

Subject Line: FREE Data for AREUEA Members

Dear AREUEA Member,

Many of you have seen our recent “lock-in” paper quoted in the NY Times, San Francisco Chronicle, Boston Globe, Financial Times (London) and Washington Post. We’ve taken the time to create two datasets from the National Mortgage Database® (NMDb), a nationally representative five percent sample of closed-end first-lien residential mortgages in the United States, that you can use in your original research. We only ask in return that you cite our paper, but otherwise, there are no strings attached.

Below, we share FAQs so you know exactly what we have created for you. All of this information is also available on the working paper page (the link is right below).

Sincerely,

Ross M. Batzer (FHFA), Jonah R. Coste (FHFA), William M. Doerner (FHFA), and Michael J. Seiler (FHFA/William & Mary)

Frequently Asked Questions

FHFA WORKING PAPER:

- <http://www.fhfa.gov/papers/wp2403.aspx>

DATASETS:

- <https://www.fhfa.gov/PolicyProgramsResearch/Research/PaperDocuments/wp2403-lock-in-data.xlsx>
- <https://www.fhfa.gov/PolicyProgramsResearch/Research/PaperDocuments/wp2403-lock-in-figures.xlsx>

Q1: What is mortgage rate lock-in?

A: In the United States, nearly all mortgages have fixed rates. In times of rising rates, such as the present, most mortgages have interest rates below prevailing market rates. In almost all cases, selling would require forfeiting this below-market rate. Consequently, many borrowers become “locked-in” to their current home and mortgage.

Q2: Are these official FHFA statistics?

A: No, they should be considered as developmental statistics. These metrics and data have been produced as part of [FHFA Working Paper 24-03](#) to stimulate discussion and comment. Please cite this working paper when using the data so we can more easily follow how they are used and make improvements if needed.

Q3: What is contained in each file?

A: The wp2403-lock-in-data.xlsx file contains estimates of lock-in exposure, sensitivity, and the effect on sales over time for different geographies and demographic groups. The wp2403-figures.xlsx file presents the data from several figures in the working paper in tabular form. Figures showing output of the theoretical model or distribution densities are not included.

Q4: What data are used to make the estimates?

A: Market-wide measures of mortgage activity come from the National Mortgage Database® (NMDb), a nationally representative five percent sample of closed-end first-lien residential mortgages in the United States.¹ The effects of lock-in on the probability of sale are estimated using proprietary Government Sponsored Enterprise (GSE) data. Data for recent quarters are preliminary and subject to revision.

Q5: Are there variable descriptions for the dataset?

A: Yes. Each begins with fields for year and quarter and defines the category or segment (as discussed in Q6). Each tab then displays fields for the variables in the table below.

Variable Name	Description
<i>Loans</i>	The number of fixed-rate mortgages active at the end of the quarter. Rounded to the nearest 1,000.
<i>Fixed_Rate</i>	The average interest rate on all fixed-rate mortgages active during the quarter. Rounded to the nearest basis point.
<i>New_Rate</i>	The average predicted interest rate for all active fixed-rate mortgages if they were re-originated during the quarter. Rounded to the nearest basis point.
<i>Rate_Delta</i>	The amount by which the average fixed rate exceeds (or falls short of) the predicted new rate. Rounded to the nearest basis point.
<i>Delta_lt_neg3, Delta_neg3_neg2, ..., Delta_gte_pos3</i>	These variables show the distribution of rate deltas for all active fixed-rate mortgages in one-point buckets. Percentages are rounded to the nearest .1%.
<i>Payment_Change</i>	The average predicted change in the monthly principal and interest payment if active loans were re-originated at their expected new rate, expressed in dollars. Rounded to the nearest dollar.
<i>Payment_Change_Percent</i>	The average predicted change in the monthly principal and interest payment if active loans were re-originated at their expected new rate, expressed as a percent of the current principal and interest payment. Rounded to the nearest .1%.

¹ The NMDb is maintained jointly by the Federal Housing Finance Agency (FHFA) and the Consumer Financial Protection Bureau. More information can be found at <https://www.fhfa.gov/PolicyProgramsResearch/Programs/Pages/National-Mortgage-Database.aspx>.

<i>Relative_Sensitivity</i>	The average estimated effect on the relative probability of sale for a one-point increase in rate delta (for rate deltas ≤ 1). Rounded to the nearest .1%.
<i>Lock_In_Effect</i>	The average estimated effect on the relative probability of sale, given the estimated rate deltas in the quarter. These estimates account for the non-linear relation between sales and rate deltas. The lock-in effect for an individual loan is capped at -100%. Rounded to the nearest .1%.
<i>Sales_Change_Quarter</i>	The estimated number of arms-length sales gained (or lost) due to the lock-in effect in the quarter. Rounded to the nearest 10.
<i>Sales_Change_Since_2022Q2</i>	The cumulative number of arms-length sales gained (or lost) due to the lock-in effect since April 2022. Rounded to the nearest 10. ²

Q6: Are there descriptions for each tab in the dataset?

A: Yes. Each tab contains estimates of each variable over time segmented by geography or demographics, as described in the table below.

Tab Name	Description
<i>Overall</i>	Contains data for all fixed-rate mortgages in the United States (50 states plus the District of Columbia) as a whole.
<i>State</i>	Contains data segmented by state.
<i>MSA</i>	Contains data segmented by metropolitan statistical area (MSA). Estimates are calculated for each of the 150 MSAs with sufficient data. Together, these MSAs contain 75.2% of the quarterly mortgage records in the data. Other mortgages are classified as belonging to either “Other MSAs” or “Micropolitan areas and non-metro counties.”
<i>MSA_All</i>	Contains data segmented by MSA. Estimates of lock-in exposure are calculated for all 387 MSAs in the data. Estimates of lock-in sensitivity and the effect on sales are provided only for the 150 MSAs identified in the “MSA” tab. Other mortgages are classified as belonging to “Micropolitan areas and non-metro counties.”
<i>County</i>	Contains data segmented by county or county equivalent. Estimates of lock-in sensitivity are calculated for all 3,142 counties in the data (subject to censoring, see Q7). Estimates of lock-in sensitivity and the effect on sales are provided only for the 380 counties with sufficient data. These counties contain 72.0% of the quarterly mortgage records in the data.
<i>Loan_Type</i>	Contains data segmented by loan type. The five loan types are: <ul style="list-style-type: none"> - GSE: Loans acquired by Fannie Mae or Freddie Mac - Conventional: Non-government-insured mortgages not acquired by Fannie Mae or Freddie Mac - FHA: Loans guaranteed by the Federal Housing Administration - VA: Loans guaranteed by the Department of Veterans Affairs

² Rounding for the *Sales_Change_Since_2022Q2* variable occurs after the cumulative number has been calculated. Therefore, there can be slight differences between the cumulative change and the sum of the quarterly changes since 2022Q2.

	- USDA: Loans guaranteed by the U.S. Department of Agriculture through its Rural Housing Service (RHS) and Farm Service Agency (FSA) programs.
<i>Home_Value</i>	Contains data segmented by original appraisal value (adjusted to 2022 prices using the national all-transaction FHFA HPI®).
<i>Borrower_Income</i>	Contains data segmented by annual borrower income (adjusted for inflation using the Consumer Price Index for All Urban Consumers).
<i>Borrower_Age</i>	Contains data segmented by borrower age. Borrower age is defined as the age of the first borrower at origination plus the loan's age in years.
<i>Race_Ethnicity</i>	Contains data segmented by the primary race and ethnicity of the first borrower.
<i>Scheduled_LTV</i>	Contains data segmented by scheduled loan-to-value ratio (LTV). Scheduled LTV is the scheduled unpaid principal balance divided by the original appraisal value.
<i>Credit_Score</i>	Contains data segmented by the lowest credit score amongst all borrowers on a loan.

Q7: Why are some rows and counties missing in the “County” tab?

A: We only show estimates when there are at least 25 records (corresponding to ~500 loans) for a category in a given quarter. This removes 13.6% of all rows in the county tab but only 0.2% of all quarterly loan records. This limitation only affects the “County” tab, as all other segments contain sufficient records.

Q8: Are any statistics smoothed or adjusted for seasonality?

A: Yes. The “Sales_Change_Quarter” and “Sales_Change_Since_2022Q2” fields apply the estimated lock-in effect to an underlying sale rate that removes seasonal variation. All other fields are unadjusted and do not exhibit strong seasonal patterns.

Q9: Are filters used in producing the statistics or estimates?

A: The data contain estimates for all fixed-rate mortgages in the U.S. The models used to make these estimates are calibrated with filtered data. Please see the working paper for more information on these filters.

Q10: Data are reported by quarter. When in the quarter are calculations performed?

A: The “Loans” field contains a count of active loans at the end of the quarter. The “Sales_Change_Quarter” field is a count of estimated sales gained or lost during the quarter due to lock-in. The “Sales_Change_Since_2022Q2” expresses this as a cumulative total. All other fields are quarterly averages for loans active during the quarter.

Q11: How do the “Fixed_Rate” and “New_Rate” fields compare to other data on average mortgage rates, such as Freddie Mac’s Primary Mortgage Market Survey® (PMMS)?

A: Most data on average mortgage rates, including the PMMS data, present average interest rates for loans originating in a given period. In contrast, the “Fixed_Rate” field shows the average mortgage rate for all active loans, most of which originated in previous periods. The “New_Rate” field shows the predicted interest rate on these loans if they were to be re-originated in the period. The “New_Rate” field will closely track average interest rates but differ slightly as existing loans can have different credit characteristics than those originated in a period. These fields contain averages for all fixed-rate mortgages, so they may also differ from the PMMS data, which considers only applications for 30-year loans submitted to Freddie Mac.

Q12: Why are there some very large/small values in some fields?

A: In some cases, the estimates are performed on a very granular level, so there will be cases where statistical noise is high. However, if something seems incorrect, please let us know! We view these measures as a public good. These estimates are developmental, and we will revise the dataset, as needed, based on user feedback.

Q13: How often do you plan to produce these data, and are revisions expected?

A: These data were produced to accompany the working paper and contain estimates for 1998 through 2023. We may continue to update or revise the dataset, as needed, based on user feedback.

Q14: Are data additive? For example, can metrics for lower levels of geography aggregate upwards?

A: For the most part, yes. However, please note that the estimates of lock-in sensitivity and the effect on sales are not calculated for all MSAs and counties, so not all geographic aggregation is possible for these metrics. Please see the tab descriptions in Q6 for more information.

Q15: Information about my city is not included. Can you provide it?

A: Unfortunately, data limitations (e.g., too few mortgages) typically prevent us from providing statistics for specific areas or years. We only release information when it has been aggregated to an appropriate level. If you believe your location has been overlooked, you are welcome to let us know, and we can investigate further.

Q16: Can I obtain the underlying data used to create these statistics?

A: For privacy reasons, this is not possible. However, the FHFA produces aggregate statistics for the NMDDB that are available to the public. More information is available at <https://www.fhfa.gov/DataTools/Downloads/Pages/National-Mortgage-Database-Aggregate-Data.aspx>.