Highlights
Conversion to New Software for Address Processing

Address Processing: Background

Property address information is critical to construction of FHFA's House Price Index (HPI). The underlying "repeat-transactions" indexing model requires the identification of homes that have had two or more historical transactions. When historical transactions data are electronically processed and "transaction pairs" are constructed reflecting price changes for the same property over time (a key model input), it is important that property address information be consistently formatted. Without standardized formatting, much data would be lost in the property matching process. For example, a search procedure encountering the data below might not determine that the transactions all came from the same address.

Address	Transaction Date	House Price
123 Oak Street	January 1, 1975	$50,000
123 Oak Street	March 12, 1982	$125,000
123 Oak St.	July 27, 1990	$200,000
123 Oak Str.	January 1999	$250,000
123 Oak Strt.	December 2, 2008	$375,000

Without a standardized abbreviation for the word "street," only the first two records above would deemed to be the same address and would be used in index estimation.

Consistent with industry practice, FHFA uses software to perform the necessary "address scrubbing" for standardization. The software works by comparing input addresses against a known set of valid addresses. Where input addresses are sufficiently similar to "known" address, the known address—which is held a database and has a standardized address format—is used. Proprietary algorithms are used to determine whether the known (well-formatted) address is the same property.

In addition to address validation and standardization, the same software is also used by FHFA to "geocode" properties. Geocoding entails supplementing the address information in a given record (e.g., ZIP code, city, and state) with additional detail about the area (e.g., Census Tract, county, Metropolitan Statistical Area). Geocoding is important to FHFA's HPI production because it allows for the construction of Metropolitan Statistical Area (MSA) indexes; MSA identifiers are not available in the raw transactions data received by FHFA and thus MSA indexes could not be produced without the geocoding.

New Software Tool

As indicated in the associated press release, beginning with the production and release of the monthly HPI for November,[1] FHFA began using a new proprietary software tool for address scrubbing and geocoding. In general, the new software performs better than the old tool. With improved logic and a larger database of valid address to search through for processing records, the new software generally increases the number of valid addresses that can be used

---

[1] See http://www.fhfa.gov/webfiles/15372/Monthly%20HPI%201%2026%2010.pdf.

in index construction.[2] In addition, the software more accurately geocodes properties. FHFA's close inspection of the MSA assignments suggests that, in the limited number of cases where the old and new applications provide different MSA assignments to the same property, the new tool's assignment was consistently more accurate.[3]

Statistics: Record Counts for Old and New Software

The new software tool produces appreciable, albeit modest, increases in the number of records deemed to have "valid addresses."[4] The estimation of the HPI is restricted to such records and thus an increase in such record counts is generally desirable.

The national sample size of purchase mortgages increased by about 0.5 percent with the use of the new software.[5] The state-specific impacts varied considerably, however. Fourteen states had sample size increases exceeding one percent, with Arizona and Vermont showing the greatest improvements (4.4 percent and 3.7 percent respectively). By contrast, in two states—West Virginia and Hawaii—sample sizes declined by more than one percent. Specific results for every state are available here.

Because indexes for metropolitan statistical areas are generally constructed using both sales prices and appraisal values from refinance mortgages, the sample size changes for those areas were evaluated with observation counts from both types of mortgages. The impact of the new tool on the total sample size for most metropolitan areas was generally limited. Small increases were evident in most areas, but observation counts declined in some locations. Sample size changes were notable in a number of Florida areas; five of the ten metropolitan areas having the largest percentage increases in sample sizes were in Florida. The full list of metropolitan areas, with their respective sample size changes, is available here.[6]

Statistics: Effects on Index Estimates

Not surprisingly, given the relatively modest impact of the new software on sample sizes, there were relatively limited differences in the index values produced with the respective software applications.   For Census Divisions, for instance the four-quarter appreciation rates computed with the old and new address processors were generally within rounding error.[7] The state estimates reveal more sizeable effects in some areas, but still a relatively limited divergence in most cases. When assessed with FHFA's purchase-only house price indexes, the largest divergence in the four-quarter price change estimates was for Hawaii. The new software produced a 0.7 percent smaller four-quarter price decline than the old software. The impact of the new software on the sample size for Hawaii was relatively large and thus the relatively significant impact was not unexpected.

---

[2] An increase in the sample size can lead to greater precision in estimates of index values.
[3] After randomly selecting 30 records for which metropolitan area assignments were different across the two applications, FHFA used online property lookup tools to determine which assignment was more likely to be accurate. In 26 of the 30 cases, the new tool appeared to provide the correct assignment.
[4] Given that matching the unstandardized input addresses to a set of valid addresses is a complex, imperfect process, it should be recognized that the records that are ultimately "unmatched" are not necessarily indicative of data errors.
[5] These mortgages are used in the formation of FHFA's "purchase-only" house price indexes.
[6] When measured with the "all-transactions" data sample used for the metropolitan area indexes, the overall national sample size increased by 0.6 percent (slightly more than the 0.5 percent increase posted with the purchase-only series).
[7] The respective estimates from the purchase-only indexes are available here.

Like Hawaii, many of the other states exhibiting above-average impacts were those with relatively small starting sample sizes. A full listing of [state results](#) reveals that Nebraska, Delaware, Rhode Island, Kansas and the District of Columbia had larger-than-average differences in their four-quarter price change estimates. In absolute terms, however, the effects even in those locations were small, with absolute effects on four-quarter price change estimates of between 0.4 and 0.6 percentage points.

Conclusion

The results discussed here, it should be noted, are consistent with evaluations done on data from prior periods. Before transitioning to the new software, FHFA reviewed its impact over prior months and quarters and generally found small improvements in transaction counts and modest differences in index estimates. Inasmuch as these changes, though not dramatic, improve the reliability and precision of the HPI, FHFA has determined that the conversion to the new software tool is warranted.