Data Organization and Analysis in Mortgage Insurance: The Implications of Dynamic Risk Characteristics

Tanya Havlicek and Kyle Mrotek, FCAS, MAAA

Abstract

The capability for mortgage guaranty insurance companies to establish loss reserves conditioned on a dynamic risk characteristic, delinquency status, presents particular data issues. There is a need to collect, organize, warehouse, and analyze large data sets that contain loan-level detail over consecutive evaluation dates in order to measure the probability of claim, conditioned on delinquency status. The generally accepted methodology of reserving for mortgage guaranty insurance claim liabilities requires evaluation of dynamic risk characteristics because mortgage guaranty insurance companies need only reserve for loans currently delinquent, both known and IBNR. Because each loan's delinquency status is usually revised monthly by the mortgage servicing company, the cohort of insured loans currently delinquent changes each month and therefore is dynamic with respect to time. Coincidentally, delinquency status has been found to be a strong predictor of future losses, so it is imperative for mortgage guaranty insurance companies to estimate reserves as a function of delinquency status, a dynamic risk characteristic. Maintaining historical economic factors in step with the historical delinquency and claim data can also enhance the reserving approach.

1. INTRODUCTION

The generally accepted methodology of reserving for mortgage guaranty insurance claim liabilities is to reserve for loans currently delinquent, both known and IBNR. Mortgage guaranty insurance companies do not reserve for loans insured but not delinquent [1]. Estimating reserves requires the evaluation of dynamic risk characteristics because each loan's delinquency status is, typically, revised monthly by the mortgage servicing company. Therefore, the *cohort* of insured loans currently delinquent in a given month for which the mortgage guaranty insurance company needs to reserve changes each month and is dynamic with respect to time. Many delinquent loans do not result in a loss. However, the delinquency status of a loan is an established strong predictor of future losses [2], so it is imperative for mortgage guaranty insurance companies to estimate reserves as a function of delinquency status, a dynamic risk characteristic.

The capability for mortgage guaranty insurers to establish loss reserves conditioned on delinquency status presents particular data issues. There is a need to collect, organize, warehouse, and analyze large data sets that contain loan-level detail over consecutive monthly evaluation dates in order to measure the probability of claim conditioned on delinquency status. When a loan becomes delinquent, it can maintain the same delinquency status, become progressively more delinquent, or move back and forth between delinquency stages before eventually resolving into one of two fates: it may become current in payments and be considered cured, or it may remain in default and result in a claim. There is then a need to track the eventual fate of each delinquency over consecutive monthly evaluations to its ultimate cure or claim.

The ability to distinguish and quantify delinquency trips and subsequent fates for all delinquent loans and then aggregate that data along risk-characteristic dimensions to develop reserving factors requires data availability and storage over consecutive monthly evaluation dates. Otherwise, the capacity to track fates and calculate empirical conditional claim probabilities is lost in data uncertainties. The dynamic nature of loan delinquency status manifests itself in mortgage guaranty insurance reserving in two aspects: calculating conditional claim frequencies from historical delinquencies to create reserving factors and identifying the current cohort of delinquent loans that need reserves, both reported and unreported.

2. BACKGROUND

2.1 Mortgage Guaranty

The Mortgage Guaranty Model Act of the NAIC defines mortgage guaranty insurance as insurance against financial loss by reason of nonpayment of principal, interest, or other sums agreed to be paid under the terms of any note or bond or other evidence of indebtedness secured by a mortgage, deed of trust, or other instrument constituting a lien or charge on real estate, providing the improvement on such real estate is designed for residential occupancy or industrial or commercial purposes [3].

The rationale for the existence of mortgage guaranty insurance is to disperse the credit risk of borrowers defaulting on their mortgages [4]. Lenders can offer borrowers mortgages more cheaply when the cost of mortgage insurance is factored in [4]. This is true because, like any insurable risk, the law of large numbers makes the variance around the mean smaller for more insured risks. Investors providing funds to the mortgage lenders (through several channels and ultimately the purchase of mortgage-backed securities) require lower returns when the credit enhancement of mortgage guaranty insurance is applicable and when this cost savings is more than the additional cost to the borrower [4]. In the end, the coupling of these phenomena allows more people to buy homes than would be able to otherwise.

Mortgage guaranty insurance is considered a property and casualty line of business, but it has notable differences from more traditional property and casualty lines of business.

The NAIC requires mortgage guaranty insurers to be monoline insurers. That is, in general, mortgage guaranty insurers are only allowed to underwrite mortgage guaranty insurance and not other lines of business [5]. As a result, mortgage guaranty insurers generally diversify through geographic and temporal underwriting initiatives. Mortgage guaranty insurance losses are strongly correlated with macroeconomic events such as home price appreciation, unemployment, and interest rates [6]. Economic recessions tend to be concentrated regionally, so, to the extent a mortgage guaranty insurer is diversified geographically, the mortgage guaranty insurer's performance should vary less. Further, because mortgage guaranty insurance losses are strongly correlated with macroeconomic events, loans insured over extended underwriting periods are affected. As a result, mortgage guaranty insurers benefit from having a portfolio of insured loans underwritten over an extended period of time because the houses of loans insured years ago tend to have appreciated more in home price than houses of loans insured recently.

In contrast, traditional property and casualty insurers in general are allowed to underwrite multiple lines of business and often do so. A review of the 2006 annual statements for nearly 3,100 U.S. property casualty insurance companies indicates that only 2% of the filed direct and assumed earned premium for calendar year 2006 was from monoline insurance companies as measured by Schedule P lines of business.

Further, mortgage guaranty insurance policy terms are generally several years long and, depending on the amortization period of the mortgage, can be as long as 20 years. This is in contrast to traditional property and casualty policies with terms of one year or even six months. Additionally, the policies are generally noncancellable by the insurer except for nonpayment of premium. In other words, a policy can not be re-underwritten periodically as is common with traditional property and casualty lines of business. As a result, the premium rate schedule is stipulated at policy issuance.

In general, mortgage guaranty insurers offer three types of premium payment: monthly, annual, and up-front. The majority of mortgage loan borrowers engage in policies requiring premium payment on a monthly basis. The mortgage loan borrower submits a payment to the mortgage loan servicer each month. Depending on the mortgage product, the monthly payment includes amounts for principal, interest, hazard insurance, property taxes, and mortgage guaranty insurance premium. The mortgage loan servicer then submits the mortgage guaranty insurance premium to the mortgage guaranty insurer on a monthly basis. The premium is earned immediately by the mortgage guaranty insurer, as there is not an unearned premium reserve affiliated with monthly policies. Much less frequently, the policy may call for annual premium payments instead of monthly and, depending on regulations and specifics, there may or may not be an unearned premium reserve. Finally, single up-front premium policies generally include a provision for an unearned premium reserve.

Another feature of mortgage guaranty insurance that differs from traditional property and casualty lines of business is the relationship between the beneficiary and the premium payer. In traditional property and casualty lines of business, the premium payer is also the beneficiary. For example, auto liability insureds pay premiums and are covered against liabilities against them. By contrast, in mortgage guaranty insurance, borrowers pay the premium, and in the event of default, the mortgagee is the beneficiary and is reimbursed by the mortgage guaranty insurance company.

2.2 Reserving

Property and casualty insurance companies are generally required to maintain loss and loss-expense reserves. Mortgage guaranty insurance companies are generally classified as property casualty insurance companies, so it follows that mortgage guaranty insurers must also maintain loss and loss-expense reserves (mortgage guaranty insurance is classified as line "S" Financial Guaranty/Mortgage Guaranty in Schedule P of annual statements for NAIC property and casualty insurance companies).

The Mortgage Guaranty Model Act of the NAIC reads "A mortgage guaranty insurance company shall compute and maintain adequate case basis and other loss reserves which accurately reflect loss frequency and loss severity and shall include components for claims reported and for claims incurred but not reported, including estimated losses on:

1. Insured loans which have resulted in the conveyance of property which remains unsold;
2. Insured loans in the process of foreclosure;
3. Insured loans in default for four months or for any lesser period which is defined as default for such purposes in the policy provisions; and
4. Insured leases in default for four months or for any lesser period which is defined as default for such purposes in policy provisions."

As a note, mortgage guaranty insurance policy provisions generally stipulate that a loan is in default (a.k.a., delinquent) the moment one monthly payment is not made and until the time at which that payment and accrued interest have been repaid.

The list of four items above presents particular data and projection issues for the actuary in estimating loss and loss-expense reserves. As mentioned previously, the delinquency status of the mortgage is a strong predictor of the likelihood of claim. Conveyance, foreclosure, and length of default indicate various delinquency statuses. Further, within default, the duration in which the loan has been in default also provides information on the likelihood of a claim. In general, the longer a loan has been in default, the more likely the loan will not cure and potentially lead to a claim. As such, loans in default for three months tend to be more likely to cure than loans in default for nine months. To better estimate the reserve, it serves the actuary well to be able to differentiate the probability of claim between loans in default with various statuses (i.e., one month, two months, three months, etc.). In order to estimate the probability of claim conditioned on delinquency statuses, the actuary may want data on the resolution of historical delinquencies, given delinquency status.

A reserving method commonly employed to estimate reserves consistent with the Mortgage Guaranty Model Act of the NAIC is a frequency-severity methodology. The frequency component of the method is incorporated by applying a probability of claim given that a loan is delinquent. In choosing the frequency factor (i.e., probability of claim conditioned on being delinquent), the reserving actuary will want to consider delinquency status, underwriting risk characteristics, and macroeconomic variables. Delinquency status can be based on the number of monthly payments missed or how long the loan has been consecutively delinquent. Potential underwriting risk characteristics include loan-to-value, borrower credit rating (e.g., FICO® Score), and property geography. Consideration for economic variables is addressed later in the article.

The severity component of the frequency-severity method can be viewed either as one factor net of salvage/subrogation, or as two components: loss given default (before salvage/subrogation) and recovery (salvage/subrogation). As is typically assumed with the frequency-severity method, the severity factor is the estimate of loss given that a claim occurred. Because the severity factor is often conditioned on a claim having occurred, the actuary may not want the quantification of the severity factor to be a function of delinquency status, in which case the premise of challenges posed by a dynamic input variable is moot. However, to the extent the actuary wants to reflect the dynamic delinquency status as an input into the severity estimate, it would pose further challenges not addressed in this article. The challenge in particular is the need to track not only a binary result (i.e., cure or claim) but also a loss (potentially relative to a coverage amount) along a continuum with respect to the dynamic delinquency status from month to month.

3. DATA ORGANIZATION

There are two types of loan characteristics that must be stored: dynamic and static. Static characteristics are those that do not change over the lifetime of the loan. The static characteristics database will contain loan information that the actuary may want to use as dimensions in developing reserving factors. Examples of static characteristics include original loan-to-value and borrower original FICO score. The static characteristics can be stored in a single policy-record database that is updated as new policies are insured.

Dynamic characteristics are those that can change monthly, such as delinquency status, and should be stored in a database that contains every monthly evaluation date. The database of dynamic characteristics contains a record for every month of the loan's lifetime and is compiled by appending the revised values each successive month. The dynamic status needs to be stored for every monthly evaluation so that historical delinquency cohorts and their fates can be accurately reconstructed and analyzed.

Size requirements and processing time may make it infeasible or impractical to store all attributes of all loans at every month, thus the segregation between the static and dynamic databases. Further, it is not necessary to store the static characteristics along with the dynamic fields. Consider the database size necessary to store monthly records of 100 fields for 100,000 loans for 156 months (a single book year of business over its lifetime) when only five of the 100 fields can change during the loan's lifetime. Then repeat this to add five book years of business. Clearly the database size will grow rapidly. It suffices to have a unique primary key ID field in the two databases such that information from the databases can be merged one-to-many correctly. The field typically used for this purpose is the policy's certificate number and is generally assigned by the mortgage guaranty insurer.

Before data storage space became relatively abundant and inexpensive in recent years, historical performance data was frequently purged or overwritten. One result of this practice is a "data vacuum," where information on prior delinquencies of a loan that cured is lost when the loan again becomes delinquent later on. In this instance, "prior delinquencies" for a particular mortgage guaranty insurance policy refer to all delinquencies except the most recent one. The absence of exhaustive historical performance data will almost always occur if only a single record with key status dates is kept for each insured loan policy, as opposed to storing delinquency information from every monthly evaluation date in the dynamic characteristics database. Overwriting key historical status dates, or purging, makes it impossible to reconstruct or analyze the delinquency behavior and exposure over time.

Other data organizational challenges may occur if the mortgage guaranty insurance company does not keep delinquency and claims information in the same database environment because the two types of data are handled by different departments. This makes it challenging to match delinquency cohorts and resolution fates, particularly if the separate departments store data with incompatible database systems or key ID fields.

The preferred time window for collection and storage of loan-level data is, simply, as long as is possible. Data availability will determine, in part, the constraints of the analysis. Obviously, newer companies will have less performance data accumulated than older companies. At a minimum, to develop frequency and severity factors, there need to be enough consecutive months of complete data to observe a credible amount of delinquency resolutions. The more granular the reserving methodology, the more voluminous the data must be.

The time window of delinquency performance data should be long enough to allow many historical cohorts of delinquent loans to fully resolve into cure or claim. Although data maintained only quarterly can be used for such analyses, it reduces both the amount of data available to build credibility for analysis and the resolution of the analysis. Quarterly interval data also requires the use of additional assumptions because a loan could cure and become delinquent again during the three-month interval comprising a quarter, resulting in decay of the resolution accuracy of delinquency performance. Also, the mortgage guaranty insurance company need only carry reserves when the loan is delinquent.

Ideally, the historical time window the actuary considers for analysis should be long enough to capture an entire economic cycle, because delinquency, foreclosure, and claim rates are influenced by economic factors (e.g., unemployment rate).[Error! Bookmark not defined.] If the insurer's volume or product mix is volatile or heterogeneous, the ideal time window would capture behavior changes that could result from these changes and shifts. Also, mortgage insurance policies have extended policy terms relative to other property and casualty insurance policies, so the time window also should be long enough to observe the claims development for several policy years of business from inception to ultimate resolution.

Not only is the historical time period of data collection important, it is also imperative that the periods for which data are collected be contiguous. If there are "holes" where some months of delinquency activity are absent, it is impossible in many cases to determine the ultimate fate of any loan actively delinquent prior to and leading into the missing evaluation date. Further, the delinquency cohorts of the missing months cannot be used for data analysis. Data holes arise for various reasons. A loan-servicing company may not report to the mortgage guaranty insurer at every monthly evaluation date. Consultants have client relationships that may not be engaged in perpetuity, in which case the client would not provide data to the consultant when there is not an active engagement. If the client-consultant relationship re-engages, the performance data from the period of relationship inactivity may no longer be available in its entirety because it was purged by the client for storage reasons or because the client contractually cannot provide the consultant with data from that period.

Despite the hole of historical performance data and its negative consequences, it may be possible to salvage information from the data. Although each data hole can lead to a significant loss in full information on many loans, depending on the number of policies in force, there may still be complete information on enough loans for credible analysis. The actuary will need to determine on a case-by-case basis whether enough information remains. Unfortunately, in addition to there being less information available to harvest from an incomplete historical data set, the amount of effort required to compensate for the inadequacies can also be far more than with a full data set. The actuary's programming code may require more "do loops," "if-then-else" conditions, and likely run-time. When there are data holes, there is less information on what has happened to a delinquency and, thus, there are more fate conditions or resolution possibilities to be considered. In some cases, it may be impossible to extract any information from the vicinity of the data hole.

The desired data organization is comprised of two databases that have been maintained for a period long enough to contain a credible amount of resolutions on delinquent loans. Both databases have a unique or key ID field specific to each loan, typically loan number or certificate number. Further, the unique ID per loan should be the same across databases. One database contains a single record for each loan, sometimes called the master policy file, of static loan and borrower characteristics (e.g., underwriting characteristics) that do not change over the life of the loan or that identify when the loan became inactive (termination or claim). Records for newly issued insurance policies can be appended to this database as new policies are written.

The second database is made up of dynamic loan characteristics that can change monthly and, therefore, it is updated monthly. However, just because it is updated monthly does not mean old information should be purged to make way for the new information. For each insured policy, the database of dynamic loan characteristics contains each evaluation date applicable to that policy and the dynamic status of each loan as of each of those dates. Specifically, the database of dynamic loan characteristics contains information on whether the loan is in force (an active policy) or not in force and the loan's status (current, delinquent, claim) for each evaluation date. As a note, a current, or nondelinquent, loan that is no longer in force is a terminated loan that is no longer insured by the mortgage guaranty company.

3.1 Storage Considerations

By including dynamic date and status fields in the master policy database, which would need to be updated monthly, accurate delinquency histories can be constructed from dynamic databases that contain only the delinquent loans from each evaluation date. Some additional fields needed for reconstruction are current delinquency status and the date the loan achieved current delinquency status.

Warehousing only delinquent loans in the dynamic database requires more assumptions, merging, and date logic to program and process than using a dynamic database that tracks all loans ever written at every evaluation date; however, the decrease in data storage and program-processing-time requirements may make this organization more desirable than the "desired" organization described above. The design decision on how to organize the database will be based on the business requirements of the user and the hardware and software platforms that will support the data. If all loans ever written are included in the dynamic database, there is no ambiguity associated with an omitted loan. Examples of ambiguity include delinquency status as of evaluation dates and whether the loan is in force or terminated. By storing only delinquent loans as of each evaluation date, a loan may not appear at a given evaluation date for at least two reasons: the loan is no longer delinquent or there has been a data error. If all loans ever written are in the dynamic database, an omitted loan indicates a data error, either because the loan was accidentally omitted or because the loan did not belong to the mortgage guaranty insurance company and should be removed. However, depending on the size and age of the business, these files can rapidly become quite large. Clearly, warehousing only delinquent loans will use less storage than keeping a monthly status on all loans ever written.

4. DATA PROCESSING

A delinquency cohort is the group of all loans delinquent as of the reserving evaluation date. Consider the following table, which presents a simplistic example of five loans over six months. The group of all delinquent loans at each evaluation date comprises six delinquency cohorts. Note that a particular loan delinquent for, say, three consecutive months will be part of those three delinquency cohorts. Table 1 is an example of the record layout from a dynamic characteristics database with four fields added for processing the data. Columns A-D come from the dynamic characteristics database (column A is implicit and shown for explanatory purposes), whereas columns E-H are added on during the program processing.

Table 1.

A Record #	B Evaluation Date	C Loan ID	D Status*	E Delq	F Delq Trip	G Cure	H Claim
1	Jan-06	1	0				
2	Jan-06	2	30				
3	Jan-06	3	90				
4	Jan-06	4	60				
5	Jan-06	5	0				
6	Feb-06	1	0				
7	Feb-06	2	30				
8	Feb-06	3	90				
9	Feb-06	4	0				
10	Feb-06	5	30				
11	Mar-06	1	30				
12	Mar-06	2	30				
13	Mar-06	3	120				
14	Mar-06	4	30				
15	Mar-06	5	60				
16	Apr-06	1	0				
17	Apr-06	2	30				
18	Apr-06	3	FCL				
19	Apr-06	4	60				
20	Apr-06	5	30				
21	May-06	1	30				
22	May-06	2	30				
23	May-06	3	FCL				
24	May-06	4	FCL				
25	May-06	5	0				
26	Jun-06	1	0				
27	Jun-06	2	30				
28	Jun-06	3	CLM				
29	Jun-06	4	CLM				
30	Jun-06	5	0				

* 0 = Current; 30, 60, 90, 120 = days past missed mortgage payment; FCL = foreclosure; CLM = claim

(Note: Loans need not progress through delinquency categories consecutively or unidirectionally. For example, a loan can go from 90 days delinquent to 120 days delinquent to 30 days delinquent over three consecutive months. The jump backward from 120 days delinquent to 30 days delinquent in just one month can occur when the borrower makes up for several missed monthly mortgage payments at once).

The goal in processing the data is to determine the fate of each loan for every month it is delinquent, while distinguishing delinquency trips, so that claim ratios can be calculated for each cohort of loans. Delinquency trips are important because if a loan cures, it no longer needs a reserve. If a loan cures on a given delinquency trip but then becomes delinquent and results in a claim at a later date on a subsequent delinquency trip, the former delinquency trip should not result or tally as a claim. The earlier delinquency trip should get full credit for the cure because, as discussed earlier, mortgage guaranty insurance companies do not reserve on ultimate claims for all insured loans, but only for losses related to loans that are currently delinquent and will not cure before leading to the insurance loss. Once delinquency fates are determined, the empirical conditional probability of claim for each monthly delinquency cohort and each delinquency status can be calculated via aggregation. Tallies are summed by delinquency cohort and risk characteristics and then claim probability is calculated as number of claims divided by number of delinquencies. This process is illustrated later.

Delinquency fates are determined by looking forward in time from each evaluation month to determine the resolution of each delinquency. Table 2 shows Table 1 condensed and tallied for Loan ID 3.

Table 2.

A Record #	B Evaluation Date	C Loan ID	D Status*	E Delq	F Delq Trip	G Cure	H Claim
3	Jan-06	3	90	1	1	0	1
8	Feb-06	3	90	1	1	0	1
13	Mar-06	3	120	1	1	0	1
18	Apr-06	3	FCL	1	1	0	1
23	May-06	3	FCL	1	1	0	1
28	Jun-06	3	CLM	0	1	0	1

* 0 = Current; 30, 60, 90, 120 = days past missed payment; FCL = foreclosure; CLM = claim; columns E-H quantified via binary 0/1

Considering the delinquency cohort as of January 2006 (from record #3), Loan ID 3 is 90 days past due. Loan ID 3 becomes progressively more delinquent until Loan ID 3 results in a claim in June 2006. Loan ID 3 has a single delinquency trip that results in a claim in June 2006 (record #28). Therefore, for the delinquency cohort January 2006, delinquency category 90, Loan ID 3 results in a claim and is tallied as such in column H. Similarly, for delinquency cohort March 2006 (from record #13), delinquency category 120, Loan ID 3 results in a claim and is tallied as such in column H. This does not mean there are multiple claims on Loan ID 3, but rather, it is affiliated with multiple delinquency cohorts.

Alternatively, consider Table 3, condensed from Table 1, which highlights Loan ID 4.

Table 3.

A Record #	B Evaluation Date	C Loan ID	D Status*	E Delq	F Delq Trip	G Cure	H Claim
4	Jan-06	4	60	1	1	1	0
9	Feb-06	4	0	0	NA	NA	NA
14	Mar-06	4	30	1	2	0	1
19	Apr-06	4	60	1	2	0	1
24	May-06	4	FCL	1	2	0	1
29	Jun-06	4	CLM	0	2	0	1

* 0 = Current; 30, 60, 90, 120 = days past missed mortgage payment; FCL = foreclosure; CLM = claim

In delinquency cohort January 2006 (from record #4), Loan ID 4 is 60 days past due on its first delinquency trip and results in a cure. This is because Loan ID 4 becomes current on payments during February 2006 (from record #9). However, for the evaluation months and delinquency cohorts that follow, Loan ID 4 tallies fate as a claim because its resolution from delinquency trip 2 results in a claim. Note that the hindsight delinquency segregation, categorization, and tallying can only occur because there is a contiguous history of delinquency status and evaluation dates. As previously mentioned, in practice, the mortgage guaranty insurance company only needs to reserve for a loan whenever it is delinquent or during any of the monthly cohorts in the tables where Delq = 1 (column E).

For completeness, Table 4 presents all the fate tallies from the dynamic characteristics database presented in Table 1.

Table 4.

A Record #	B Evaluation Date	C Loan ID	D Status*	E Delq	F Delq Trip	G Cure	H Claim
1	Jan-06	1	0	0	NA	NA	NA
2	Jan-06	2	30	1	1	0	0
3	Jan-06	3	90	1	1	0	1
4	Jan-06	4	60	1	1	1	0
5	Jan-06	5	0	0	NA	NA	NA
6	Feb-06	1	0	0	NA	NA	NA
7	Feb-06	2	30	1	1	0	0
8	Feb-06	3	90	1	1	0	1
9	Feb-06	4	0	0	NA	NA	NA
10	Feb-06	5	30	1	1	1	0
11	Mar-06	1	30	1	1	1	0
12	Mar-06	2	30	1	1	0	0
13	Mar-06	3	120	1	1	0	1
14	Mar-06	4	30	1	2	0	1
15	Mar-06	5	60	1	1	1	0
16	Apr-06	1	0	0	NA	NA	NA
17	Apr-06	2	30	1	1	0	0
18	Apr-06	3	FCL	1	1	0	1
19	Apr-06	4	60	1	2	0	1
20	Apr-06	5	30	1	1	1	0
21	May-06	1	30	1	2	1	0
22	May-06	2	30	1	1	0	0
23	May-06	3	FCL	1	1	0	1
24	May-06	4	FCL	1	2	0	1
25	May-06	5	0	0	NA	NA	NA
26	Jun-06	1	0	0	NA	NA	NA
27	Jun-06	2	30	1	1	0	0
28	Jun-06	3	CLM	0	1	0	1
29	Jun-06	4	CLM	0	2	0	1
30	Jun-06	5	0	0	NA	NA	NA

* 0 = Current; 30, 60, 90, 120 = payment days past due; FCL = foreclosure; CLM = claim

In practice, there are not 30 records for five loans to analyze, but potentially millions of records for hundreds of thousands of loans. At the end of 2006, the private mortgage insurance industry had nearly $800 billion of primary insurance in force [7]. This tallying procedure is executed with a programming language that can handle the logic of do loops and consecutive record comparison, so that key ID fields, delinquency statuses, and evaluation dates can be compared and processed. Two examples of programming languages that can accomplish these tasks are C++ and Visual Basic. For each record, tallies depend on what happens in later records for the same certificate number and delinquency trip. Delinquency trip is determined by delinquency status and evaluation date, as was illustrated with Loan ID 4 in Table 3.

Table 5 illustrates the aggregation of tallies and calculation of empirical claim rate for one delinquency cohort, March 2006. Column 3 shows three delinquent loans of status 30. These are Loan IDs 1, 2, and 4 from record numbers 11, 12, and 14. Loan ID 1 results in a cure, for a sum of 1 for cure, status 30, in column D. Loan ID 2 is still delinquent at the end of the time window under consideration. The empirical claim rate can only be calculated based on those loans whose fate, or resolution, is known. Therefore, unresolved loans should be excluded from the calculation. Loan ID 4 results in a claim, for a sum of 1 for claim, status 30, in column E. Column G is calculated as the number of claims for the status divided by the number of resolved delinquencies, or the sum of cures and claims.

Table 5.

A Delinquency Cohort	B Status*	C Delqs	D Cures	E Claims	F = D+E Resolved Delqs	G = E/F Claim Rate on Resolved Delinquencies
Mar-06	30	3	1	1	2	50%
Mar-06	60	1	1	0	1	0%
Mar-06	90	0	0	0	0	NA
Mar-06	120	1	0	1	1	100%
Mar-06	FCL	0	0	0	0	NA

*30, 60, 90, 120=payment days past due; FCL= foreclosure

As a note, it may also be of interest to the reserving actuary to calculate the maximum possible claim rate for a delinquency category. In the previous example, the max claim rate would be 67% (two-thirds). The ratio is calculated by summing every claim plus unresolved delinquencies (assumes all unresolved loans with claim) divided by number of loans in the delinquency cohort (3).

When fates are comprehensively tallied, the loan risk characteristics from the static database can be merged onto each record, such that resolution ratios (i.e., probability of claim versus probability of cure) for each cohort can be calculated along various risk dimensions. The fewer fields within each record to be processed, the more program run performance is optimized; therefore, record-by-record tallying is best done prior to merging the static characteristics. The risk dimensions that can or should be used depend on the robustness of the data and the judgment of the actuary (and are beyond the scope of this discussion). Table 5 shows the most basic risk-dimension calculation based only on delinquency status and not including other characteristics.

Table 6 presents an example of what summarized tallies might look like for a single delinquency cohort aggregated along the risk dimension loan-to-value (the ratio of loan amount to purchase price). In general, the higher the percentage of loan relative to the home's value, the larger the likelihood of default and, similarly, claim. In general, higher loan-to-value ratios result in borrowers with less equity in the property and therefore less to lose in the case of default, versus borrowers with loans that have low loan-to-value. As mentioned previously, in general, the more severely a loan's delinquency status has progressed along the spectrum of delinquency status (i.e., 30, 60, 90+, FCL), the higher likelihood of claim. The authors have observed exceptions to this, but even then, the anti-intuitive empirical result is not significant. Table 6 is similar to Table 5 but with the addition of a second, albeit static, risk characteristic that allows the actuary to analyze the interaction of these two risk characteristics, delinquency status and loan-to-value.

Table 6.

A Status*	B Loan-To-Value	C Delqs	D Cures	E Claims	F = D+E Resolved Delqs	G = E/F Claim Rate
30	90	1000	930	70	1000	7%
	95	1200	1092	108	1200	9%
	100	1400	1232	168	1400	12%
60	90	800	720	80	800	10%
	95	900	792	108	900	12%
	100	1000	860	140	1000	14%
90	90	600	528	72	600	12%
	95	700	595	105	700	15%
	100	800	664	136	800	17%
120	90	300	240	60	300	20%
	95	350	266	84	350	24%
	100	400	288	112	400	28%
FCL	90	100	65	35	100	35%
	95	120	72	48	120	40%
	100	140	77	63	140	45%

*30, 60, 90, 120 = payment days past due; FCL = foreclosure

Claim-rate frequency indications can be calculated using summary statistics of the actuary's choice by using different groupings of delinquency cohorts. From these indications, along with other sources for consideration, the actuary can select frequency factors to be applied to the current, and potentially future, cohort of delinquent loans for loss-reserving purposes.

5. ECONOMIC VARIABLES

As mentioned previously, mortgage guaranty insurance performance is strongly dependent on macroeconomic factors. Macroeconomic factors found to be predictive of mortgage default include home price appreciation, unemployment, and interest rates (this list is not exhaustive). As such, the actuary may choose to include a loss-reserving methodology dependent on forecasted macroeconomic factors such as these.

Depending on the granularity of the modeling approach, the actuary may want to have available selected macroeconomic factors associated with historical mortgage loan defaults, loss given default, and recoveries. Collection of the corresponding macroeconomic variables is relatively easy. Generally, a high-speed Internet connection and time to gather and download the information is all that is required. The first pass at collecting all the historical information may require a fair amount of time up front, but updating the series periodically should be less onerous.

For example, assume the actuary wishes to estimate loss reserves each month and incorporate interest rates, home price appreciation, and unemployment into the loss-reserving process as leading factors.

The actuary may want to estimate loss reserves as a function of forecasted market mortgage interest rates, in addition to the dynamic delinquency status and other static underwriting risk characteristics. One possibility is to collect Freddie Mac's Primary Mortgage Market Survey® (PMMS) as a historical information source for mortgage interest rates. It provides a proxy for market mortgage rates for four mortgage products and also reports for the nation and five geographic regions. According to Freddie Mac, "Freddie Mac's Primary Mortgage Market Survey surveys lenders each week on the rates and points for their most popular 30-year fixed-rate, 15-year fixed-rate, 5/1 hybrid amortizing adjustable-rate, and 1-year amortizing adjustable rate mortgage products." Additionally, "Average rates and points (and margin for ARMs) for each product are reported for the nation and the five Freddie Mac regions."

The actuary can evaluate PMMS historical interest rates as predictors of claim probability, loss given default, and recovery rates. Possible models include logit models for default where the input variables include economic variables such as interest rate, as well as underwriting characteristics and delinquency status. Once a model relating interest rates as a leading indicator to mortgage loan default and mortgage insurance loss is developed, interest rates can be incorporated into the reserving process. Interest rates can be forecast using various interest rate models, or the actuary can rely on readily available deterministic estimates of future mortgage interest rates. Freddie Mac offers mortgage rate forecasts in its weekly "Economic and Housing Market Outlook." The Mortgage Bankers Association (MBA) offers on its website an economic forecast of Treasury interest rates and unemployment in its "MBA Long-term Economic Forecast."

As mentioned earlier, home price appreciation and unemployment are other economic variables that can be collected and tested for significance of estimating loss reserves. Sources for historical home price appreciation data include the Office of Federal Housing Enterprise Oversight (OFHEO) House Price Index, Freddie Mac's Conventional Mortgage Home Price Index (CMHPI) and the S&P/Case-Shiller® Home Price Indices. OFHEO's House Price Index is published quarterly and geographically for the U.S. as a whole, nine U.S. Census divisions, state, and metropolitan statistical area (MSA). Freddie Mac's CMHPI is also provided for the same geographic regions, while the S&P/Case-Shiller® Home Price Indices are only available for 20 large MSAs (and two composites), but broken out monthly. Finally, historical unemployment data can be obtained from the U.S. Department of Labor's Bureau of Labor Statistics monthly and at the state level.

Depending on the granularity of the historical economic data along dimensions of frequency and geography (i.e., monthly versus quarterly or state versus Census division), preparing it for mapping to the preferred reserving methodology may require additional consideration. Conceptually, this tends to be straightforward. For example, using loan-level performance data where each loan record contains a field for property state but historical Freddie Mac mortgage rates provide only geographic regions (where these geographic regions contain multiple states) would require mapping the states to Freddie Mac's geographic regions. In practice, this requires another step in the approach and generally leads to fewer field categories (i.e., 50 states, Washington, D.C., and territories get aggregated into five geographic regions).

Next, merging the collected historical economic data to test its predictive significance on default, loss given default, and recovery will require further effort. The actuary may want to test the historical economic variables with respect to the mortgage-loan performance at various time leads (e.g., one month, one quarter, or one year), and this adds another dimension to the considerations for historical economic data manipulation.

6. CONCLUSIONS

Mortgage guaranty insurance loss reserves are provisions for losses due to insured loans currently delinquent, both reported and unreported. Specifically, there need not be a provision for losses due to loans insured but not delinquent. As a result, the status of whether a loan is delinquent or not is integral to the reserve estimate. The extent of a loan's delinquency has been found to have significance as a predictor of loan default and therefore insured loss. Because of the dynamic nature of each loan's delinquency status over time, the reserving actuary will want a contiguous historical performance data set with enough information to reconstruct the month-by-month status of each insured loan so as to quantify the relationship between delinquency status (dynamic) and other characteristics (generally static but potentially dynamic, such as borrower's current FICO® Score) to ultimate fate and claim loss. The ability to reconstruct this history requires monthly database updating, relational database fields with integrity (i.e., unique ID keys that can be referenced across different data sets) and maintenance without purging.

7. REFERENCES

[1] "Mortgage Guaranty Insurance Model Act" Model #630-1, National Association of Insurance Commissioners, Section 16.

[2] DeFranco, Ralph, "Modeling Residential Mortgage Termination and Severity Using Loan Level Data", Spring 2002, page 74.

[3] "Mortgage Guaranty Insurance Model Act" Model #630-1, National Association of Insurance Commissioners, Section 2.

[4] Dennis, Marshall W. and Robertson, Michael J., Residential Mortgage Lending, Fourth Edition, 1995, page 131.

[5] "Mortgage Guaranty Insurance Model Act" Model #630-1, National Association of Insurance Commissioners, Section 9.

[6] Siegel, Jay, "Moody's Mortgage Metrics: A Model Analysis of Residential Mortgage Pools", April 1, 2003, page 10.

[7] *Inside Mortgage Finance*, Feb. 16, 2007.

Abbreviations and notations

ARM, adjustable rate mortgage
CMHPI, Freddie Mac's Conventional Mortgage Home Price Index
IBNR, incurred but not reported
MBA, Mortgage Banker's Association
MSA, Metropolitan Statistical Area
NAIC, National Association of Insurance Commissioners
OFHEO, Office of Federal Housing Enterprise Oversight
PMMS, Freddie Mac's Primary Mortgage Market Survey

Glossary of Terms

1-year amortizing adjustable rate mortgage, a mortgage with an interest rate that changes annually
5/1 hybrid amortizing adjustable-rate mortgage, a mortgage with an initial five-year fixed-interest rate; thereafter the interest rate begins to adjust on an annual basis
Conveyance, the transfer of property from one person to another
Delinquent, mortgage overdue in payment
Delinquency cohort, group of loans with the same accident month
Delinquency status, categorical classification of a mortgage's overdue payment
Delinquency trip, series of monthly delinquency statuses beginning on a loan's accident month and only ending with a status of cure or claim
Fate, ultimate resolution of delinquent loan
Foreclosure, proceeding in which the financer of a mortgage seeks to regain property
Length of default, time elapsed between evaluation date and accident month

Biographies of the Authors

Tanya Havlicek is an Actuarial Assistant and Statistician at Milliman in Milwaukee, WI. She has a B.S. in mathematics from The Ohio State University and an M.S. in Land Resources from the University of Wisconsin - Madison. She is an expert in SAS and has 15+ years of programming experience from coding in a variety of languages. She works on projects involving loss reserve analysis, reinsurance, government and international markets. Prior to joining Milliman, Tanya was a research assistant at the University of Wisconsin - Madison and developed models to analyze complex interactions in a natural resource management context.

Kyle Mrotek is an Actuary at Milliman in Milwaukee, WI. He has B.B.A. degrees in both Actuarial Science and Finance from the University of Wisconsin - Madison. Kyle is a Fellow of the CAS and a Member of the American Academy of Actuaries. He has performed work for mortgage insurers, mortgage lenders, state housing finance agencies and other government guaranty insurers. Recently, Kyle worked out of Milliman's London office for one year.