



Loan Performance Data Tutorial

Getting Started with Fannie Mae's Single-Family Dataset

February 2021

Disclaimer

The information provided in this presentation is intended to provide an introduction to the analysis and understanding of single-family mortgage loan performance data published by Fannie Mae. The tools and methods presented in the tutorials are not intended to provide comprehensive instruction as to the use and analysis of the data, and may not reveal trends in performance or other information that may be meaningful to particular users of the data. Data users, together with their financial and other advisors, must determine for themselves the most appropriate methods to use to analyze the data and should ensure they are comfortable with the sufficiency of such analysis before using the data to assist in making investment decisions. Fannie Mae shall have no liability for any errors or misunderstandings resulting from misapplication of the information presented in the tutorials. Data users should also note that all loan performance data referenced in the tutorials is historical performance data and may not be predictive of future loan performance.



Table of Contents

Background	4
Available Resources	9
Creating a Loan-level Performance Dataset	11
Home Affordable Refinance Program® (HARP®) Data	20





Background

History

- In support of its credit risk transfer programs, Fannie Mae released an extensive dataset beginning in 2013 that provides insight into the credit performance of a portion of Fannie Mae's single-family book of business.
- The dataset provides monthly loan-level detail and is offered to help investors gain a better understanding of the credit performance of a portion of single-family loans owned or guaranteed by Fannie Mae.
- The public dataset includes a subset of Fannie Mae's 30-year, fixed-rate, fully documented, single-family amortizing loans that the company owned or guaranteed on or after January 1, 2000.
- In October 2020, Fannie Mae enhanced the data offering to include additional credit performance data.



General characteristics of the data

Data is a subset of Fannie Mae's single-family, conventional mortgages and includes the following attributes only:

- 30-year fixed-rate
- Fully amortizing
- Fully documented
- Acquired by Fannie Mae on or after January 1, 2000

Data does not include:

- Home Affordable Refinance Program (HARP) mortgage loans
- Refi Plus™ mortgage loans
- Mortgage loans with LTVs greater than 97%
- Alt-A or other mortgage loans with reduced documentation and/or streamlined processing
- Loans sold to Fannie Mae with lender recourse or subject to certain other third-party, risk-sharing arrangements (other than primary mortgage insurance)
- Loans acquired under certain programs or negotiated variances that are no longer eligible for delivery to Fannie Mae
- Adjustable-Rate mortgage loans
- Balloon mortgage loans
- Interest-only mortgage loans
- Mortgage loans with prepayment penalties
- Government insured mortgage loans
- Loans sold to Fannie Mae on a negotiated bulk basis



Changes in the October 2020 release

- The October 2020 Single-Family Loan Performance data update included acquisition data and performance data through Q2 2020.
- The acquisition data is no longer be published with a 13-month lag, and instead will be on a four-month lag.
- In addition, Fannie Mae has enhanced its dataset to a single-file format with a new file layout of 108 fields, which will mirror our Single-Family Credit Risk Transfer monthly remittance disclosure files.
- Please refer to the next slide for a full list of new data fields included in the October 2020 release.
- Users may also reference additional resources including:
 - [File Format Comparison](#)
 - [Glossary and File Layout](#)
 - [Single-Family Loan Performance Data Frequently Asked Questions \(FAQs\)](#)
 - [Sample File](#)



New data fields

Field Name	Description
<i>Prepayment Penalty Indicator</i>	An indicator that denotes whether the borrower is subject to a penalty for early payment of principal.
<i>Interest Only Loan Indicator</i>	An indicator that denotes whether the loan only requires interest payments for a specified period of time beginning with the first payment date.
<i>Interest Only First Principal and Interest Payment Date</i>	For interest-only loans, the month and year that the first monthly scheduled fully amortizing principal and interest payment is due.
<i>Months to Amortization</i>	For interest-only loans, the number of months from the current month to the first scheduled principal and interest payment date.
<i>Loan Payment History</i>	The coded string of values that describes the payment performance of the loan over the most recent 24 months. The most recent month is located to the right.
<i>UPB at the Time of Removal</i>	The unpaid principal balance of the loan at the time of removal. Populated as a loan hits a Zero Balance Code or is otherwise liquidated.
<i>Total Principal Current</i>	The change between the prior reporting period's disclosed Current Actual UPB and the current reporting period's disclosed Current Actual UPB.
<i>HomeReady® Program Indicator</i>	An indicator that denotes if the borrower participated in Fannie Mae's HomeReady program. HomeReady is our affordable, low down payment mortgage product designed to expand the availability of mortgage financing to creditworthy low-to-moderate-income borrowers.
<i>Property Valuation Method</i>	The method by which the value of subject property was obtained.
<i>High Balance Loan Indicator</i>	An indicator that denotes if the original principal balance of a mortgage loan is greater than the general conforming loan limit and up to the high-cost area loan limit.
<i>Borrower Assistance Plan</i>	The type of assistance plan that the borrower is enrolled in that provides temporary mortgage relief or an opportunity to cure a mortgage delinquency over a defined period.
<i>High Loan to Value (HLTV) Refinance Option Indicator</i>	An indicator that denotes if an eligible original reference loan is refinanced under Fannie Mae's HLTV refinance option, which results in such mortgage loan remaining in the Reference Pool, as further defined in each individual CRT document, if applicable.
<i>Alternative Delinquency Resolution</i>	An indicator that denotes that the borrower has been granted a specified plan that provides for an alternative form of mortgage payment relief which allows the borrower to cure a mortgage delinquency by deferring all or a certain portion of past due amounts on the mortgage loan until the maturity date of the loan.
<i>Alternative Delinquency Resolution Count</i>	The total number of Alternative Delinquency Resolutions as reported by the servicer for a specific loan.
<i>Total Deferral Amount</i>	The total amount related to one or more Alternative Delinquency Resolutions which will not accrue interest.





Available Resources

Glossary, FAQs and code resources

- [File Glossary and Layout](#)
- [FAQs](#)
- [Code \(Primary\)](#)
- [Code \(HARP\)](#)
- [Statistical Summary \(Primary\)](#)
- [Statistical Summary \(HARP\)](#)





Creating a loan-level performance dataset

How to download the data

- The data is grouped according to Fannie Mae's acquisition year. For each available period there is one file containing static acquisition characteristics and dynamic performance activity.
- Files are updated quarterly – download all files every quarter.
- Data for individual acquisition quarters may be downloaded, or users can download all files at once by utilizing the link at the top of the data portal.

The screenshot displays the 'Data Dynamics' interface for downloading 'Single-Family Loan Performance Data'. The left sidebar contains navigation options: HP, REPORTS, MESSAGES (33), DOWNLOAD DATA, and RESOURCES. The main content area is titled 'Download Single-Family Loan Performance Data' and is divided into two sections. The top section, 'Full datasets', includes a 'Primary Dataset' with a link for '2000Q1-2020Q2 Acquisition and Performance File' and a 'HARP Dataset' with a link for 'HARP Acquisitions, Performance and Mapping Files'. The bottom section, 'Individual acquisition quarters', is titled 'Quarterly Single-Family Loan Performance (Primary) Dataset' and features a table with columns for Year, Q1 Records, Q2 Records, Q3 Records, and Q4 Records. Each cell in the table contains a link labeled 'Acquisition and Performance'.

Year	Q1 Records	Q2 Records	Q3 Records	Q4 Records
2000	Acquisition and Performance	Acquisition and Performance	Acquisition and Performance	Acquisition and Performance
2001	Acquisition and Performance	Acquisition and Performance	Acquisition and Performance	Acquisition and Performance
2002	Acquisition and Performance	Acquisition and Performance	Acquisition and Performance	Acquisition and Performance



From files to dataset

Creating a dataset that's ready for analysis

- To aid analysis, we provide R code to guide the creation of an analytical dataset that condenses key information into one row per loan.
- This dataset combines static loan and borrower characteristics with monthly and “last status” variables pulled from performance data, including:
 - Flags for intermediate performance milestones (ex. date of first 180-day delinquency)
 - Values from the most recent performance reporting period (ex. zero-balance code or current status)
- Please reference the R codes in conjunction with this tutorial.



Reading in the data

- ‘LPPUB_Infile.R’ provides a basic method for importing Fannie Mae’s Loan Performance data into an R session.
- For this tutorial we will utilize the ‘LPPUB_StatFile.R’ and ‘LPPUB_StatFile_Production.R’ codes, building upon the basic data import method and transforming the data into our analytical dataset.
- ‘LPPUB_StatFile.R’ will call the ‘Production’ code to sequentially process the Loan Performance files into our preferred format. You will not need to edit the Production code, but **please ensure ‘LPPUB_StatFile.R’ is configured correctly.**
- For this tutorial we will be analyzing data from Q1 2000 through Q2 2020, which is the default configuration of the code.



Dataset calculations

Exact calculations for each of these variables are included in the associated R code

- Borrower minimum credit score
- Original Home Value

Derived from acquisition characteristics

- Last Status
- Last Date
- Credit Event Date
- Credit Event UPB
- First 180 Date
- First 180 UPB
- Foregone Interest Cost
- Net Loss
- Net Severity

Derived from performance data



Calculated field definitions

Field Name	Description
Borrower minimum credit score	If there is both a borrower credit score and co-borrower credit score on a mortgage loan, the minimum credit score is determined by taking the lowest of the two credit scores in order to create one column for analysis
Original Home Value	Original Loan-to-Value Ratio (OLTV) is the proportion of the original loan amount relative to the price of the property <ul style="list-style-type: none"> • Purchase mortgages: Original home value = minimum of purchase price or appraisal • Refinance mortgages: Original home value = appraisal • Purchase price = Original Loan Amount / (OLTV/100)
Last Status	For loans in a “continuous” status (i.e., non-liquidated) the last date will be the latest monthly reporting period.
Last Date	For loans in a “terminal” status (i.e., liquidated) the last date will be the date on which the terminal status occurred
Credit Event Date	The date that the credit event occurred
Credit Event UPB	The outstanding unpaid principal balance corresponding to the credit event date
First 180 Date	The date that the first 180-day delinquency occurred.
First 180 UPB	The outstanding unpaid principal balance corresponding to the First 180 Date
Foregone Interest Cost	This represents the foregone interest Fannie Mae would otherwise earn on a performing loan
Net Loss	This represents the loss to Fannie Mae, including delinquent interest, net of any proceeds. It is represented in dollars
Net Severity	This represents the net total loss to Fannie Mae, as a percentage of defaulted UPB. It is usually represented as a percentage and is sometimes referred to as a “loss given default” statistic



Creating the statistical summary

- After we have read in and processed the Loan Performance data, we turn to ‘LPPUB_StatSummary.R’ to help analyze the data.
- This code will generate a statistical summary identical to the summary provided in the Loan Performance data portal. Replicating the summary is a good way to ensure the data has been properly and completely downloaded.
- The final output of the code is an Excel file that summarizes several key metrics of the Loan Performance data, grouped by acquisition year of the loan.



Running the code

A step-by-step guide

1. Download and unzip the most recently available Single-Family Loan Performance data. Ensure the files are all located in the same directory and accessible by your R application.

Note: The full Loan Performance dataset is very large. Consider working with subsets of the data as necessary if you experience memory constraints while using R.

2. Download the R codes, available from a link in the Loan Performance data portal, and copy them into the same directory as the data. You will need 'LPPUB_StatFile.R', 'LPPUB_StatFile_Production.R', and 'LPPUB_StatSummary.R' to recreate the statistical summary.
3. Ensure the *starting_file* and *ending_file* parameters within 'LPPUB_StatFile.R' are correctly defined. The default values for the October 2020 release are 0 and 81, respectively.
4. Start your R session and set the working directory to the folder containing the data and code. Run `source('LPPUB_StatFile.R')` to begin processing the data.



Running the code (continued)

A step-by-step guide

5. Ensure that 'LPPUB_StatFile.R' completes without error. There will be one ".csv" file for each quarter of the dataset in the working directory.
6. Run `source('LPPUB_StatSummary.R')` to create the statistical summary. This will output an Excel workbook to the working directory.
7. Inspect the statistical summary and confirm it is identical to what is provided in the Loan Performance data portal. If the statistical summary is not identical, check whether the Loan Performance data was downloaded in full and that each R code ran without error.
8. If you encounter any difficulty while working with the dataset or code, please utilize the contact information available at the end of this tutorial.





HARP Data

What is included in the HARP data?

- The **HARP** dataset contains approximately one million 30-year fixed rate mortgage loans that are in the primary dataset that were acquired by Fannie Mae from January 1, 2000 through September 30, 2015 and then subsequently refinanced into a fixed rate mortgage loan through HARP from April 1, 2009 through September 30, 2016.
- The dataset includes a mapping key to assist users in aligning the loan ID in the primary dataset with the corresponding loan ID in the HARP dataset after the loan was refinanced.
- Loan performance characteristics derived from the HARP dataset reflect events that occurred at any point during the loan's life-span, before and after HARP refinancing.



How to utilize and download

Download Single-Family Loan Performance Data

Primary Dataset

HARP Dataset ← HARP dataset

2000Q1-2020Q2 Acquisition and Performance File

HARP Acquisitions, Performance and Mapping Files

Quarterly Single-Family Loan Performance (Primary) Dataset

Year	Q1 Records	Q2 Records	Q3 Records	Q4 Records
2000	Acquisition and Performance	Acquisition and Performance	Acquisition and Performance	Acquisition and Performance
2001	Acquisition and Performance	Acquisition and Performance	Acquisition and Performance	Acquisition and Performance
2002	Acquisition and Performance	Acquisition and Performance	Acquisition and Performance	Acquisition and Performance

- Users must download the HARP Dataset files marked above *in addition* to the regular dataset in order to reproduce our statistical summary including HARP.
- The download includes two files: (1) the HARP acquisition and performance data and (2) the loan mapping file, which matches loan IDs from before and after HARP refinancing.



Creating the HARP stat summary

- In addition to our regular R code, we also provide code specific to processing and summarizing the HARP data. This code is denoted by the suffix “_HARP”.
- This code follows the same steps as the regular code. Users can edit and run ‘LPPUB_StatFile_HARP.R’ to produce the dataset, then run ‘LPPUB_StatSummary_HARP.R’ to create the HARP statistical summary.
- For speed and convenience, the HARP code re-uses the files created during the regular statistical summary process. **Please follow the steps necessary to create the regular statistical summary before proceeding with the HARP code.**



Contact us

If you encounter difficulty utilizing Fannie Mae's Single-Family Loan Performance dataset, please contact the Fixed Income Marketing department by [email](#).





Thank you

