can be used for comparisons of the vendor files to the jurisdictions files and to evaluate our ability for direct collection. The letter requested calendar year 2009 data for variables including, but not limited to, parcel ID number, 2009 gross assessed value of land and improvements, valuation date, property use class, parcel address, property owner name, 2009 sales amount, sales date, and deed type. Though two electronic file formats were originally specified (text files, Microsoft Excel spreadsheets), any data submission was accepted.

In response to the request letters and non-response follow up, 51.7 percent of sampled assessment jurisdictions and 43.0 percent of sampled deeds jurisdictions provided data. The response time and non-response follow-up for this survey was six weeks. Given the time and resource restrictions for this sample survey, the response rates observed for this feasibility study may not be indicative of expected response rates for a future TPV survey.

Non-response was the most frequent cause of missing data. Nonresponse jurisdictions included those that did not reply to multiple follow-up attempts, jurisdictions that agreed to participate but did not send data before the close of the survey, and jurisdictions that declined to participate. The reasons jurisdictions declined to participate ranged from a lack of resources available to fulfill the request, the voluntary nature of the survey, non-disclosure laws in the state, and questions about the authority to collect sales data.

Jurisdictions in 12 states did not provide sales information due to non-disclosure of sales data policies. In non-disclosure states, transaction sales prices are not available to the public; therefore, sales data were not provided by the sampled jurisdictions. Another survey would have to be developed to collect sales price information directly from buyers and sellers of sold parcels in these states.

Costs associated with providing jurisdiction data were another cause of missing data. During the direct data collection, 14 jurisdictions quoted a cost associated with data access, ranging from \$60 to \$3,000. Due to set payment limits and complications with payment processing, no jurisdictions were paid for data. Three jurisdictions opted to supply data without payment, while the other jurisdictions declined to do so.

Timeliness was another cause of missing data. During nonresponse follow-up, a number of jurisdictions were willing to participate, but could not provide data during the requested time period due to resource constraints. In other situations, jurisdictions did not maintain historical records and could only provide current year data (2010), rather than the 2009 calendar year data that were requested.

A lack of relevant variables also limited the usability of jurisdiction data. We experienced difficulty using the file if they had missing variables, particularly the primary matching variable (APN). Missing APN variables make it difficult to identify properties or match assessment and sales parcel information. The second most common missing variable was the land use code. A missing land use code would make it difficult to segregate assessment variables by land use.

The formats of data provided by the jurisdiction also proved to be challenging. These issues usually involved unknown formats of jurisdiction data files, as these files were often transmitted with incorrect filename extensions or no filename extensions at all. Each time this occurred, an analyst had to investigate possible formats for the jurisdiction file. Sometimes, these files could be opened as an Excel spreadsheet or text file, but for a limited set of jurisdiction files the format could not be determined, even after hours of investigation.

Usability problems with the jurisdiction files point to a larger issue that once afflicted the historical TPV survey and cannot be ignored now: a lack of standard data formats. If standard formats could be established for taxable property data in jurisdiction files, direct data collection could become a more practical option. Without standard formats, the Governments Division faces a formidable task: parsing relevant data from the great diversity of jurisdiction files and then forcing the data into a consistent format. Unfortunately, the required workload for this task increases with the number of jurisdictions in a sample. The workload needed to

process data from the diverse array of jurisdiction file formats means it would be very challenging to scale the required resources for a TPV survey to the point it could produce national estimates.

CONCLUSIONS

In this feasibility study, we evaluated three data collection methods that could be used for a TPV data product. Jurisdiction websites, data provided by a third party vendor, and direct collection from local jurisdictions were evaluated by availability, resource requirements, data quality, and compliance with Census Bureau statistical standards. None of the three methods independently offered the completeness and quality needed for TPV. A combination of collection methods would be needed to produce a comprehensive data product that meets statistical standards. This would require a level of effort beyond current resources. From this feasibility study, we concluded the following:

- The lack of jurisdiction websites that can provide sufficient data (both current and historical) in a usable format make website data collection an unsuitable option for a TPV survey.
- The Census Bureau would not be able to confirm the quality of the vendor data each year until a series of edit tests are run on each of the files. The tests would help to clarify what type of data are reported, missing, and what each value actually represents (i.e., market value, assessed value, etc.). This would be a time-consuming process requiring a manual, detailed analysis of each vendor file.
- Exemptions given to homeowners in select categories that reduce assessment value are typically not reflected in the assessment value in the vendor data file. The Census Bureau would have to consider those when defining taxable property values.
- The primary challenges in using a direct data collection method included: nonresponse, non-disclosure, format of data files, cost, timeliness of data, and availability of data variables. Collecting data through government surveys is resource intensive, and these factors would have to be considered when using direct data collection methods to gather or supplement data.
- Jurisdictions in twelve states did not provide sales information because of non-disclosure of sales data policies; therefore, sales data would have to be provided willingly from the buyer or seller of the parcel via a separate survey, which has not been developed.

Acknowledgements

The author(s) would like to thank Lisa Blumerman, Carma Hogue, and Carrie Dennis for their comments on this paper. Any errors that remain in this paper are the sole responsibility of the authors.

Notes

- ¹ For more information see: <http://www.census.gov/govs/estimate/>
- ² Additional TPV literature can be found in the U.S. Census Bureau technical paper, *Taxable Property Values Exploratory Research Study* (released in June 2011), and *Taxable Property Values Feasibility Study Literature Review*, pending release in the Governments Division Report Series.

References

National Research Council. *State and Local Government Statistics at a Crossroads*. Washington, DC, 2007: 68.