**Mortgage Insurance Factor**

**A Proposal to Simplify Disallowances and Curtailments**

**July 8, 2016**

**EXECUTIVE SUMMARY**

While disallowances and curtailments are responsible for only a relatively small reduction in the overall mortgage insurance (MI) claim payments, they are present on the vast majority of all MI claim settlements. Disallowances and curtailments in the MI claims process create uncertainty with MI claim payments as well as significant processing and staffing burdens. They also extend settlement timelines due to the rebuttal process for curtailments with Fannie Mae’s servicers, which can take up to two or more years.

Fannie Mae is proposing to eliminate these issues by using an “MI Factor” algorithm to self-disallow and to self-curtail claims prior to their submission to the MI companies. This would eliminate the uncertainty, the processing and staffing burdens, and the extended timelines since the MI claim amount would already be inclusive of disallowance and curtailment adjustments. The MI Factor would be applied against certain loan-level information shortly after loan liquidation through a foreclosure sale or alternative to foreclosure.

To self-disallow, the proposed algorithm would leverage a standardized list of disallowed costs to remove all unclaimable expenses from the algorithm’s estimation population. To self-curtail, the proposal is to leverage Fannie Mae’s published foreclosure timelines, including allowable delays, to cap the actual timelines. This capping would apply to both the interest and foreclosure cost components of the MI claim amount calculation.

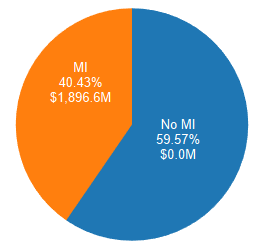
The proposed MI Factor algorithm estimates average foreclosure costs using Fannie Mae’s actual foreclosure cost data, adjusted for disallowances and curtailments. The algorithm uses key loan characteristics such as disposition type, geography, property value, and property type in its specification. This granularity provides the flexibility needed to capture changing market dynamics. The proposed algorithm performs well from a statistical perspective and should be easy to update going forward. Since the average costs will be calculated by key loan characteristics, a grid implementation, updated annually, is recommended.

**MORTGAGE INSURANCE OVERVIEW**

Background

Borrower-paid mortgage insurance is an important tool used by mortgage investors to mitigate the credit losses that occur when a borrow defaults on their mortgage.[[1]](#footnote-1) Borrowers are required to pay for mortgage insurance policies on the mortgage investor’s behalf on any conventional mortgage where the origination loan-to-value ratio (LTV) is greater than 80%. Should the borrower default on their mortgage, resulting in a credit loss for the mortgage investor, the MI company pays the investor a claim amount based on the MI policy that is in place for the mortgage. In 2015, 40% of all of Fannie Mae’s loan defaults were covered by MI policies allowing for $1.9 billion in borrower-paid mortgage insurance recoveries.

Figure 1: Mortgage Defaults by MI Coverage



MI Claim Amount & MI Expected Payment

When a mortgage defaults with an MI policy in place, the mortgage servicer files the claim with the MI company. The filed MI claim amount is equal to the sum of the unpaid principal balance (UPB), the accrued delinquent interest, and the foreclosure costs.

* UPB: The balance of the unpaid principal at the time of loan liquidation.
* Interest: The total accrued interest at the time of loan liquidation, which is the product of the UPB, the note rate on the mortgage divided by 365.25, and the number of days in delinquency.
* Foreclosure Costs: The sum total of all of the costs that the servicer incurred related to the mortgage and its underlying collateral during the delinquent time period through loan liquidation. Some foreclosure costs may be one-time costs, such as eviction costs, which are classified as fixed costs. Other foreclosure costs may be periodic, such as property preservation items, which are classified as variable costs.

Since both the interest and foreclosure cost components accumulate while the mortgage is delinquent, they are both significantly impacted by the length of time the mortgage was in delinquency, also known as the foreclosure timeline.

* Foreclosure Timeline: The mortgage’s delinquency time period, defined as the number of days between the beginning of delinquency, known as the last paid installment (LPI) date, and the liquidation event.

The MI claim’s expected payment can be derived from the MI claim amount by multiplying it by the MI coverage percent.

* MI Coverage Percent: The percent of the MI claim amount, specified in the MI policy, that the MI company will cover in the case of mortgage default.

MI Settlement Types

There are three ways in which an MI claim may be settled: an option settlement, a presale with influence, or an MI acquisition.

* Option Settlement: The MI company pays the MI coverage percentage of the MI claim amount. In 2015, claims were settled as option settlements approximately 96% of the time.

* Presale with Influence: This settlement type takes into account the actual sale proceeds and sale costs for the foreclosed property. Rather than paying a specified percentage, the property’s sale proceeds are subtracted from the claim amount and the difference is paid. This settlement type also includes a supplemental bill for real estate owned (REO) costs. In 2015, claims were settled as presales with influence approximately 4% of the time.
* MI Acquisition: The MI company, rather than making a claim payment, chooses to make Fannie Mae whole on the loss and takes the foreclosed property into its inventory. In 2015, claims were settled as MI acquisitions less than 1% of the time.

Disallowances & Curtailments

A number of issues can occur that result in an actual MI payment that is less than expected. The MI Factor proposal concerns the simplification of two such issues: disallowances and curtailments[[2]](#footnote-2).

* Disallowances: If the servicer erroneously includes foreclosure costs in the MI claim amount that the MI policy does not cover, like the MI premium payments, the MI company will exclude these costs from the claim amount when calculating the actual MI payment. Disallowances result in a reduction in the overall MI claim amount, and thus the actual MI payment.
* Curtailments: When issues are found with the servicing of the mortgage that cause the overall MI claim amount to be unreasonably high, such as the servicer taking too long to foreclose, the MI company will curtail the MI claim amount to a reasonable level. This results in a reduction in the overall MI claim amount, and thus the actual MI payment.

In 2015, 72% of all MI claims were subject to disallowance or curtailment, which translates to an average MI payment reduction of 7.7%.

Figure 2: MI Claims with Disallowances and Curtailments

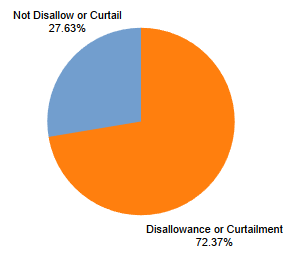
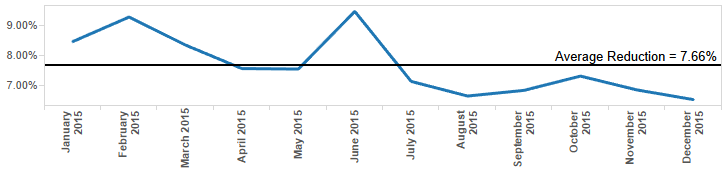


Figure 3: MI Benefit Payment Reductions due to Disallowances and Curtailments



Causes of Curtailments

Disallowances have a single cause, which is the inclusion of unclaimable costs in the MI claim amount. Curtailments, however, are caused by several different kinds of servicing issues: foreclosure timeline issues, the absence of loss mitigation, and cost threshold issues.

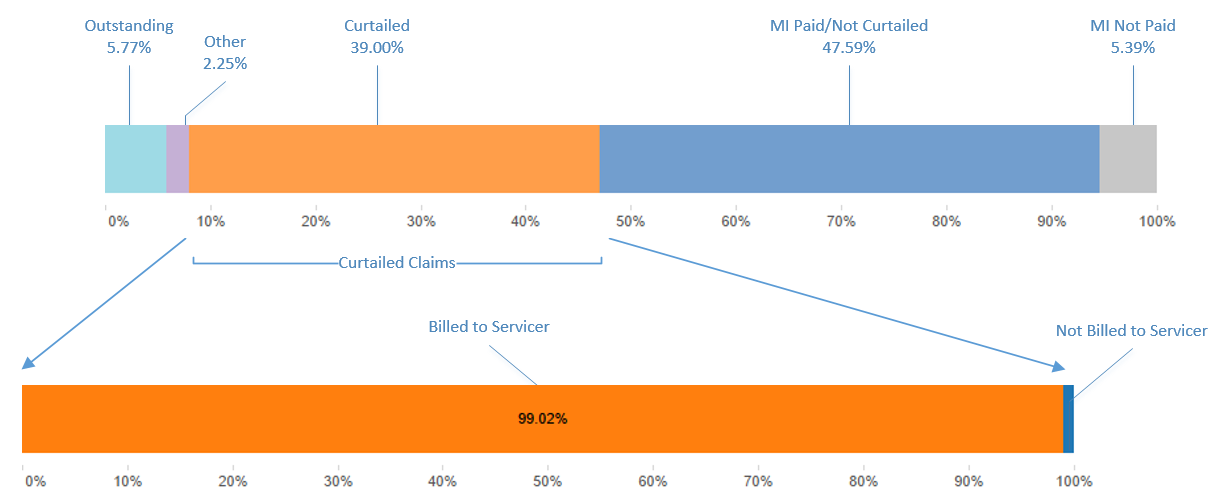
* Foreclosure Timeline Issues: If the foreclose timelines are unreasonably long, this results in upward pressure on both the interest as well as the foreclosure cost components of the MI claim amount equation. If the foreclosure timeline is deemed unreasonable, the MI company will curtail the timeline to a reasonable length of time, thereby reducing both the interest and foreclosure cost components of the total MI claim amount. Fannie Mae publishes guidance on allowable foreclosure timelines and allowable delays, e.g., bankruptcy, which many MI companies choose to follow; however, several MI companies curtailed historically based off of their own in-house timelines.
* Absence of Loss Mitigation: Loss mitigation activities, such as loan modifications, are designed to prevent mortgage defaults and to keep borrowers in their homes. If the mortgage default is prevented due to loss mitigation, no loss would result and the MI company would not be required to make an MI payment. Not all loss mitigation activities are successful, however. If a foreclosure occurs without any attempt at loss mitigation, the MI company will curtail the payment as the loss may have been avoided had loss mitigation activities occurred.
* Cost Threshold Issues: Cost threshold issues result in curtailments of the foreclosure cost component of the MI claim amount equation. When the amount of any given allowable cost is deemed to be unreasonably high, the MI company imposes a cost threshold that caps the expense at a reasonable level. Cost threshold issues are closely related to foreclosure timeline issues since extended timelines can result in a substantial increase in variable costs.

As the MI Factor algorithm results will show, Fannie Mae’s research indicates that foreclosure timeline issues explain the vast majority of all curtailments.

Servicer Billing

Since curtailments are associated with servicing issues, if the MI company curtails a claim, Fannie Mae will send a supplemental bill to the servicer for the total amount of the curtailment. The servicer frequently disagrees with the curtailment in part or in total and rebuts the curtailment with the MI company. The rebuttal process can take up to two or more years before all parties agree on what is owed and all claim payments are made. Practically all MI curtailments include a supplemental bill to the servicer; to date, only 22% of the total amount billed to servicers for 2015 MI claims has been paid.

Figure 4: Servicer Billings for Curtailments



MI Claims Process Issues

There are a number of issues in the MI claims process caused by disallowances and curtailments such as claim payment uncertainty, processing and staffing burdens, and extended timelines. Each are explained below.

* Claim Payment Uncertainty: Disallowances and curtailments create uncertainty in the MI claim process since the actual claim payment amount is not known until the payment occurs. This uncertainty is aggravated by a lack of standardization across the MI industry. Each MI company follows their own internal policies to determine their approach to curtailments, making actual claim payments difficult to predict industry-wide, although generally the there have been trends toward closer alignment in recent years.
* Processing and Staffing Burdens: MI companies review, decision, and pay each claim individually for disallowance and curtailment issues. As a result, each claim payment is also processed individually by Fannie Mae and its servicers. Processing every claim individually creates a significant staffing burden for all parties involved. This burden is aggravated by the additional staff needed to process curtailment rebuttals. In 2015, Fannie Mae’s outsourcing cost for managing the MI claims process, including rebuttals, was $1.6 million, not counting any ancillary expenses associated with other teams at Fannie Mae indirectly who exist with coordinating servicer conversations .
* Extended Timelines: Curtailment rebuttals cause significant delays in receiving the full MI claim payment. Rebutting of curtailments involves explaining in detail any events associated with the foreclosure process which had adverse effects on timelines, which is document intensive. Settling rebuttals for final decisions on payments requires significant back and forth between the servicers, the MI companies, and Fannie Mae that can take up to two or more years.

Historical Approach

Historically, a different factor approach was in use to address curtailment issues. This approach, known as streamline factor, settled claims by using a UPB multiplier that accounted for interest and foreclosure costs. For example, if the UPB was $100,000 and the streamline factor was 116%, then the final claim amount would have been $116,000. The factor was updated annually based on a rolling two-year window of actual default data, and it was applied as a single multiplier in each year for each state.

There were a number of issues with the historical streamline factor approach that limited its ability to capture changing market dynamics. The housing collapse caused major changes in default volumes, foreclosure timelines, foreclosure procedures, and foreclosure costs. Additionally, with higher-value properties defaulting, there were changes in the types of loans that were defaulting. The streamline factor’s specification was unable to reflect these market changes, and almost all of the MI companies using this approach opted out during the crisis.

The historical specification for streamline factor was unable to capture market changes for three reasons: it lacked granularity by loan characteristics, it did not capture the extending foreclosure timelines or the MI companies’ foreclosure timeline policies, and it ignored preferred liquidation strategies.

* Lack of Granularity: The lack of granularity is the primary issue with the historical streamline factor. Since state was the only loan characteristic used, no changes in the default distribution by any other important loan characteristic were captured. This would often cause substantial variation between claims processed using actual data versus those using the factor.

* Foreclosure Timelines: As the housing collapse protracted, many states were unable to efficiently process the increased foreclosure volumes, resulting in significantly extended foreclosure timelines. Since the historical factor did not account for these extreme variations in the timelines, at the beginning of the crisis, it was not able to keep up with the market. As the housing crisis matured, the extended timelines began to be reflected in the estimation data and thus in the streamline factor. At this point, however, the actual foreclosure timelines extended beyond what was allowed by the MI companies’ policies, and there was no mechanism in place to cap the timelines per the MI policies.
* Preferred Liquidation Strategies: During the housing collapse, liquidation alternatives such as short sales, third party sales (TPS), and deeds in lieu of foreclosure (DIL) became standard options for Fannie Mae and its servicers to pursue; however, the historical streamline factor only considered the REO disposition type. These alternative liquidation strategies result in savings on foreclosure costs as compared to an REO outcome, and the historical factor did not capture these savings.

The specification for the proposed new MI Factor is designed to address these weaknesses in the historical streamline factor’s approach. Increased granularity by key loan characteristics provides the proposed MI Factor with the flexibility to capture changes in market dynamics, foreclosure timelines, and liquidation strategies.

**MI FACTOR ALGORITHM PROPOSAL**

Fannie Mae proposes to simplify the MI claims process by leveraging an algorithm to self-disallow and to self-curtail the MI claim amounts prior to filing the claims with the MI companies. Leveraging an algorithm streamlines the MI claims process by removing the need to determine disallowances and curtailments on a claim-by-claim basis. It also creates greater certainty in the MI claims process and greatly reduces the claims processing burden, thereby also reducing the staffing burden. Additionally, the claim payment timeline would be greatly reduced since Fannie Mae further proposes to leverage the new proposed process to eliminate servicer curtailment billings. Fannie Mae would continue to manage servicer performance at its discretion through the Compensatory Fee billing process rather than through both a Compensatory Fee and a curtailment billing, which would address servicer concerns about a duplicative penalty being associated with extended timelines.

As a note, this document only covers the proposal for option settlements. Any proposals to change ‘Presales with influence’ and ‘MI acquisitions’ are still being discussed for consideration.

Proposed MI Claim Amount Adjustments

The MI claim amount is calculated as the sum of the UPB, interest, and the foreclosure costs for any given loan default.

*MI Claim Amount = UPB + Interest + Foreclosure Costs*

In order to simplify disallowances and curtailments, the following adjustments to the MI claim amount equation are proposed.

* UPB: No changes to the UPB component of the MI claim amount equation are proposed since UPB is not impacted by disallowances or curtailments.
* Interest: Curtailments of the interest component of the MI claim amount calculation are driven by foreclosure timeline issues. The recommendation is to leverage Fannie Mae’s published foreclosure timelines, including allowable delays, to cap the actual foreclosure timelines. Currently, interest is calculated as follows.

*Interest = UPB \* (Note Rate / 365.25) \* Actual Days*

The proposed new interest rate calculation is as follows.

*Interest = UPB \* (Note Rate / 365.25) \* minimum (Actual Days, Allowable Days)*

* Foreclosure Costs: The proposal for standardizing foreclosure costs is to leverage a gross up factor, or MI Factor, to estimate average fixed and variable costs by key loan characteristics, similar to the prior method. These estimated costs would replace the actual foreclosure costs in the MI claim amount calculation. Fannie Mae’s actual foreclosure cost data will be used as the estimation data, and the data will be adjusted to account for disallowances and curtailments. First, to account for disallowances, a standardized list of allowed and disallowed costs will be used to remove all disallowed costs from the data. Second, to account for curtailments, Fannie Mae’s published foreclosure timelines, including allowable delays, will be used to cap the variable costs. Since the average costs will be calculated by key loan characteristics, a grid implementation is recommended. The proposed new foreclosure cost calculation is as follows.

*Foreclosure Costs = Fixed Costs + (Variable Costs \* minimum (Actual Days, Allowable Days)*

The fixed and variable costs would be calculated based on a percentage basis of UPB, with controls for important factors for each claim.

Foreclosure Timeline Methodology

A joint alignment effort among Fannie Mae, Freddie Mac, and their regulator, the Federal Housing Finance Agency (FHFA) sets guidance on reasonable state-level foreclosure timelines. Alignment is important because these foreclosure timelines are used in Fannie Mae’s compensatory fee process, which is part of Fannie Mae’s larger servicer performance management strategy. Updates to the published foreclosure timelines have historically been reviewed on an annual or semi-annual basis. In the most recent update in early 2016, the aligned parties agreed to use the median (50th percentile) of observed foreclosure timelines as the new guidance.

There are events, such as bankruptcy, that cause a loan’s actual foreclosure timeline to reasonably extend beyond the published guidance, ‘Allowable Delays’. As such, the published foreclosure timelines also includes guidance on allowable delays, which are included in the MI Factor algorithm proposal. Certain of these allowable delays are associated with borrower events (bankruptcy filings of various types, contesting of the foreclosure, or probate), while other are statutory in nature (military indulgences). There are also others that control for certain events associated with the loan such as work-out attempts that are to the mutual benefit of Fannie Mae and MI companies if successful.[[3]](#footnote-3) The algorithm will not, however, include any one-off delays that the servicer often is able to substantiate during manual, case-level review in rebuttal of Compensatory Fee billings.

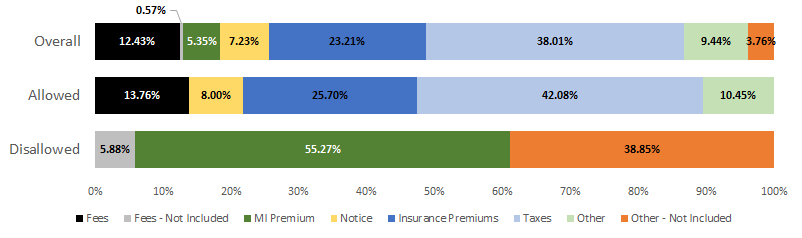
It is important to note that during the compensatory fee process, foreclosure timeline overages are offset by timelines that come in below guidance, and servicers are only billed for the net timeline overages in a given state for a given billing cycle. The MI Factor proposal does not include such state-level performance-based netting. Additionally, as Fannie Mae and servicers would through sampling approaches to arrive at rebuttal adjustments to Compensatory Fees, we are also not proposing to include these adjustments to the calculations of allowable days for purposes of the factor.

As the foreclosure process continues to evolve and different states return to normal, new foreclosure timeline methods may be recommended to the FHFA-led working group. The current published timelines and allowable delays are posted on Fannie Mae’s website [here](https://www.fanniemae.com/content/guide_exhibit/foreclosure-timeframes-compensatory-fees-allowable-delays.pdf).[[4]](#footnote-4)

Standardized Disallowances

At present, different MI companies use their own internal policies to determine allowed and disallowed foreclosure costs. The largest expense groupings that compose the allowed costs are insurance premiums, taxes, and fees, while disallowed costs consist of MI premiums and some specific uncollectable fees and other expenses. In order to apply an algorithmic approach to disallowances, a standardized list of allowed and disallowed costs is needed. A table listing the proposed allowed and disallowed foreclosure costs by expense grouping can be found in Appendix A.

Figure 5: Allowed and Disallowed Foreclosure Costs Distributions by Expense Grouping



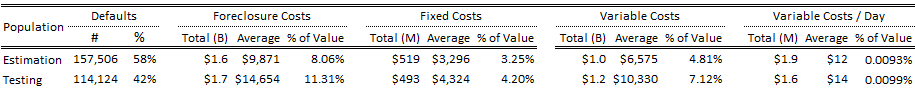
Fixed & Variable Foreclosure Costs

The allowed foreclosure costs are then segregated and analyzed as fixed and variable costs.

* Fixed Costs: Costs that generally occur only once during the foreclosure timeline are considered to be non-timeline varying.
* Variable Costs: Periodic costs occurring multiple times during the foreclosure timeline are considered functions of the foreclosure timeline.

It is important to analyze fixed and variable costs separately since the total amount of the variable costs is directly related to the length of the foreclosure timeline, making variable costs subject to foreclosure timeline-related curtailments. A table listing the allowable foreclosure costs that are considered to be fixed and variable can be found in Appendix B. While there are notable differences in absolute magnitude of costs between the estimation and testing samples in Table 1, the variable costs are actually similar when expressed on a *per diem* basis.

Table 1: Fixed and Variable Foreclosure Costs



The largest expense groupings that compose the fixed costs are various fees and notification costs, while variable costs consist predominately of hazard or flood insurance premiums and taxes.

Figure 6: Fixed and Variable Foreclosure Costs Distributions by Expense Grouping

MI Factor Estimation Data & Adjustments

The estimation data for the MI Factor algorithm proof of concept contains all of the actual foreclosure costs incurred by the servicer on all defaults or dispositions from July 2013 through June 2014. For short sales and TPS, the population includes all loan liquidations that occurred during the estimation time period. For REO, the population includes all property dispositions during the time period. Therefore, the base date used to determine the population is liquidation date for short sales and TPS and disposition date for REO. REO dispositions were used instead of liquidations to allow time for trailing pre-liquidation expenses to be submitted to Fannie Mae.

As default inventories are declining across the country, in order to ensure a robust population size going forward, all loan liquidations and property dispositions were used regardless of whether the loan was covered by MI. While the figures below indicate differences at a summary level in the average cost factors between loans with and without MI, this difference is diminished substantially once control factors are taken into account.

Figure 7: Foreclosure Costs by MI Coverage

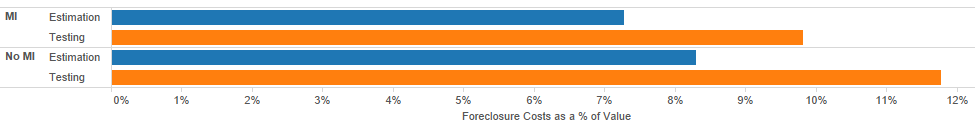
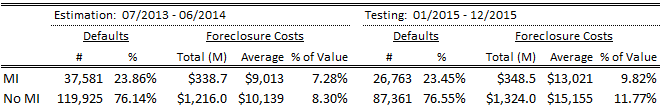


Table 2: Foreclosure Costs by MI Coverage



Adjustments to account for disallowances and curtailments were made as follows.

* Disallowance Adjustment: All costs on the standardized disallowed list were removed from the estimation data.
* Curtailment Adjustment: All variable costs incurred after the allowed foreclosure timelines, including allowable delays, were removed from the estimation data.

The estimation data was used to predict foreclosure costs from January through December 2015. The six-month lag in the data from July 2014 – December 2015 is in place to mimic the MI Factor implementation. When updating the algorithm going forward, a six-month data lag will be needed to allow time for all foreclosure costs to be submitted to Fannie Mae and for distribution of the proposed updated grid to MI companies for evaluation prior to implementation.

MI Factor: Key Loan Characteristics

The loan characteristics that were found to best explain foreclosure costs are disposition type, geography, property value, property type, and an urban/rural indicator. The specifications for each loan characteristic is detailed below with alternative specifications provided for geography, property value, and the urban/rural indicator.

* Disposition Type: Disposition type has a significant impact on foreclosure costs since there are costs that only occur in relation to the foreclosure event. Short sales, which dispose without going through a full foreclosure, do not have a full share of foreclosure costs, whereas TPS and REO, which do go through full foreclosure, do. Due to this, TPS and REO are grouped separately from short sales in the MI Factor algorithm. Although DIL generally has a cost structure that is more similar to short sales, for practical reasons of low frequency and information availability, DIL is grouped with TPS and REO. The property value specification for short sales, which is covered later, requires a retail sales price for short sale dispositions. Since DIL liquidations dispose as REO, a sales price is not available when the MI claim is filed. DIL makes up a small percent of all default liquidations so they cannot be grouped on their own in the longer term without introducing volatility. Due to their small size relative to REO, grouping them with REO and TPS has a minimal impact on the overall performance of the MI Factor algorithm. If in the future DIL volumes become a larger share relative to REO or TPS, this bucketing may be revisited.

Figure 8: Foreclosure Costs by Disposition Type

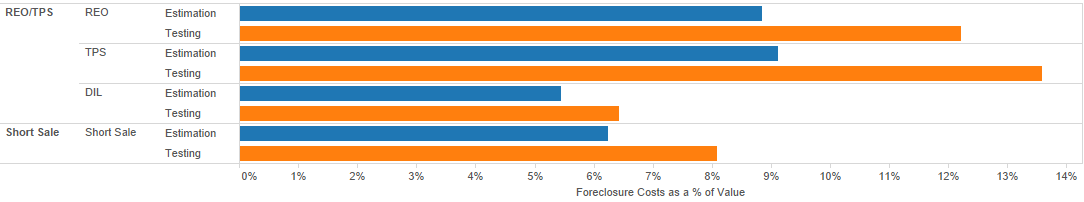
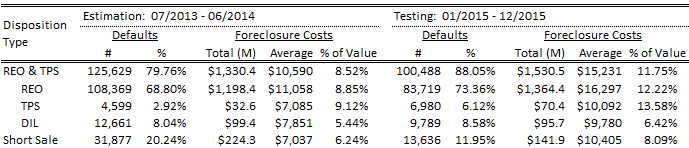


Table 3: Foreclosure Costs by Disposition Type

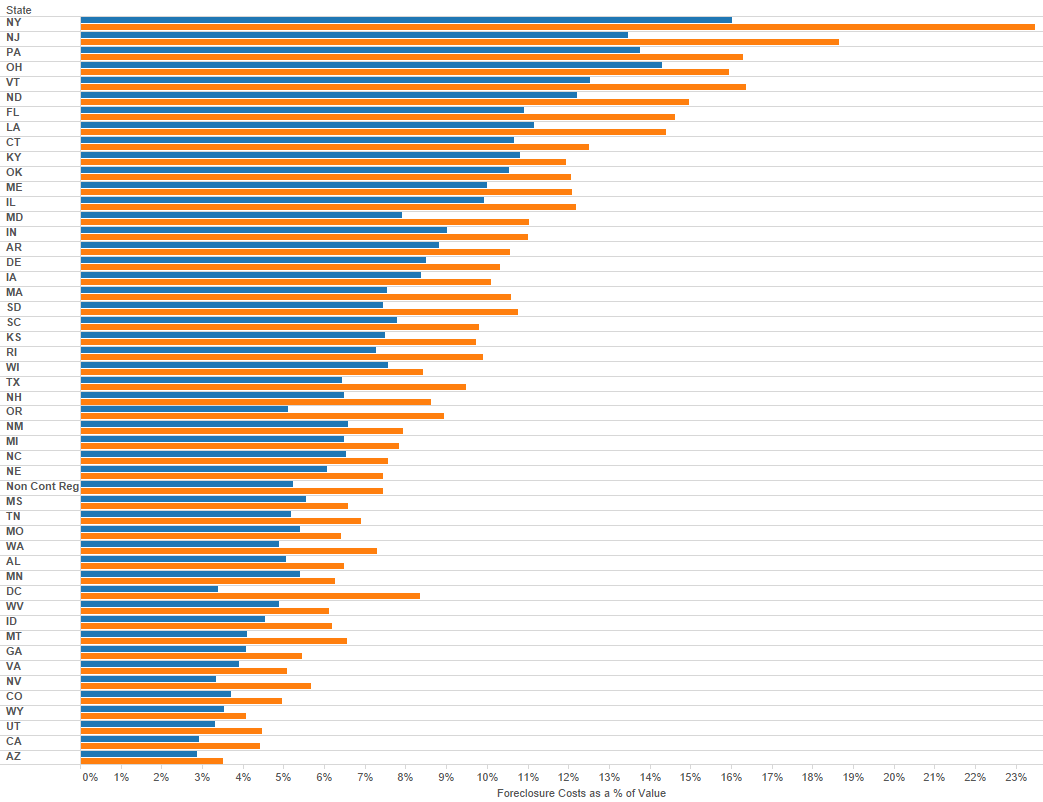


* Geography: Differences that vary by geography, such as costs of living, judicial processes, etc., also have a significant impact on foreclosure costs. To capture these geographical differences, two specification alternatives were researched.

Geography Alternative 1: 50 States

The first specification alternative for geography uses 50 state-level buckets. The 50 buckets are comprised of the 48 continental states; Washington, DC; and a consolidated grouping of the non-continental states and territories, which were grouped in order to ensure that sufficient default volumes are available for estimation. Due to the size, the full “Foreclosure Costs by 50 States” table can be found in Appendix C.

Figure 9: Foreclosure Costs by 50 States



Geography Alternative 2: State Clusters

State clusters were an alternative approach researched rather than using the full 50-state grouping for geography. A clustering methodology groups states with similar patterns together, which provides a more streamlined structure for the overall framework. This has an obvious implementation benefit and also provides the algorithm with greater longevity as default volumes continue to decline.

A *k*-means cluster algorithm was used to group like states. The *k*-means algorithm was chosen for two reasons: 1) it allows the researcher to specify the desired number of clusters, and 2) it also allows the researcher to remain neutral in the assignment the clusters. In order to assign the clusters, the *k*-means cluster algorithm first requires the *k*, or number of output clusters, to be defined. Several possible groupings were tested, but repeated testing resulted in eight clusters as the optimal recommendation. With *k* defined as eight, the algorithm associates the states individually based on the fixed and variable foreclosure costs mean reference points in aggregate for each state, eight at a time. With each iteration, the next wave of eight state are associated with the prior wave of eight clusters based on distance to the individual mean. The mean position, or centroid, of the eight points of the iterative cluster becomes the new mean for association in the next wave of eight associations. This process repeats until convergence has been reached and the end result is eight different clusters of the greatest possible distinction based on the fixed and variable foreclosure costs.

Figure 10: State Cluster Analysis

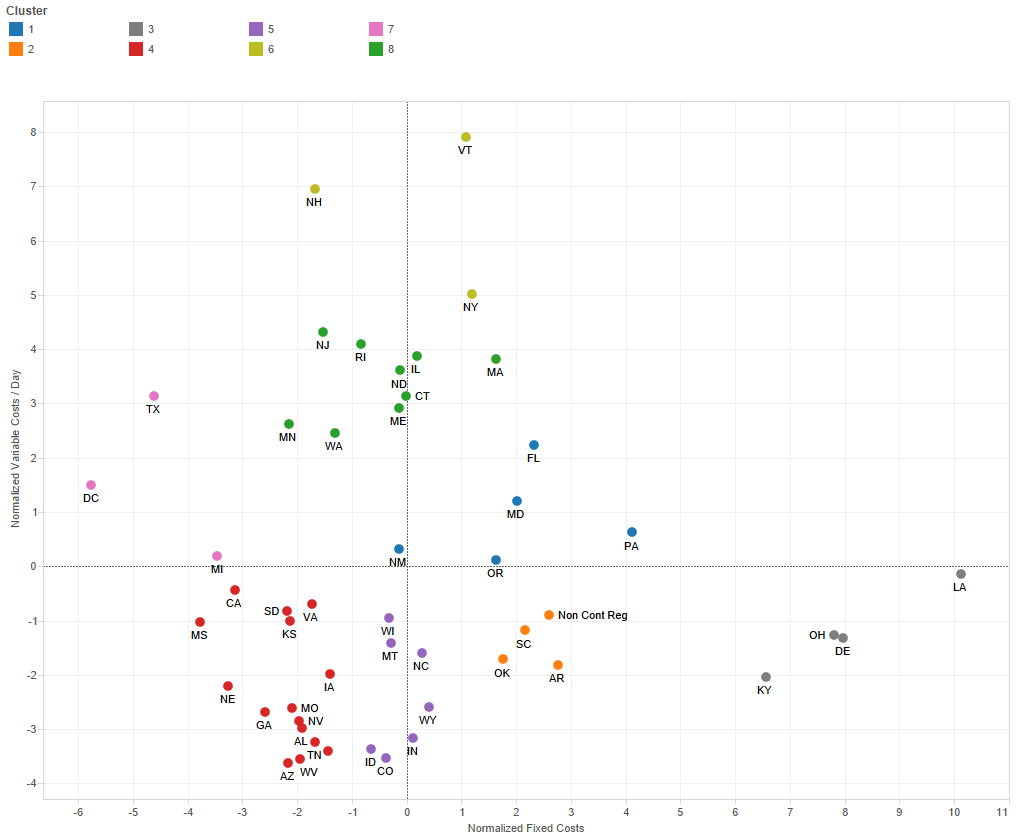


Table 4: Proposed Geography Clusters

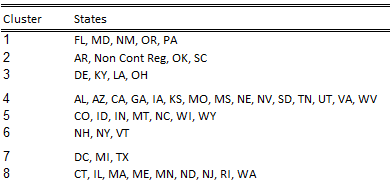


Figure 11: Foreclosure Costs by Geography Clusters

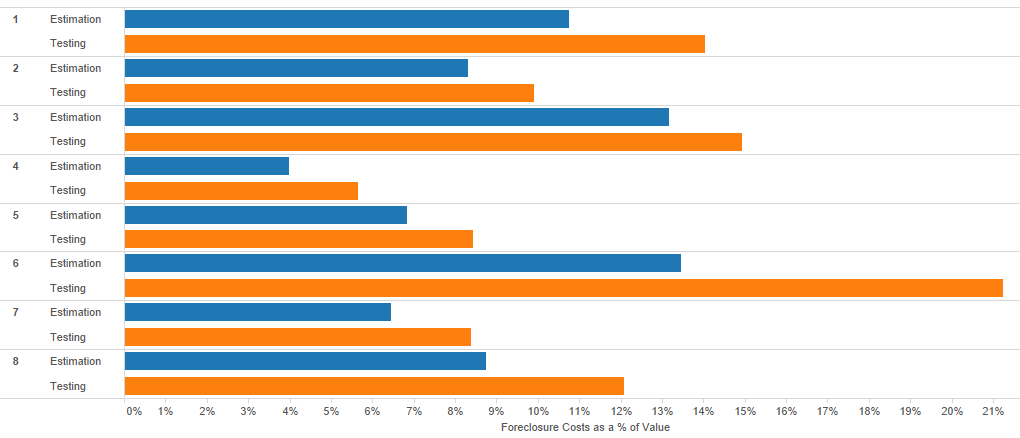
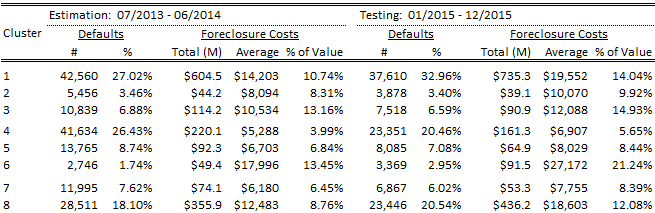


Table 5: Foreclosure Costs by Geography Clusters



The first 50-state alternative has the advantage of being easy to understand; however, due to declining default inventories, there may soon be insufficient data in many of the state buckets to continue to use this specification in the future. Should this become the case, updating the algorithm becomes problematic as judgment will be needed to determine future geographic groupings. The second geographic clustering alternative is perhaps less easy to understand; however, it provides a clear methodology for future algorithm updates as the *k*-means cluster algorithm is entirely data-driven. This removes judgment from the algorithm update process, which allows for the initial algorithm assumptions to be maintained consistently going forward. Also, in the future, as default volume declines, we would quickly run into situations where default volume within states would result in low cell frequency, especially once other control factors are considered. The cluster algorithm would help to avoid this problem. To the extent that future updates result in a different optimal definition of *k* other than eight, research for the proposed change would be provided in the annual update process.

* Property Value: Many foreclosure costs are influenced by value, the most obvious and largest of which are property taxes, which are a function of assessed value that is closely correlated with market value. Property values are determined in the open market by the sale of the property. At the time that the MI claim is filed, however, only short sales generally have a retail sales price. This presents an implementation issue for TPS and REO dispositions. Because of this, a hybrid approach was chosen to specify value. Since a retail sales price is available for short sales, the algorithm uses the short sale price as the value for short sale dispositions. TPS and REO dispositions use UPB as a value proxy since a retail sales price is not available.

UPB as a value proxy does have its short comings, as while its positively correlated with property value, there could be significant differences. Origination LTVs, changing home prices, or principle paydowns may cause default UPBs to be significantly greater or significantly less than the actual value of the property. While it is possible to use a home price index to mark the origination value to the liquidation date, this would significantly increase the complexity of the MI Factor implementation without providing any significant improvement in algorithm performance. Using broker price opinions (BPOs) or appraisals is also an option; however, both are subject to judgment and therefore also subject to dispute. As simplifying the MI claims process by reducing rebuttals is one of the goals of the MI Factor proposal, BPOs and appraisals are not the best solution for the goal. All things considered, using UPB as a value proxy for TPS and REO dispositions provides the most balanced solution. There are two specification alternatives researched for property value:

Property Value Alternative 1: $100K Buckets

The first alternative for specifying value is to simply bucket in increments of $100,000 in order to capture low-value, normal value, and high-value properties. Three buckets are specified: 1) $0 - $100K, 2) $100K - $200K, and 3) $200K and up.

Figure 12: Foreclosure Costs by Property Value - $100K Buckets

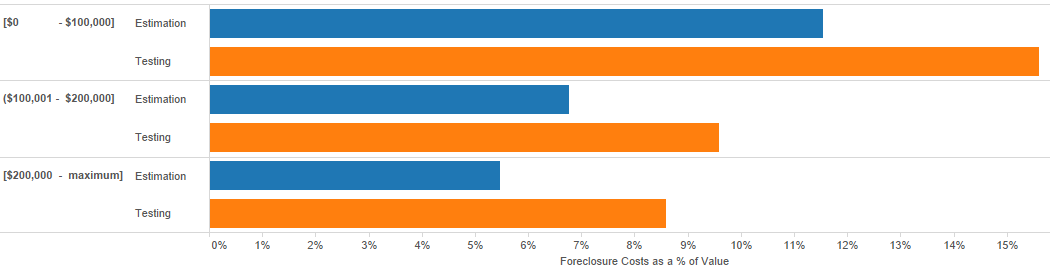
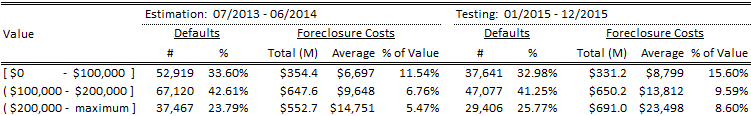


Table 6: Foreclosure Costs by Property Value - $100K Buckets



Property Value Alternative 2: Statistical Buckets

The second value alternative researched uses means and standard deviations to specify the value buckets, again with the intention of capturing low, normal, and high-value properties. Since value is a hybrid of sales price and UPB, mean and standard deviation are calculated separately for each disposition type bucket. In other words, the short sale mean and standard deviation will be calculated separately from TPS and REO. There are four value buckets specified for each disposition type bucket making eight value buckets in total. We also researched calculating the value means and standard deviations within a geography (both the 50-state and clustered state groupings), but performance differences of the algrothim were marginal.

* + Short Sales
    - 0 through Mean – One Standard Deviation
    - Mean – One Standard Deviation through Mean
    - Mean through Mean + One Standard Deviation
    - Mean + One Standard Deviation through Maximum
  + TPS & REO
    - 0 through Mean – One Standard Deviation
    - Mean – One Standard Deviation through Mean
    - Mean through Mean + One Standard Deviation
    - Mean + One Standard Deviation through Maximum

Figure 13: Foreclosure Costs by Property Value – Statistical Buckets

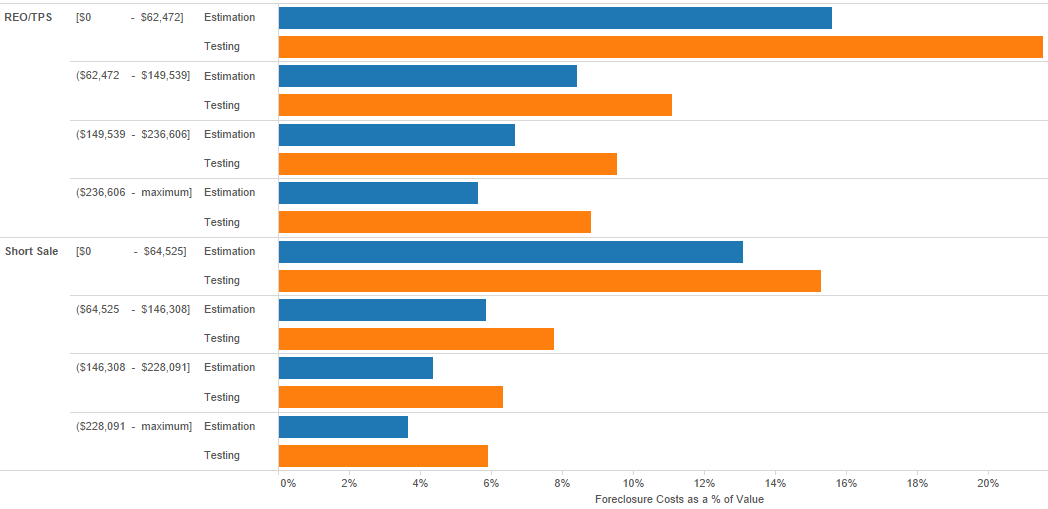
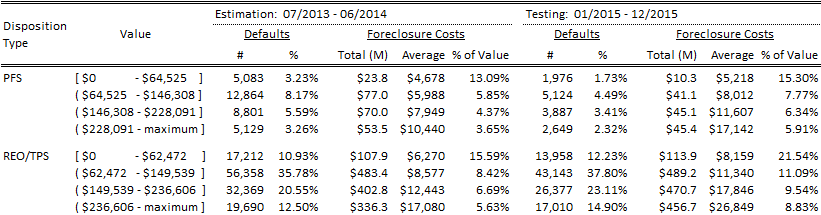


Table 7: Foreclosure Costs by Property Value – Statistical Buckets



The pros and cons of the two property value alternatives are similar to those of the geography alternatives. The $100K buckets are easy to understand, but require researcher judgment to update. The statistical specification is perhaps less easy to understand; however, it provides a clear methodology for future algorithm updates and is data-driven in its approach to maintain long-term dynamics.

* Property Type: Maintenance requirements, which impact the overall foreclosure costs, vary by property type. The proposed property type specification is grouped into three buckets. Eighty five percent of all defaults are one-unit properties, which make up the first bucket. Condos make up the second bucket since their multifamily-like structure has different maintenance requirements than do other single family properties. The third bucket is made up of all the other property types, including two to four unit properties as well as manufactured housing.

Figure 14: Foreclosure Costs by Property Type

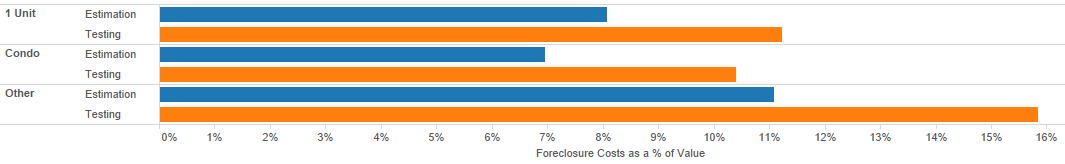
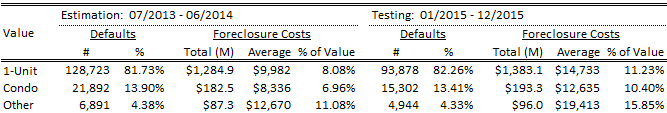


Table 8: Foreclosure Costs by Property Type



* Urban/Rural Indicator: It has already been established that the geographic location has an impact on foreclosure costs. However, within major metropolitan areas cost functions are markedly different than the more rural areas, even within a state. The urban/rural indicator is a further specification of geography. Metropolitan statistical areas (MSA) were used to determine urban locations. Any property not located in an MSA is considered rural.

The two alternatives for the urban/rural indicator are 1) using the indicator or 2) not using the indicator. Using the indicator allows the algorithm to further specify geography, but it increases the complexity of the implementation by increasing the grid size.

Figure 15: Foreclosure Costs by Urban/Rural Indicator

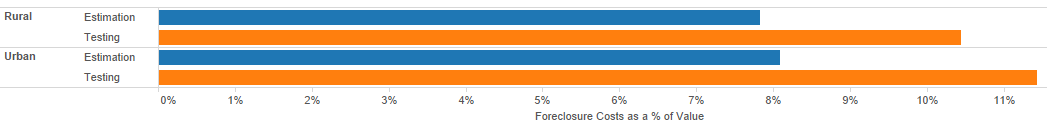
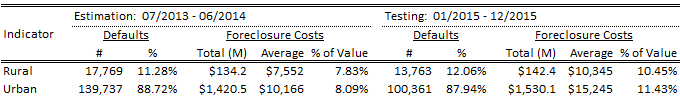


Table 9: Foreclosure Costs by Urban/Rural Indicator



MI Factor Outputs: Fixed and Variable Foreclosure Costs

The MI Factor algorithm has two outputs: the value-weighted average fixed foreclosure costs and the value-weighted average variable foreclosure cost per day. Both fixed and variable costs are specified as a percentage of value, which is a hybrid of sales price and UPB depending on disposition method. Specifying the costs as a percent of value has the advantage of normalizing the costs for differences in the MI and no MI default populations. Since fixed costs are one-time costs, they can be compared directly to the value. Variable costs are periodic and are thus expressed on a *per diem* basis of the value. Fixed and variable costs are calculated as follows.

*Calculated Days = minimum of (Actual Days, Allowable Days)*

*Fixed Costs = ∑ Allowable Fixed Costs \* Value*

*∑ Values*

*∑ Allowable Variable Costs*

*Variable Costs per Day = ∑ Values a\* Value*

*Calculated Days*

Algorithm Performance Metrics

Two categories of statistical measures were used to judge the MI Factor algorithm performance: goodness of fit and precision. The measures of goodness of fit track the variance between the observed values and expected values. The measures of precision track the remaining error at an aggregate level. Two additional important considerations in choosing the best algorithm are the ease of implementation and the ease of updating the methodology, for a total of four important criteria for measuring algorithm performance. When weighing the MI factor algorithm options, the algorithm that strikes the best balance between all four criteria was chosen.

* Goodness of Fit: There are three measures used to track the goodness of fit: the mean squared errors (MSE), the sum of squared errors (SSE), and r-squared (R2).

Mean Squared Errors

The average of the squared error, or Mean Squared Error (MSE) tracks the algorithm’s average variance. From a goodness of fit perspective, algorithms with lower MSEs perform better than algorithms with higher MSEs. MSE equals the average (*1/n ∑)* of the observed value (*y)* minus the expected value (*Yhat)*, squared.

*MSE = 1/n ∑ (y –Yhat)2*

Sum of Squared Errors

The sum of the squared error (SSE) tracks the goodness of fit by measuring the algorithm’s total variance. As with MSE, the lower the SSE value, the better the algorithm’s performance from a goodness of fit perspective. SSE equals the sum of the observed value (*y)* minus the expected value (*Yhat)*, squared.

*SSE = ∑ (y –Yhat)2*

*R*-Squared

The coefficient of determination, or *r*-squared, measures how well the independent variables explain the dependent variable, or in this case, how well the loan characteristics explain the foreclosure costs. *R*-squared is expressed as a number between zero and one; the closer that the *r*-squared value is to one, the better the algorithm’s performance. The formula for *r*-squared is as follows.

*R2 = 1 – (SSE / SST)*

Wherethe total sum of squares (SST) is equal to the sum of the observed value minus the average observed value (*Ybar)*, squared. The formula for SST is as follows.

*SST = ∑ (y –Ybar)2*

* Precision: To measure precision, the estimated and the actual outcomes were compared. Monthly totals, averages, and percent differences for both foreclosure costs and MI claim payments were compared.
* Ease of Implementation: Since the MI Factor proposal is intended to simplify the MI claims process, the ease of implementation is an important factor in choosing the best algorithm. The two measures of ease of implementation are the overall grid dimensions and the inverse of the roll-up percent. The grid dimensions are determined by the number of loan characteristics’ bucket combinations. The inverse roll-up percent is the percent of loans using all of the loan characteristics. [[5]](#footnote-5) The algorithm with the smallest grid dimensions and the largest inverse of the roll-up percent is the best from an implementation perspective.
* Updateable Methodology: If the MI Factor proposal is implemented, going forward, there will be no actual data on disallowances and curtailments. Without actual data, there will be no ability to back test future updates. This creates a significant responsibility to specify an MI Factor methodology that can be updated in a reasonable and unbiased manner without the benefit of actual data for back testing and without the use of goodness of fit or precision measures. Furthermore, the methodology also needs to be able to adjust for declining default inventories. The statistical specifications of the loan characteristics do this by providing a clear methodology for dynamic algorithm updates. This eliminates reliance on human judgment that needs to be informed by back testing, goodness of fit, and precision measures and reduces substantially any concerns about bucket-misspecification that may result under these conditions.

**ALGORITHM STRUCTURE OPTIONS**

Of the five loan characteristic inputs, three have specification alternatives: geography, value, and the urban/rural indicator.

* Geography: The algorithm can specify geography by using 1) state, or 2) a statistical clustering methodology to determine state groupings.
* Value: The algorithm can specify value by using 1) $100K buckets, or 2) buckets specified by the value’s mean and standard deviation.
* Urban/Rural Indicator: The algorithm can be specified 1) with or 2) without an urban/rural indicator.

The other two loan characteristics, disposition type and property type, each have only one specification.

* Disposition Type: The algorithm specifies two buckets for disposition type: 1) short sales and 2) TPS and REO.
* Property Type: The algorithm specifies three buckets for property type: 1) one-unit single family properties, 2) condos, and 3) all other property types.

Each of the loan characteristic options have been iteratively analyzed resulting in eight scenarios, changing one factor of the overall calculation method at a time and then analyzing the performance metrics of goodness of fit, precision, ease of implementation, and updatable methodology. Additionally, Appendix D contains a table that compares the performance metrics for all of the algorithms, which may be a helpful reference when reviewing the individual algorithm results.

Option 1: 50 State Geography, $100K Values, With Urban/Rural Indicator

Table 10 provides the algorithm performance metrics for the first option. This option uses the 50-state specification for geography, the $100K value buckets, and includes the use of the urban/rural indicator.

Table 10: Algorithm Option 1



The algorithm specification for option one is the easiest to understand out of all the algorithm options, and the metrics for goodness of fit and precision all indicate a highly effective algorithm. The ease of implementation and the ease of updating the methodology are both negatives, however. The grid dimension count is quite large (at over 1,200 different values), and the methodology for updating is judgment based.

Option 2: 50 State Geography, $100K Values, and No Urban/Rural Indicator

Table 11 provides the algorithm performance metrics for the second option. This option uses the 50-state specification for geography, the $100K value buckets, and excludes the urban/rural indicator.

Table 11: Algorithm Option 2



Option two also uses an algorithm specification that is very easy to understand as the only difference between option two and option one is the exclusion of the urban/rural indicator. The metrics for goodness of fit and precision all indicate a highly effective algorithm and are negligibly different from option one, suggesting that the urban/rural indicator is not terribly predictive. While the grid dimensions decrease due to the exclusion of the urban/rural indicator, the grid dimension count is still relatively large (around 700 possibilities), and the update methodology remains judgment based, making both the ease of implementation and the ease of updating the methodology negatives for this scenario. Based on the negligibly different results for goodness of fit and for precision, we can conclude that the urban/rural indicator does not add significant value to the MI Factor algorithm when using the 50-state geography specification and the $100K value buckets. Given that excluding the urban/rural indicator provides a decrease in the grid dimensions, option two is preferred over option one.

Option 3: 50 State Geography, Statistical Values, With Urban/Rural Indicator

Table 12 provides the algorithm performance metrics for the third option. This option uses the 50-state specification for geography, the statistical value buckets, and includes the use of the urban/rural indicator.

Table 12: Algorithm Option 3



Option three differs from option one in its use of statistical values rather than the $100K value buckets. Using the statistical values makes the algorithm slightly more difficult to understand; however, it provides an increase in the ease of updating due to using statistically driven value updates instead of judgment-based updates. The overall ease of updating is not materially improved, however, because geography is still judgment based. The metrics for goodness of fit and precision all indicate a highly effective algorithm and are negligibly different from option one and two. This option has the largest grid dimension count at 1,398 and one of the smallest inverses of the roll-up percent at 95% making its ease of implementation the worst out of all the options. Given these results, there is really no compelling reason to use option three’s algorithm specification.

Option 4: 50 State Geography, Statistical Values, and No Urban/Rural Indicator

Table 13 provides the algorithm performance metrics for the fourth option. This option uses the 50-state specification for geography, the statistical value buckets, and excludes the urban/rural indicator.

Table 13: Algorithm Option 4



Option four differs from option three only in its exclusion of the urban/rural indicator, with similar impact as noted in option two. From a goodness of fit and a precision perspective, the metrics indicate that this algorithm performs well and is negligibly different from the other three algorithms. While excluding the urban/rural indicator provides a decrease in the grid dimension count, option four’s cell frequency is still larger than option two’s. Essentially, option four has all of the same pros and cons as option three, and there is no compelling reason to use this algorithm specification.

Option 5: Statistical Geography, $100K Values, With Urban/Rural Indicator

Table 14 provides the algorithm performance metrics for the fifth option. This option uses the geography clusters, the $100K value buckets, and includes the use of the urban/rural indicator.

Table 14: Algorithm Option 5



The clustering methodology reduces the number of geography buckets from 50 to eight resulting in option five having a significant decrease in its grid dimensions count – 367 – as compared to the previous four options, the least of which was 725. The clustering methodology also allows for statistically driven updates of geography; however, property value is back to having judgment based updates as we used the $100,000 bucketing. As for goodness of fit and precision, the metrics indicate that this algorithm performs well and is negligibly different from the other four algorithms.

Option 6: Statistical Geography, $100K Values, and No Urban/Rural Indicator

Table 15 provides the algorithm performance metrics for the sixth option. This option uses the geography clusters, the $100K value buckets, and excludes the urban/rural indicator.

Table 15: Algorithm Option 6



Option six differs from option five only in its exclusion of the urban/rural indicator, which provides the smallest grid dimension count out of all of the options at 188. Other than the decrease in the grid dimensions there is no material difference between option six and option five.

Option 7: Statistical Geography, Statistical Values, With Urban/Rural Indicator

Table 16 provides the algorithm performance metrics for the seventh option. This option uses the geography clusters, the statistical value buckets, and includes the use of the urban/rural indicator.

Table 16: Algorithm Option 7



Thus far, option seven is the best from an ease of updating perspective as both geography and value are statistically driven. The ease of implementation is also very good with the grid dimension count at 449, although the count is slightly higher than options five and six due to the statistical values having four buckets as compared to the three $100K value buckets. From a goodness of fit and a precision perspective, again, the metrics indicate that this algorithm performs well and is negligibly different from all the other algorithms.

Due to the significant increase in the ease of updating, the good performance from an ease of implementation perspective, and the metrics for goodness of fit and precision indicating a highly effective algorithm, so far, option seven does the best job of balancing the four performance criteria.

**MI FACTOR RECOMMENDED ALGORITHM**

Option 8: Statistical Geography, Statistical Values, and No Urban/Rural Indicator

Table 17 provides the algorithm performance metrics for the eighth and final option. This option uses the geography clusters, the statistical value buckets, and excludes the urban/rural indicator.

Table 17: Algorithm Option 8



Option eight is the same as option seven except that is does not use the urban/rural indicator. Other than bringing the grid dimension count down to 237, excluding the urban/rural indicator does not have a material impact on the algorithm performance as compared with option seven.

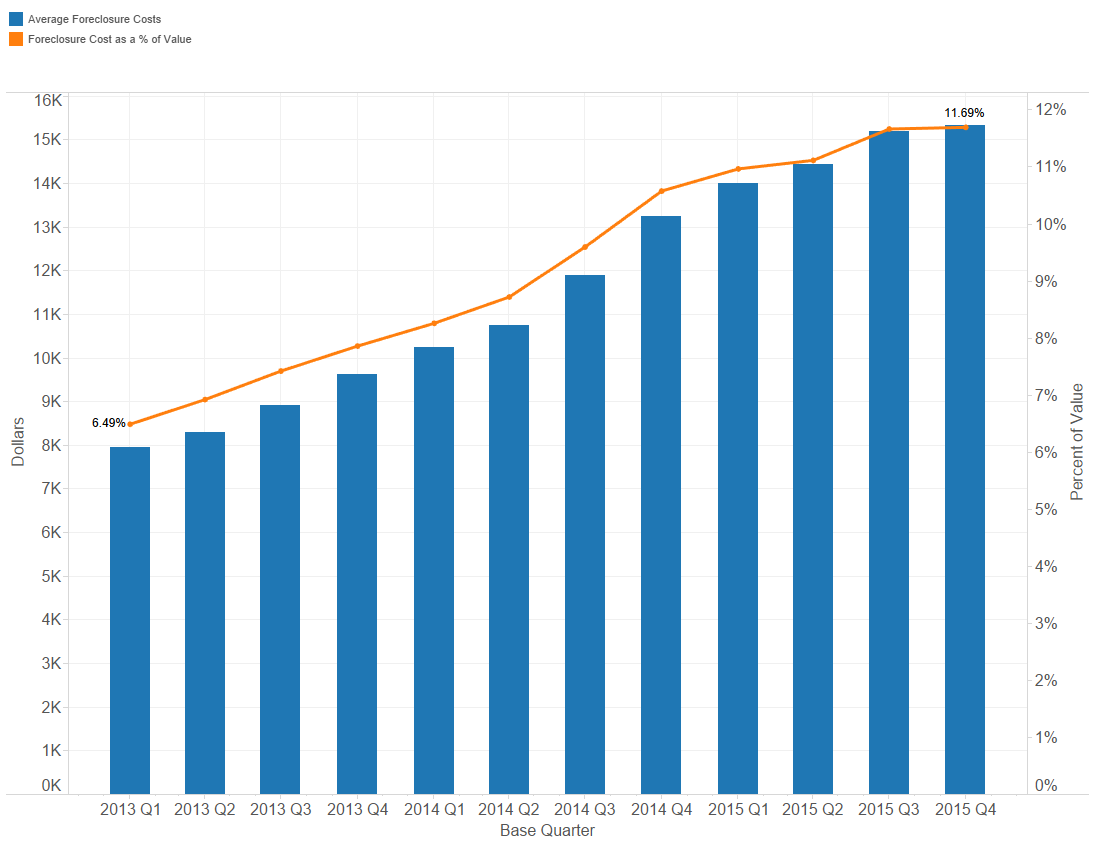
Out of all of the options, eight strikes the best balance between each of the four performance criteria. From a goodness of fit and from a precision perspective, all of the algorithms perform very well, and none of these metrics differ materially between the options. This fact leaves us free to choose the algorithm that provides the easiest implementation and updates. Having both geography and value statistically driven is a significant advantage from an ease of updating perspective. Additionally, excluding the urban/rural indicator provides an improved ease of implementation as compared to option seven, with little cost to overall performance. Because of this, option eight is the recommended MI Factor algorithm specification.

**MI FACTOR IMPLEMENTATION**

Implementation Overview

A grid implementation was chosen for the MI Factor algorithm for its advantages of being straight-forward and transparent. A weakness of this implementation; however, is that it cannot capture trends in the way that, for example, a predictive regression using continuous dimensions can. In other words, MI Factor can only capture average foreclosure costs at a point in time. If costs are increasing or decreasing rapidly, the algorithm is not able to apply this trend into the future. While a regression would not have this weakness, in addition to being less transparent, a regression implementation is not possible for MI Factor since there will no longer be any actual disallowance or curtailment data going forward, which is necessary for rigorously back testing regression updates. The impact of this inability to capture foreclosure cost trends is small, however. A number of geographies showed rapidly increasing foreclosure costs during the estimation time period of July 2013 – June 2014, which had little impact on the algorithm’s overall precision in estimating 2015 foreclosure costs and MI claim payments.

Figure 16: Foreclosure Costs over Time



Roll-up Methodology

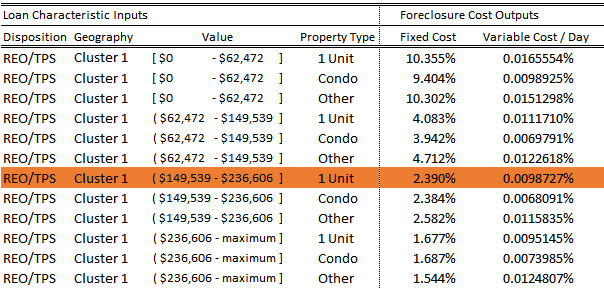
The implementation employs a roll-up methodology whereby if there are less than 30 observations in one of the grid’s buckets, the next highest bucket will be used, a common structural choice in statistical applications. This is to prevent reliance on very limited data points from skewing the overall results. Therefore, overall structure, or hierarchy, of the order in which buckets are applied matter for a limited number of cases in the proof of concept. For example, if there are only 100 defaults during the estimation time period that disposed as a short sale located in geographic cluster 6 with a value greater than $200,000. Of these, 85 were one-unit properties, 15 were condos, and 5 were some other property type. Since both the condos and the “other” property types have less than the 30 observation minimum, the property type loan characteristic would not be used when calculating the foreclosure cost averages for these defaults, and the average foreclosure costs of all 100 defaults would be used instead. The property type loan characteristic would, however, be used for the one-unit population since 85 is greater than the 30 observation minimum threshold. We proposed the ordering where, in the event of lower than 30 frequency of events, property type would be disregarded first, then the value-hybrid bucketing, then geographic clusters, and finally disposition method. This aligns with the explanatory power of each dimension; property type contributes the least to explanatory power while disposition method the most. Additionally, the inclusion of loans with and without MI coverage in the estimation data helps diminish this problem.

Grid Implementation Example

As foreclosure costs are a function of the loan characteristics, the proposed grid implementation uses the loan characteristics as the inputs and has fixed and variable costs as the outputs. The following provides an example of how to use the grid to solve for fixed and variable foreclosure costs. For example, a borrower defaults on a mortgage with the following loan characteristics.

* Disposition Type: REO
* Geography: Cluster 1
* Property Value (UPB): $150,000
* Property Type: 1-Unit
* Urban/Rural Indicator: NA

Table 18: Grid Example



Using the loan characteristics to look up the fixed and variable cost percentages in the grid, the fixed and variable cost amounts can be calculated as follows.

*Fixed Cost Amount = Fixed Cost Percent from Grid \* Property Value*

*Variable Costs per Day = Variable Cost Percent from Grid \* Property Value*

*Fixed Cost Amount = 2.390% \* $150,000 = $3,585*

*Variable Cost Per Day = 0.0098727% \* $150,000 = $14.81*

MI Factor Calculation & Example

Once the grid values have been used to calculate the fixed and variable cost amounts, MI Factor can be calculated. Continuing the example, the defaulted mortgage also has the following loan characteristics.

* LPI Date: January 1, 2015
* Liquidation Date: September 30, 2015
* Allowable Days: 240

MI Factor is calculated as follows.

*Actual Days = Liquidation Date – LPI Date*

*MI Factor = Fixed Costs + (Variable Costs per Day \* minimum (Actual Days, Allowable Days))*

*Actual Days = September 30, 2015 – January 1, 2015 = 272 Days*

*MI Factor = $3,585 + ($14.81 \* minimum (272, 240)) = $7,139.17*

MI Claim Amount Calculation & Example

MI claim amount is the sum of the UPB, interest, and the foreclosure cost. The proposed changes to the interest and the foreclosure cost components are covered below. There are no changes proposed for the UPB component of the MI claim amount calculation.

* Interest: Assuming that our example loan has a note rate of 5%, the proposed new curtailed interest rate calculation is as follows.

*Interest = UPB \* (Note Rate / 365.25) \* minimum (Actual Days, Allowable Days)*

*Interest = $150,000 \* (5% / 365.25) \* minimum (272, 240) = $4,931.51*

* Foreclosure Costs: The foreclosure cost component is equal to the MI Factor.

*Foreclosure Costs = MI Factor = $7,139.17*

The proposed new MI claim amount equation is as follows.

*MI Claim Amount = UPB + Curtailed Interest + MI Factor*

*MI Claim Amount = $150,000 + $4,931.51 + $7,139.17 = $162,070.68*

MI Claim Settlement Example: Option Settlements

This example only covers the proposal for option settlements since presales with influence and MI acquisitions are still being researched. Assuming that our example loan has an MI coverage percent of 20%, option settlements are calculated as follows.

*MI Claim Payment = MI Claim Amount \* MI Coverage Percent*

*MI Claim Payment = $162,070.68 \* 20% = $32,414.14*

Update Schedule & Methodology

Annual updates are proposed for the MI Factor algorithm. As described on earlier in this document, the estimation data used in the updates will be Fannie Mae’s actual foreclosure cost data, and the timeframe used will be a rolling 12-month window with a six-month lag. The lag allows time for all of the foreclosure costs to be submitted to Fannie Mae. For example, if the MI Factor grid is estimated on July 2015 – June 2016 data, the results will be implemented for all MI claims filed in 2017. The base date used to define the population will be liquidation date for short sales and TPS and will be disposition date for REO. Additionally, the estimation data will be adjusted so that Fannie Mae is self-disallowing and self-curtailing the MI claims.

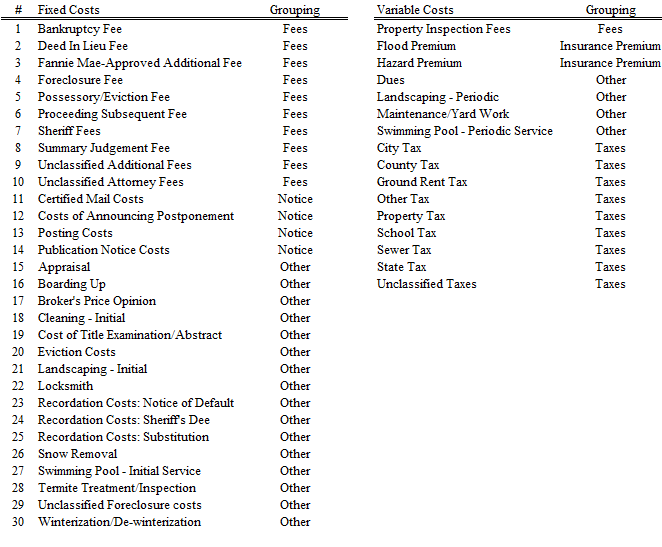
**PROOF OF CONCEPT OVERVIEW**

In order to provide the greatest possible transparency into the MI Factor algorithm research, Fannie Mae is providing a proof of concept package. This package includes the estimation data and the SAS code that was used to generate the algorithm research. The package also includes output data for each of the algorithm options as well as their corresponding grids. This proof of concept allows the MI companies to do their own research and to develop their own educated opinions on the MI Factor proposal and on the recommended algorithm of option eight. A detailed overview of the files included in the proof of concept package can be found in Appendix E. A similar set of material would be provided as part of the annual update process to the factor settings for on-going vetting activities.

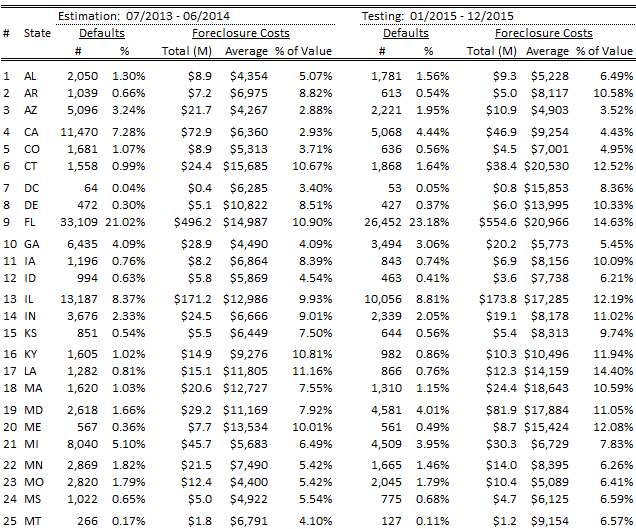
**APPENDIX A: ALLOWED & DISALLOWED FORECLOSURE COSTS**



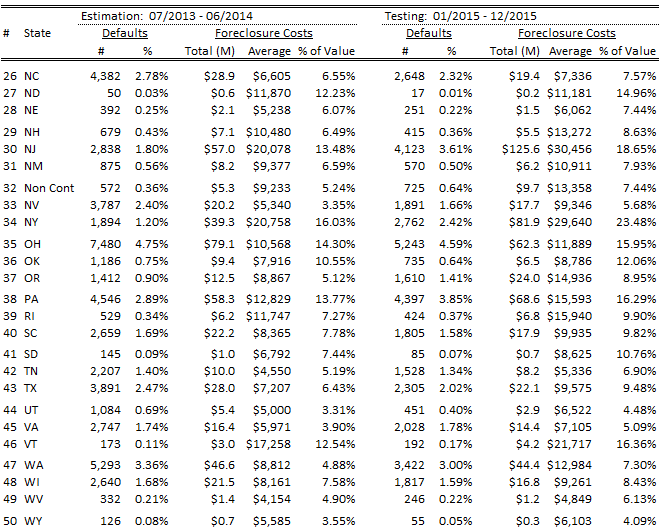
**APPENDIX B: FIXED & VARIABLE FORECLOSURE COSTS**



**APPENDIX C: FORECLOSURE COSTS BY 50 STATES (STATES 1 – 25)**



**APPENDIX C, CONTINUED: FORECLOSURE COSTS BY 50 STATES (STATES 26 – 50)**



**APPENDIX D: ALGORITHM PERFOMANCE METRICS COMPARISON**



**APPENDIX E: PROOF OF CONCEPT OVERVIEW**

To afford transparency, a proof of concept is provided that includes the input data, SAS code, and the output data that Fannie Mae used to create the algorithm options. Using the proof of concept, Fannie Mae’s MI Factor algorithm research should be able to be reconstructed.

**INPUTS**

The first step in reconstructing the MI factor analysis is understanding the input data provided. The file MI\_Factor\_Sample.csv contains all of the input data that was used to develop the recommended algorithm and has two main attributes: population criteria and included fields.

Population Criteria

In order to analyze the algorithm’s performance, the input data is split between the estimation and the test populations.

* Estimation Population: This population contains all of the loans used to estimate the MI Factor values. It includes all short sale and TPS liquidations and all REO dispositions occurring from July 2013 through June 2014 regardless of the presence of MI coverage.
* Testing Population: This is the population that the estimated MI factor results are tested against. It includes MI-covered short sale and TPS liquidations and REO dispositions occurring from January through December 2015.

To test the algorithm, the MI Factor results from the estimation population are compared with the actual MI claim results from the test population. This mirrors the future implementation of the algorithm as historical MI claim data will be used to settle future claims.

Included Fields

There are five categories of input data fields: distinct id, algorithm inputs, claim amount inputs, expected claim proceeds inputs, and key indicators.

* Distinct ID: Row ID is the unique key used to join the input and output data sets. It is simply a record counter; all distinguishing attributes such as Fannie Mae’s loan number have been removed.
* Algorithm Inputs: The below fields are required by the SAS code to generate all eight MI factor algorithm options.

Disposition Type State UPB

Sale Proceeds Property Type Urban/Rural Indicator

Fixed Costs Variable Costs Base Date

Calculated Timeline Allowable Timeline Allowable Delays

Population Indictor Post Guidelines Date

* Claim Amount Inputs: The below fields are also required to calculate the MI claim amount.

UPB Note Rate Liquidation Date

LPI Date Allowable Timeline Allowable Delays

Algorithm Inputs (above)

* Expected Claim Payment Inputs: In addition to the claim amount inputs listed above, the MI coverage percent is needed as an input for the expected claim payment.
* Key Indicators: There are two key indicators that do not fall into one of the prior categories: MI indicator and diagnostics flag.
  + MI Indicator: This field indicates whether or not the loan is covered by an MI policy at the time of mortgage default, excluding loans under a Delayed Payment Option (DPO) contract. An allowable value of “1” means that the loan is covered by an MI policy; a value of “0” means that the loan is not MI covered.
  + Diagnostics Flag: This flag was designed to simplify the process of evaluating algorithm performance. Loans with an allowable value of “1” should be included when calculating algorithm performance metrics, and loans with a value of “0” should not. A value of “1” indicates that the loan has the following characteristics.
    - Population Indictor = Test Population
    - MI Indicator = 1
    - Actual MI Proceeds > 0
    - 1 ≤ MI Coverage Percent ≤ 40
    - Not Settled as a Presale with Influence

**SAS CODE**

We leveraged SAS, a statistical and data processing suite of software, in order to research the proposals in this document. A streamlined set of this SAS code, MI\_Factor\_Algorithm.sas, that was used to create the recommend MI factor algorithm is also provided as part of the proof of concept.

After reading in the provided input data, the code performs the algorithm analysis. The following bullets provide a high level breakdown of the main analytical portions of the SAS code.

* Statistically Driven State Clustering: This portion of the code uses the cost characteristics of states to group similar states into clusters. Clustering is used to decrease the number of grid dimensions needed while having a minimal reduction in algorithm performance. Additionally, this process is setup to automatically update every year without judgment-based interference.
* Statistically Driven Value Bucket: This portion of the code uses the mean and standard deviation of the value bucket to group values after accounting for the disposition type bucket. This process is setup to automatically update every year without judgment-based interference.
* Calculation of Cost Components: This portion of the code calculates the average fixed costs and variables costs per day for every combination of the loan characteristics.
* Roll-Up Type: This portion of the code evaluates whether or not there is the sufficient loan coverage of 30 observations within each bucket of the MI Factor grid. If there is not enough coverage, then the next highest level of the grid is used.
* Calculation of Estimated Foreclosure Cost and Claim Proceeds: This portion of the code calculates the estimated foreclosure cost and claim proceeds for every loan. This is calculated by applying the correct roll-up type and cost components at the loan level.

**OUTPUTS**

The code outputs several .csv and .pdf files that contain the algorithm outputs and performance metrics. Each MI factor algorithm option has its own set of .csv and .pdf files. The algorithm output data and MI Factor grids are provided as .csv files. The .pdf files contain the algorithm performance metrics. These metric files are not provided as part of the proof of concept package since all of the important metrics are included in this paper. Should the algorithm metric files be desired, the SAS code may be used to create them.

Algorithm Performance Metric Files

These .pdf files contain the algorithm performance metrics used to analyze the algorithm options. They are not provided as part of the proof of concept but can be created using the SAS code provided.

Proof of Concept Output Files

There are three kinds of output files provided as part of the proof of concept, the state cluster statistics, the algorithm output data files, and the MI Factor grids. The output data and the grid files are provided for each of the eight algorithm options making 17 output files in total.

* State Cluster Statistics: This .pdf file contains the analytical output that supports the specific state cluster outcome.
* Output Data File: The output data file MI\_Factor\_Option[#].csv includes all of the fields created by the SAS code. Joining the input and output data provides the information needed to evaluate the performance of each algorithm. The below fields are included in this file.

Row ID Property Type Value

Diagnostics Flag Value Amount

Liquidation Bucket Geography – State Buckets

Value Amount – $100K Buckets Property Type

MSA Indicator (Urban/Rural) MSA NA (placeholder)

Foreclosure Timeline Geography – Cluster

Value Amount – Statistical Roll-up Type

Actual Foreclosure Cost Actual MI Payment

Variable Amount Per Day Average Fixed Amount

Average Variable Amount Per Day Average Variable Amount Per Day Percent

Estimated MI Proceeds Estimated Foreclosure Costs

Estimated Interest

* MI Factor Grid: The data file MI\_Factor\_Grid\_Matrix\_Option[#].csv contains the MI factor grid, which includes the loan characteristics and the resulting fixed and variable foreclosure cost components. The following fields are included.

Disposition Type Geography

Value Property Type

Urban/Rural Indicator Percentage of Avg. Fixed Costs to Value

Percentage of Avg. Daily Variable Costs to Value

1. For the purposes of this paper, mortgage default is defined as a loan liquidation via a short sale (PFS), third party sale (TPS), or real estate owned (REO) sale. [↑](#footnote-ref-1)
2. The other types of issues which can impact MI claim payment are typically referred to as ‘Denials’ and ‘Rescissions’. Denials are typically associated with documentation missing from the claims filing process or extreme servicing issues. Rescissions are typically associated with violations of the MI company’s guaranty policy similar to a Fannie Mae Representation & Warranty violation. Both of these issues are *not* considered in the MI Factor proposal discussed here. [↑](#footnote-ref-2)
3. A full listing of the allowable delays and the amount of time associated with these per standard policy is available on the second page of our foreclosure timelines exhibit. [↑](#footnote-ref-3)
4. <https://www.fanniemae.com/content/guide_exhibit/foreclosure-timeframes-compensatory-fees-allowable-delays.pdf> [↑](#footnote-ref-4)
5. Grid buckets containing fewer than 30 loans are not used since the number of observations are deemed too small to be representative. As a result, loans in these buckets are “rolled-up” to the next highest grid level. Roll-up is explained further in the implementation section. [↑](#footnote-ref-5)