

## **Neighborhood Stabilization Program - Revised 10-20-08**

### **Methodology and Data Dictionary for HUD Provided Data**

#### *Background*

Using data from the Mortgage Bankers Association National Delinquency Survey as of June 2008, HUD has calculated the approximate number of foreclosure starts for all of 2007 and the first six months of 2008 ("Foreclosure Starts over 18 months") at the statewide level.

The Mortgage Bankers Association (MBA) data are not available for geographic areas smaller than states. As such, HUD has identified data collected by other federal agencies that prove to be good predictors of where foreclosures are likely. HUD has used those data to "distribute" the statewide counts of foreclosure starts among the neighborhoods, places, and counties within each state.

To test the reliability of HUD's estimated foreclosure rate at the local level, HUD asked the Federal Reserve to compare HUD's estimate to data the Federal Reserve had from Equifax showing the percent of households with credit scores that were delinquent on their mortgage payments 90-days or longer. The Equifax data are based on a 5 percent sample of all credit records in the United States. As such, they are more reliable for counties with higher population counts (a larger sample size reduces sampling error) than those with smaller population counts. At the statewide level, 90-day delinquencies from Equifax and the MBA data on foreclosure starts are closely related, that is they have a very high correlation with one another (0.90 where 1 is a perfect correlation).

Analysis by the Federal Reserve staff found that when comparing the HUD predicted county foreclosure rates to the Equifax county level rates of delinquencies, HUD's data and the Equifax data had high intrastate correlations. For example, within the state of California, the correlation was 0.835 (where 1 is a perfect correlation). The county level intrastate correlations were higher when the analysis was restricted to counties with greater than 15,000 households. There are reasons that either the HUD estimated foreclosure rate or Equifax delinquency data could be wrong, but when they are very similar to one another in a particular community we have a greater confidence that HUD's estimated foreclosure rate and the Equifax delinquency data are accurately targeting the problem.

HUD also obtains data from the United States Postal Service (USPS) on addresses that have been vacant for 90-days or longer. The USPS collects these data to reduce delivery of bulk mail to homes where no one is picking up the mail. While there are many reasons for homes being vacant for 90-days or longer, HUD believes that if a Census Tract is found to be estimated to have a higher rate of foreclosures and it has a high rate of homes 90-days or more vacant, abandonment risk associated with the foreclosure crisis is higher in those neighborhoods.

HUD is providing its data on estimated foreclosures (based on risk) and vacancy data to assist state and local governments in their efforts to target the communities and neighborhoods with the greatest needs. HUD recommends that if states and local governments have local data, such as county data on foreclosure filings, that those data also be given serious consideration in identifying areas of greatest needs.

HUD has created data files at several areas of geography to assist local and state governments:

- (1) County
- (2) County-Place
- (3) Census Tract
- (4) Block Group (part)

The County, County-Place, and Census Tract files contain the same data:

- Estimated number and percent of foreclosure starts over the past 18 months through June 2008
- Number and percent of vacant addresses in June 2008

- Data used to calculate the estimated foreclosure rates
  - Federal Reserves Home Mortgage Disclosure Act Data on high cost loans
  - Office of Federal Housing Enterprise Oversight Data on falling home prices
  - Bureau of Labor Statistics data on place and county unemployment rates

The Block Group (part) file includes:

- Number and percent of persons estimated at less than 120 percent of median income
- A “foreclosure and abandonment risk score” that is a function of the estimated foreclosure rate and percent of addresses vacant
- Percent of foreclosure starts over the past 18 months through June 2008
- Percent of vacant addresses in June 2008
- Data used to calculate the estimated foreclosure rates
  - Federal Reserves Home Mortgage Disclosure Act Data on high cost loans
  - Office of Federal Housing Enterprise Oversight Data on falling home prices
  - Bureau of Labor Statistics data on place and county unemployment rates

### *Methodology*

All of the files provide estimates of foreclosures based on a formula that calculates the rate of foreclosure starts over the past 18 months as a function of:

- Metropolitan area decline in home values as of June 2008 against peak home values in June of any previous year between 2000 and 2008. If home values have not declined, it is zero. These data are from the Office of Federal Housing Enterprise Oversight (OFHEO)<sup>1</sup> Home Price Index. Data for non-metropolitan balances of states are from the March 2008 Home Price Index.
- County or Place Level unemployment rate as of June 2008 from the Bureau of Labor Statistics Local Area Unemployment Rate data.
- Census Tract Level Data on number of loans made between 2004 and 2006 from the Home Mortgage Disclosure Act (HMDA) data and the number of those loans that are high cost (where the rate spread is 3 percentage points above the Treasury security of comparable maturity).

A regression using statewide data on foreclosures from the Mortgage Bankers Association National Delinquency Survey against the factors above creates the following model:

Predicted Foreclosure Start Rate= -2.211

- (0.131\*Percent change in MSA OFHEO current price relative to the maximum in past 8 years)
- + (0.152\*Percent of total loans made between 2004 and 2006 that are high cost)
- + (0.392\*Percent unemployed in the place or county in June 2008).

The regression used to calculate this model found that it predicted 75 percent (R-square of 0.750) of the variance in foreclosure start rates between states. This means that this is a very strong model for predicting foreclosure starts, but there are other reasons not accounted for in the model (the 25 percent of the variance we don't account for) that could lead to a community having a higher or lower foreclosure rate than what is predicted by the model.

The number of mortgages for a jurisdiction is based on its proportional share of loans made between 2004 and 2006 within the state (from HMDA) times the total number of mortgages in the state (from

<sup>1</sup> Now the Federal Housing Finance Agency (FHFA). Data available from [www.ofheo.gov](http://www.ofheo.gov).

American Community Survey 2006 on homeowners with a mortgage adjusted by HMDA data on fraction of investor loans).

The number of foreclosures for a jurisdiction is weighted to reflect the statewide totals of foreclosure starts over 18 months from the Mortgage Bankers Association National Delinquency Survey through June 2008.

As noted above, staff at the Federal Reserve Board compared HUD's estimated foreclosure start rates at the county level with the Equifax data which are based on a 5 percent sample of credit records (and thus suffer from sampling error). If the pattern of high rates of 90-day delinquencies from the Equifax data matches with the pattern of high foreclosure rates as estimated by HUD, we have a higher degree of confidence that the both the HUD estimate and the Equifax data are reasonably accurate.

States with very high rates of correlation between HUD's foreclosure rate estimates and Equifax 90 day delinquencies (correlation of 0.80 or higher) are California, Connecticut, Hawaii, Maryland, New Jersey, Rhode Island, and South Carolina. States with a modestly high rate of correlation (correlation 0.60 to 0.79) are Arizona, Florida, Massachusetts, Michigan, and South Dakota.

While most of the remaining states had correlations that were positive and significant, the correlations were lower. The reason for the lower rate of correlations could be because the model HUD is using to estimate foreclosure rates does not account for the factor or factors most contributing to foreclosures in that state, the sampling errors in the Equifax data makes the comparison data inaccurate, there is not enough variation between counties on the data in the model to show significant variations in county foreclosure rates, or some other reason. Notably, intrastate correlations between the HUD estimated foreclosure rate and the Equifax data improve dramatically when only counties with more than 15,000 households are included in the analysis. When making this restriction, 23 intrastate correlations are greater than 0.6 (see Appendix 1). Since the Equifax data are sample data, their accuracy is improved by having a larger N while the HUD model is also more accurate for communities within the metropolitan areas that OFHEO calculates price change information.

All grantees are advised to look to other local data when considering their areas of greatest need, particularly if they are not among the states listed as having high rates of intrastate correlation between the HUD estimated foreclosure rate and the Equifax 90-day delinquency data. Even in states with relatively low correlation, HUD believes that the data it is providing are useful for identifying areas state and local governments should review as possible candidates for targeting funds because they have underlying characteristics that make them at significant risk for foreclosures and abandoned homes.

#### *Data Dictionary for County, County-Place, and Tract Files*

Geographic Identifiers in Each File Are As Follows:

##### County Level File

- **countycode** - 5 character combination of state and county FIPS codes
- **state** - 2 character state FIPS code
- **sta** - 2 character state alphanumeric abbreviation
- **county** - 3 character county FIPS code
- **countyname** - county name

##### County-Place Level File

- **countyplace** - 10 characters. For CDBG Entitlement Cities, this is the CDBG ID. For Urban Counties and State Nonentitlement Areas, is a concatenation of state, county, and place FIPS codes
- **cdbguogid** - the unique ID for a CDBG Entitlement Area
- **name** - Name of the CDBG Entitlement Area

- **state** - 2 character state FIPS code
- **sta** - 2 character state alphanumeric abbreviation
- **county** - 3 character county FIPS code
- **countyname** - county name
- **place** - 5 character place FIPS code
- **placnm** - place name

#### County Level File

- **Tractcode** - 11 character combination of state, county, and Census Tract codes
- **state** - 2 character state FIPS code
- **sta** - 2 character state alphanumeric abbreviation
- **county** - 3 character county FIPS code
- **countyname** - county name
- **tract** - 6 digit Census Tract Code

#### The County, County-Place, and Tract Level Files All Have the Following Variables

- **hhuniv** - count of households from Census 2000
- **estimated\_number\_foreclosures** - HUD model, estimated count of foreclosure starts over 18 months through June 2008. Note caveats above.
- **estimated\_number\_mortgages** - HUD estimated number of mortgages as described above.
- **estimated\_foreclosure\_rate** - number of foreclosure starts divided by number of mortgages times 100.
- **total\_90\_day\_vacant\_residential\_addresses** - United States Postal Service data from June 2008 on residential addresses vacant 90-days or longer.
- **total\_residential\_addresses** - United States Postal Service Data on total addresses as of June 2008
- **estimated\_90\_day\_vacancy\_rate** - addresses vacant 90 days or longer divided by total addresses times 100.
- **total\_hicost\_2004\_to\_2006\_HMDA\_loans** - Total number of conventional loans made between 2004 and 2006 where Home Mortgage Disclosure Act where the rate spread is 3 percentage points above the Treasury security of comparable maturity.
- **total\_2004\_to\_2006\_HMDA\_loans** - Total number of conventional loans made between 2004 and 2006 according to data from the Home Mortgage Disclosure Act
- **estimated\_hicost\_loan\_rate** - percent of loans made between 2004 and 2006 shown to be high cost according to HMDA data.
- **bls\_unemployment\_rate** - June 2008 place or county unemployment rate. If unemployment data available for a place, the place level unemployment rate is used. For places and balance of county place level unemployment data unavailable, the county level unemployment rates is used.
- **ofheo\_price\_change** - a measure of price decline in home values that uses data from the Office of Federal Housing Enterprise Oversight (OFHEO) Housing Price Index (HPI) to calculate price decline from peak value in the second quarter of any year between 2000 and 2008 and the second quarter home price in 2008.

#### *Data Dictionary for Block Group (part) file*

Note, the unique geographic identifier for a record is the combination of:

State, county, place, county subdivision, tract, Urban/Rural (UR), block group

- **Cdbguogid** - the unique ID for a CDBG Entitlement Area
- **Cdbgname** - Name of the CDBG Entitlement Area
- **cdbgtype**
- **sta** - 2 character state alphanumeric abbreviation

- **logrecno** - unique record identifier from Census Bureau data files
- **state** - 2 character state FIPS code
- **county** - 3 character county FIPS code
- **countyname** - county name
- **cousub** - 5 character FIPS county subdivision code
- **cousubname** - county subdivision name
- **place** - 5 character place FIPS code
- **placename** - place name
- **tract** - 6 character Census Tract Code
- **blkgrp** - 1 character block group code
- **UR** - Urban/Rural classification where, an Urban Area is defined to encompass densely settled territory, which consists of (1) core census block groups or blocks that have a population density of at least 1,000 people per square mile and (2) surrounding census blocks that have an overall density of at least 500 people per square mile. In addition, under certain conditions, less densely settled territory may be part of each UA or UC. Any area not meeting this classification is rural. Part of a block group within a city can be rural if it does not have a density of 1,000 people per square mile and the surrounding blocks have a density of less than 500 per square mile. In other words a city "block" where few or no people lived in 2000 surrounded by other blocks with relatively few people could lead to a "rural" designation. For example, an area that is largely industrial might get categorized as "Rural". Parkland within a city might be counted as rural. Not surprisingly, you will notice that when an area is defined as "Rural" within a city, it often shows up with 0 for "total persons".
- **middle\_low\_mod\_eligible** - "Y" if area qualifies for Low- Moderate- Middle-Income area benefit
- **Estimated\_foreclosure\_abandonment\_risk\_score** - a score of 1 to 10, where 10 indicates that the area is in the highest 10 percent of risk nationwide for foreclosure and abandonment based on the combination of HUD's foreclosure risk estimate and vacancy rate. 1 indicates the lowest risk.
- **Percent\_lt\_120\_AMI** - percent of persons estimated to be less than 120 percent of Area Median Income in the area.
- **Persons\_lt\_120\_AMI** - number of persons estimated to be less than 120 percent of Area Median Income in the area.
- **Total\_Persons** - Total persons in 2000 in the area.
- **OFHEO\_CBSA\_home\_price\_decline\_since\_peak** - a measure of price decline in home values that uses data from the Office of Federal Housing Enterprise Oversight (OFHEO) Housing Price Index (HPI) to calculate price decline from peak value in the second quarter of any year between 2000 and 2008 and the second quarter home price in 2008.
- **BLS\_place\_or\_county\_unemployment\_rate\_0608** - June 2008 place or county unemployment rate. If unemployment data available for a place, the place level unemployment rate is used. For places and balance of county place level unemployment data unavailable, the county level unemployment rate is used.
- **HMDA\_hi\_cost\_loan\_rate** - Percent of conventional loans made between 2004 and 2006 from Home Mortgage Disclosure Act data where the rate spread is 3 percentage points above the Treasury security of comparable maturity. Calculated at Census Tract level.
- **predicted\_18\_month\_underlying\_problem\_foreclosure\_rate** - HUD model, estimated count of foreclosure starts over 18 months through June 2008 divided by estimated number of mortgages times 100. Calculated at Census Tract level. Note caveats above.
- **USPS\_residential\_vacacancy\_rate** - United States Postal Service data from June 2008 on residential addresses vacant 90-days or longer divided by total residential addresses. Calculated at Census Tract level. Note caveats above.

#### Appendix 1: Pearson Correlation Comparison of HUD County Foreclosure Rate Estimate to Equifax 90-day mortgage delinquency sample data for Counties with over 15,000 Households

Correlation when restricted to counties above 15,000			Correlation when restricted to counties above 15,000			Correlation when restricted to counties above 15,000		
State	households	N (counties greater than 15,000 households)	State	households	N (counties greater than 15,000 households)	State	households	N (counties greater than 15,000 households)
AK	***	4	KY	***	21	NY	0.545	55
AL	***	32	LA	***	27	OH	0.556	62
<b>AR</b>	<b>0.606</b>	<b>18</b>	<b>MA</b>	<b>0.779</b>	<b>12</b>	OK	***	22
<b>AZ</b>	<b>0.823</b>	<b>11</b>	<b>MD</b>	<b>0.874</b>	<b>18</b>	OR	0.587	18
<b>CA</b>	<b>0.862</b>	<b>45</b>	ME	***	10	<b>PA</b>	<b>0.632</b>	<b>55</b>
<b>CO</b>	<b>0.735</b>	<b>15</b>	<b>MI</b>	<b>0.777</b>	<b>40</b>	PR	***	0
<b>CT</b>	<b>0.840</b>	<b>8</b>	MN	0.466	22	<b>RI</b>	<b>0.942</b>	<b>5</b>
DC	***	1	MO	0.587	25	<b>SC</b>	<b>0.788</b>	<b>25</b>
DE	***	3	<b>MS</b>	<b>0.625</b>	<b>16</b>	<b>SD</b>	<b>1.000</b>	<b>2</b>
<b>FL</b>	<b>0.799</b>	<b>43</b>	MT	***	6	TN	0.570	37
<b>GA</b>	<b>0.676</b>	<b>42</b>	<b>NC</b>	<b>0.641</b>	<b>58</b>	TX	0.428	63
<b>HI</b>	<b>0.970</b>	<b>4</b>	ND	***	4	UT	***	6
<b>IA</b>	<b>0.750</b>	<b>17</b>	<b>NE</b>	<b>0.821</b>	<b>5</b>	VA	0.531	40
<b>ID</b>	<b>0.876</b>	<b>7</b>	NH	-0.678	9	VT	***	6
IL	0.593	34	<b>NJ</b>	<b>0.920</b>	<b>21</b>	<b>WA</b>	<b>0.619</b>	<b>23</b>
IN	0.591	37	NM	***	13	WI	0.468	35
<b>KS</b>	<b>0.976</b>	<b>11</b>	<b>NV</b>	<b>0.883</b>	<b>5</b>	WV	0.594	14
***If data not shown, it was not statistically significant at the .05 level or better						WY	-1.000	2

States with correlations 0.6 or higher shown in bold.



























