# Assessing the Delinquency and Default Risk of Insured and Non-Insured High LTV Mortgages

**July 15, 2011**

## Table of Contents

## Executive Summary

During the recent housing bubble, many borrowers who lacked a 20% down payment used second mortgages (so-called piggyback loans) as a way of avoiding private mortgage insurance on a first lien with a higher than 80% loan-to-value ratio. In a typical "piggyback" transaction, a borrower would take out a first mortgage for 80% of the home's value, a second for 10%, and make a 10% down payment.

First mortgages with a piggyback second were the most prevalent alternative to the use of mortgage insurance over the past decade. At the request of Genworth Financial, Promontory Financial Group conducted an independent study to assess the relative default performance of piggyback and insured loans. For this study, Promontory analyzed the loan-level details on a sample of 5.6 million mortgages originated from 2003 to 2007. The dataset, provided by First American CoreLogic, included several borrower and loan-level characteristics. Serious delinquency was evaluated using a definition corresponding to a loan having ever been 90 or more days past due (or worse) at any given time.

Using this measure, 29.09% of the non-insured, piggyback loans were ever delinquent, compared to 19.44% of insured loans. For the 2007 origination year, the rates were 34.80% and 27.75%, respectively. For each of the provided loan-level variables, insured loans were found to have lower ever delinquent rates. For example, insured loans with a combined LTV of 95 to 100% had a delinquency rate of 21.97%, compared to 33.47% for non-insured, piggyback loans. Similarly, insured loans with FICO scores below 620 had a delinquency rate of 34.56%, well below the 50.05% rate for non-insured loans. Low-doc insured loans had a delinquency rate of 24.70%, compared to 33.67% for non-insured loans.

Because the rich dataset included loan-level, monthly performance indicators, it was possible to study not only the presence of delinquency, but the timing as well. Using a widely known statistical technique known as survival analysis, Promontory assessed the relative performance of insured and non-insured, piggyback loans over time, while simultaneously controlling for loan characteristics that are indicators of the risk of delinquency, including documentation level, loan purpose, owner-occupied status, combined LTV, and FICO score. In its analysis, Promontory also included several time-varying factors including local unemployment rates, market interest rates, and home price indices, all of which helped to significantly explain borrower propensities to default. After controlling for this wide variety of factors, Promontory still found that MI was associated with lower default rates for both fixed rate and adjustable rate first mortgages. Overall, across both fixed and adjustable rate loans, the proportion of non-insured loans surviving to 72 months was .798, compared to .833 for insured loans. Significantly, this difference implies that the baseline cumulative default rate of non-insured loans is 20.98% percent higher than that of insured loans.

Promontory's approach can quantify the extent to which MI serves as a proxy for unobserved aspects of the mortgage underwriting process, which when implemented serve to lower default risk for observed combinations of borrower and loan characteristics. However, the survival analysis regression methodology does not measure the impact that MI-related underwriting may have on adjusting the factors which are controlled for in the study, such as LTV. Any impact that MI may have on mitigating the risk associated with such factors is likely to be embedded in the model covariates, and would not be reflected in the estimated baseline performance differences between insured and non-insured loans.

Questions or comments relating to this study should be directed to C. Erik Larson, PhD, Director, Promontory Financial Group, email: elarson@promontory.com, phone: 202-384-1200.

## 1. Introduction

This study presents the results obtained by Promontory Financial Group in its review and assessment of the performance of mortgage loans originated with a second “piggyback” lien compared to first-lien MI-insured mortgage loans originated in the years 2003 to 2007.

Section 1 begins by illustrating the performance differences though descriptive tabular analysis of severe (ever 90 days-past-due) delinquency rates and through graphical comparison of vintage cumulative delinquency curves. A conclusion from the tabular and vintage curve analysis is that it will be important to control simultaneously for a potentially large number of risk factors, and to do so in a way that is sensitive to the time-varying impact that such factors may have over the life of the mortgage. An appropriate framework by which to control for such effects in a time-sensitive manner will require a relatively sophisticated modeling approach, that of statistical survival analysis.

Section 2 discusses the need to employ survival analysis in order to control for the presence of “censored” observations in the mortgage data. In the present context, censored observations correspond to the measured time-to-default of those accounts which have not defaulted and remain open at the end of a study period. For a censored observation, it is only known that the actual time to default or payoff will exceed the observed value. Since longer-lived accounts are more likely to be censored, analysis based solely on non-censored observations is likely to result in biased statistical estimates. Note that there are two “events” which may end a mortgage account lifetime: the first is default; the second is payoff. Since either of these two events may impact the probability of observing the other, we consider a “competing risks” survival analysis, though we continue to focus on the risk of extreme delinquency (i.e., default).

Section 3 presents the results from estimation from both simple and extended versions of MI-stratified Cox proportional hazards models, estimated by mortgage interest rate type (fixed rate and adjustable rate). Risk factor parameter estimates are generally in line with expectations as to sign. We also compare the implied baseline survival curves from the estimated models to smoothed Kaplan-Meier estimates of the empirical survival function. Our modeling approach allows us to produce separate baseline survival estimates for insured and non-insured (with piggyback) mortgages. These baseline curves have been controlled for the impact of risk factors on performance in a way that cannot accomplished by simple tabular or graphical analysis of empirical data. Overall, our analysis is supporting of the assertion that the historical performance of first lien MI-insured loans has been associated with lower rates of extreme delinquency or default, when compared to non-insured first lien loans accompanied by a piggyback second lien, and when controlling for various risk factors.

Section 4 concludes.

## 2. Mortgage Performance Data

The data obtained by Promontory for this study contain performance information for 5,676,428 individual residential mortgages. The data were provided by Genworth Financial in 2011, who obtained them from First American CoreLogic’s servicing database.

There are a number of reasons why the loans in the Genworth-provided dataset might not mirror those in the population as a whole.

- First, and most importantly, both the current and original Genworth study focus exclusively on loans with <20% down payment (>80% Loan-to-Value), which is only a portion of the first-lien origination market. Loans with LTV in excess of 80% represent approximately 20% of the overall market.
- Second, the CoreLogic database does not cover 100% of the loan market, as not all servicers are CoreLogic customers. Their coverage over the study period is over 60% of loans originated. This fact reduces both the number of piggyback and insured loans in the Genworth dataset, relative to the population. However, the missing servicers during the study period were mainly large diversified national-level players, and there is no reason to think that their omission should have a systematic selectivity bias on the representativeness of mortgage types in our dataset.
- Third, CLTV is not reported on 100% of loans in the CoreLogic dataset. Genworth's definition of a "loan with a piggyback" is a first lien loan with LTV=80 and with reported CLTV >80. This definition serves to reduce the number of piggybacks potentially included in the study, while not reducing insured loans.
- Finally, certain exclusions had already been applied to the dataset before Promontory received it. These included excluding records with missing FICO at origination.

To limit and ensure the comparability of our analysis, Promontory further excluded loans with:

- Missing region;
- Combined loan-to-value (CLTV) greater than 105%;
- Categorization of 'Non Insured, Sold'; and
- A mismatch between the origination date in the dataset and the origination date as calculated from the performance history.

Of the records provided by Genworth, 5,492,097 were used in the benchmarking and vintage curve analysis described below.

## a. Descriptive Statistics

This section presents summary tabular analyses illustrating how insured vs. non-insured (with piggyback) mortgage performance differs with various risk factors that are typically thought to be indicative of borrower or product risk.

Promontory used the performance definition of "ever 90 days past due or worse" (including foreclosure and "real estate owned"), a loan-level variable calculated by Genworth and provided on the analysis dataset. This variable is a measure of severe delinquency and is closely related to the definition of default used by most servicers.

Table 1 presents the lifetime cumulative delinquency rates corresponding to our performance definition (ever 90 days past due or worse). In all years except for 2003, the calculated piggyback delinquency rates are higher than the insured delinquency rates. The overall bad rate on the analysis dataset was 19.44% for insured loans and 29.09% for piggyback loans.

**Table 1: Delinquency Rates by Origination Year**

| Origination Year | 2003 | 2004 | 2005 | 2006 | 2007 | 2003-2007 |
|---|---|---|---|---|---|---|
| Insured | 12.10% | 16.15% | 20.49% | 24.34% | 27.75% | 19.44% |
| Non-Insured with Piggback | 9.40% | 16.18% | 27.47% | 36.73% | 34.80% | 29.09% |

Table 2 illustrates how delinquency rates increase with Combined Loan-to-Value (CLTV). For the insured mortgages, the CLTV value is the same as the LTV of the first lien; for non-insured mortgages, the CLTV represents the combined LTV of both the first and second (piggyback) liens.

**Table 2: Delinquency Rates by CLTV**

| Combined LTV at Origination | 80-85 | 85-90 | 90-95 | 95-100 |
|---|---|---|---|---|
| Insured | 16.14% | 17.29% | 17.57% | 21.97% |
| Non-Insured with Piggback | 30.90% | 29.77% | 21.80% | 33.47% |

As expected, increasing FICO scores are associated with lower delinquency rates, with piggyback loans having higher delinquency rates in all FICO score bands, as documented in Table 3.

**Table 3: Delinquency Rates by FICO Score**

| Origination FICO | 350-619 | 620-659 | 660-699 | 700-719 | 720-739 | 740-759 | 760+ |
|---|---|---|---|---|---|---|---|
| Insured | 34.56% | 24.29% | 18.53% | 15.25% | 12.47% | 9.90% | 7.04% |
| Non-Insured with Piggback | 50.05% | 46.35% | 37.34% | 32.83% | 28.11% | 22.74% | 15.77% |

Table 4 shows little difference in severe delinquency rates between purchase and refinance purposes for insured loans, while non-insured (with piggyback) loans supporting refinance are significantly riskier than loans supporting a new purchase. These patterns run against the traditional thinking that a loan supporting a new purchase is riskier than one supporting a refinance; however one may need to control for other factors to see the expected relationship in these data.

**Table 4: Delinquency by Loan Purpose**

| Loan Purpose | Purchase | Refinance |
|---|---|---|
| **Insured** | 19.76% | 18.66% |
| **Non-Insured with Piggyback** | 26.42% | 38.00% |

Table 5 illustrates that low documentation loans are more risky than full-documentation loans for both insured and non-insured loans.

**Table 5: Delinquency by Documentation Level**

| Documentation Level | Full | Low |
|---|---|---|
| **Insured** | 17.56% | 24.70% |
| **Non-Insured with Piggyback** | 21.07% | 33.67% |

And finally, Table 6 illustrates the dramatically lower delinquency rates for adjustable rate mortgages that are insured, compared to those that are non-insured. The difference is much smaller for fixed rate loans.

**Table 6: Delinquency by Rate Type**

| Rate Type | Fixed Rate | Adjustable Rate |
|---|---|---|
| **Insured** | 19.33% | 22.45% |
| **Non-Insured with Piggyback** | 20.15% | 41.96% |

## b. Vintage Curves

Vintage curves provide powerful summaries of the performance of insured and piggyback loans. To construct our vintage curves, we plot the cumulative monthly severe delinquency rate over time for loans originated in a given year. For each vintage, we present curves for sub-segments of insured and piggyback loans. We segment using origination FICO (<=620 is SubPrime, >620 Prime) and CLTV (less than or equal to 90% and greater than 90%). The early vintages (2003 through 2005) have 72 months of performance. Vintages 2006 and 2007 have 60 and 48 months of performance, respectively. As shown in Figures 1 and 2, below, for the 2007 vintage, piggyback loans have significantly accelerated and higher lifetime cumulative delinquency. Appendix A presents additional curves.

**Figure 1**

Cumulative Bad Rates for 2007 Vintage and CLTV LE90

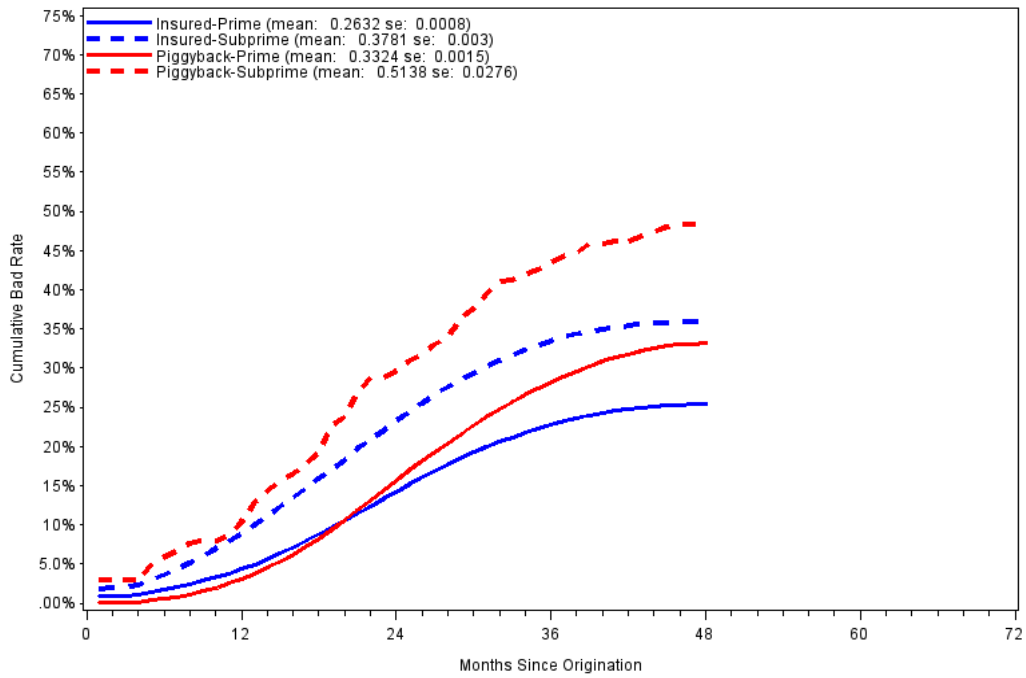
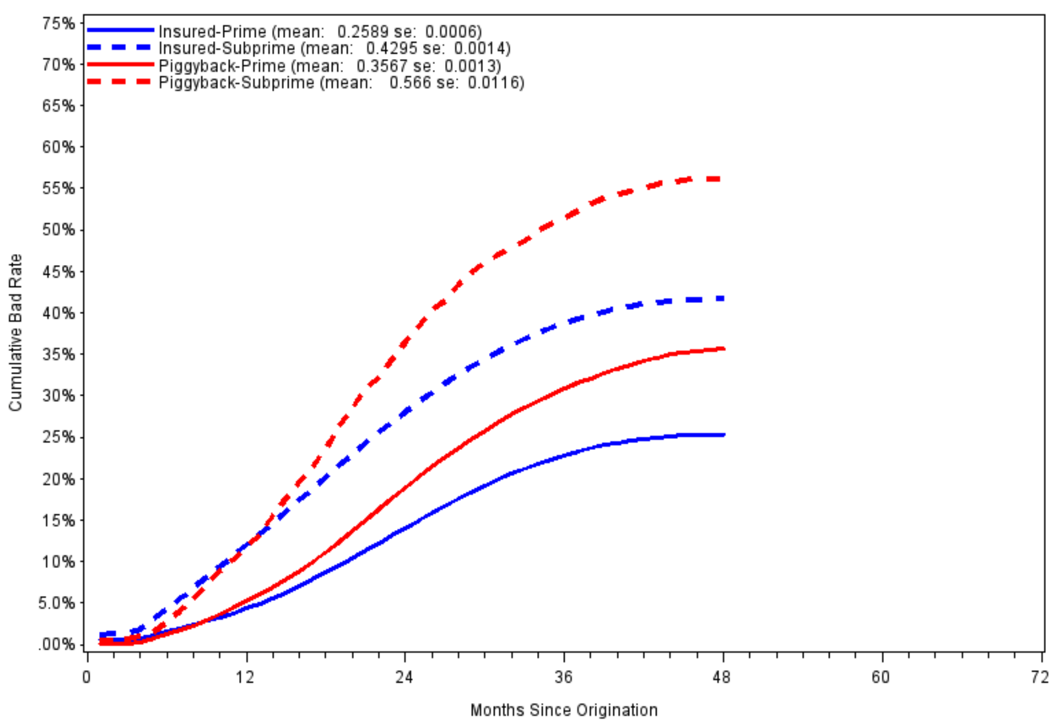



**Figure 2**

Cumulative Bad Rates for 2007 Vintage and CLTV GT90



The tabular analysis and the vintage curve analysis are both strongly suggestive of differing performance characteristics for insured and non-insured (with piggyback) mortgages. However, it is undoubtedly the case that other risk factors, whose level and impact may differ for insured and non-insured (with piggyback) groups, should be controlled for before any conclusions are drawn or stylized facts established.

For instance, while the vintage curves generally illustrate that non-insured loans with piggyback seconds may have cumulative long-term delinquency rates that are higher than their insured counterparts, the vintage curves do at times cross, with insured loan cumulative severe delinquency rates often being greater during the first 12, and in some instances, first 48 months. This occurs even with vintage curves that attempt to control – albeit weakly -- for factors such as origination FICO and CLTV. One potential explanation for this reversal in risk is that differences in payments between the two mortgage types may significantly impact the observed delinquency. In our dataset, and in the population, insured mortgages overwhelmingly have fixed-rate payment structures, while non-insured (with piggyback) mortgages are almost evenly split between fixed- rate and adjustable-rate payment structures. Since initial rate levels of adjustable-rates loans are usually significantly below those carrying a fixed-rate, and because they remain so for months or years before any ARM reset, the initial payments for the fixed rate loans are likely to be significantly higher than the adjustable rate loans. Consequently, it would not be surprising if the higher initial payments of fixed rate mortgages (controlling for CLTV) were associated with an initial higher risk of delinquency for insured, predominantly fixed rate, mortgages.

An obvious takeaway is that it will be important to control simultaneously for a potentially large number of risk factors, and to do so in a way that is sensitive to the time varying impact that such factors may have over the life of the mortgage. Our dataset will allow us to control for such effects, but an appropriate framework in though which to control for such effects in a time-sensitive manner will require a relatively sophisticated modeling approach.

## 3. Survival Models and Analysis

The statistical methods of survival analysis (also called life-table analysis or failure-time analysis) have been developed to analyze the time-to-occurrence of an event as well as the fact of its occurrence. For example, survival analysis has been employed to study the time-to-failure of machine components, time-to-death of patients in a clinical trial, and the duration of unemployment spells of workers.

Introductions to the statistical literature on survival analysis may be found in texts by Kalbfleisch and Prentice (1980), Lawless (1982) and Cox and Oakes (1984). Here, we use survival analysis to model the "lifetimes" of mortgages. Note that there are two "events" which may end a mortgage account lifetime: the first is default, which we have been studying above; the second is payoff. Since either of these two events may impact the probability of observing the other, we consider a "competing risks" survival analysis.

A common feature of survival data is the presence of censored observations. In the present context, censored observations correspond to the measured time-to-default of those accounts which have not defaulted and remain open at the end of a study period. For a censored observation, it is only known that the actual time to default or payoff will exceed the observed value. The study of survival data typically employs information from both censored and non-censored observations. Since longer-lived accounts are more likely to be censored, survival analysis based solely on non-censored observations is likely to result in biased statistical estimates. Indeed, simple regression analysis of account bad-rates which fails to take account for the impact of censoring is likely to produce biased estimates of the explanatory variables if the censoring is not random or if the mixture of effects is not distributed randomly across censored and uncensored accounts.

## a. Survival and Related Functions

Suppose the population under study consists of mortgage lifetimes for N relatively homogeneous accounts. Each lifetime in the population can be represented by a random variable, $T_i$, where i=1,...,N. If n account lifetimes are to be randomly sampled from the target population, each account will have a potential censoring time (or censoring age) $a_i$ (i=1,...,n). The potential censoring time is determined using the opening date for the account and the closing date for the period during which observations are collected. The sample data consists of n pairs ($c_i$,$s_i$), where $s_i=\min(T_i,a_i)$ is the observed lifetime of account i, and $c_i$ is an indicator variable taking the values $c_i=1$ if $T_i \leq a_i$ ($s_i$ is an uncensored observation) and $c_i=0$ if $T_i > a_i$ ($s_i$ is a censored observation).

For the moment, ignore the possibility of censoring. Distributional characteristics of a population of random account lifetimes $T_i$ are summarized by a distribution function, F(t), and survival function, S(t), here defined as

$$F(t) = 1 - S(t) = Probability(T_i < t).$$

F(t) and S(t) are both defined for $0<t<\infty$. Using statistical survival analysis, one can use sample data to make reliable inferences about these population functions.

Note that F(t) reports the proportion of accounts in the population with lifetimes less than t, while S(t), reports the proportion of accounts with lifetimes greater than or equal to t. Also, as t increases from zero, F(t) monotonically increases from zero toward one, while S(t) monotonically decreases from one toward zero.

Closely related to the distribution function, F(t), is the density function, f(t). When t is measured in continuous units, f(t) is defined by

$$f(t) = \partial F(t)/\partial t.$$

The density function can be thought of as the instantaneous probability of the account lifetime ending at t.

The hazard function or age-specific failure rate function, h(t), is related to the distribution, survival and density functions. The hazard function is defined by

$$h(t) = f(t)/S(t).$$

The hazard, h(t), may be interpreted as the “instantaneous” conditional probability that an account will close at age t, given that it has remained open to at least age t. Hazard functions are particularly useful in the analysis of account lifetimes, since they specify the risk of immediate closure of an open account at age t. The choice of an appropriate statistical model for account lifetimes is aided by the careful study of empirical hazard functions constructed from sample data.

The distribution, survival, density and hazard functions are mathematically equivalent representations of the distributional characteristics of a population of account lifetimes, since each one of them can be derived given any of the others.

## b. Cox Proportional Hazard Models

As part of this study, Promontory estimated a Cox Proportional Hazard (PH) Model to investigate and quantify the relative performance of piggyback and insured loans while controlling for loan-level factors that are commonly thought to be important in describing loan performance. The Cox Proportional Hazard Model is originally due to David Cox (1972). The model has been extended significantly by others (see Therneau and Grambsch (2000)), and has received widespread empirical application. The model is usually written as

$$h_i\,(t) = \lambda_0(t)\ Exp(\beta_1 X_{i1t} + \beta_2 X_{i2t} + \dots + \beta_k\, X_{ikt}).$$

This model specifies that the hazard rate for individual "i" at time "t" is made up from the product of two components: a non-negative "baseline" hazard function $\lambda_0(t)$, and an individual-specific proportionality factor $Exp(\beta_1 X_{i1t} + \beta_2 X_{i2t} + \dots + \beta_k X_{ikt})$, where $X_{i1t}, X_{i2t}, \dots, X_{ikt}$ are the values of the observed , possibly time-varying, covariates (hence the indexing of the individual covariates by t.[1]) The corresponding covariate coefficients, $\beta_1, \beta_2, \dots \beta_k$, are unknown parameters which have to be estimated from the data.

Taking natural logs, the model is also written as:

$$\log h_i\,(t) = \alpha_0(t) + \beta_1 X_{i1t} + \beta_2 X_{i2t} + \dots + \beta_k\, X_{ikt},$$

The Proportional Hazards Model gets its name from the fact that the ratio of hazards for any two individuals is given by the ratio or their proportionality factors. However, there is sometimes a reason to believe that the proportionality assumption underlying the Cox specification might not be warranted, and that it is appropriate to consider extensions of the model for non-proportional hazards. One such extension is through "stratification."

In a stratified model, there is a presumption that the hazards of two (or more) groups of individuals may be written as

$$\log h_i\,(t) = \alpha_1(t) + \beta_1 X_{i1t} + \beta_2 X_{i2t} + \dots + \beta_k\, X_{ikt},\ \textit{for individuals i that are members of group 1, and}$$

$$\log h_j\,(t) = \alpha_2(t) + \beta_1 X_{j1t} + \beta_2 X_{j2t} + \dots + \beta_k\, X_{jkt},\ \textit{for individuals j that are members of group 2.}$$

These two specifications can be combined into a single specification for both groups by writing

$$\log h_i\,(t) = \alpha_c(t) + \beta_1 X_{i1t} + \beta_2 X_{i2t} + \dots + \beta_k\, X_{ikt},\ \textit{where}\ \alpha_c(t) = \alpha_1(t) D_{i1} + \alpha_2(t) D_{i2}$$

[1] In order to incorporate time-varying covariates, we utilize a representation of the survival model as a counting process; see Hosmer and Lemeshow (1999), Appendix 2.

where $D_{i1}$ and $D_{i2}$ are zero-one indicator functions identifying an individual's membership in group 1 or 2.

In order to estimate the Cox PH model, methods of partial likelihood maximization are employed (which allows one to avoid specifying the baseline hazard function.)[2] In the case of a stratified model, partial likelihood estimation requires a slightly more complex estimation procedure. Separate partial likelihoods functions are first constructed for each stratification group; these functions are then multiplied together to form an aggregate partial likelihood model that is maximized though numerical estimation of the coefficient vector **β**.

# 4. Estimation

## a. The Survival Analysis Modeling Dataset

Due to the size of the Genworth dataset and the computational demands in terms of memory and time required to estimate the partial likelihood algorithms for the alternative survival models, particularly in the presence of time-varying covariates , Promontory did not find it feasible to estimate the stratified proportional hazard models with the full dataset that had been provided by Genworth. Instead, we have utilized a 10% randomly selected subsample for use as a modeling dataset.[3] This dataset is still very large, containing 538,500 mortgage lifetimes. Summary information is given in the following table.

**Table 7: Counts and Dispositions of Observations in the Modeling Dataset**

| Rate Type | Type | Default | Paid Off | Paying | Total by Rate Type |
|---|---|---|---|---|---|
| All Rate Types | Insured | 83,641 | 144,807 | 203,240 | 538,500 |
| | Non-insured w/ Piggyback | 31,198 | 33,323 | 42,291 | |
| Fixed Rate | Insured | 73,764 | 126,260 | 188,923 | 452,026 |
| | Non-insured w/ Piggyback | 12,774 | 21,275 | 29,030 | |
| Adjustable Rate | Insured | 9,877 | 18,547 | 14,317 | 86,474 |
| | Non-insured w/ Piggyback | 18,424 | 12,048 | 13,261 | |

Appendix B contains additional summary information on loans characteristics in the modeling dataset.

## b. Results

### *Estimation of Nonparametric (Empirical) Survival Curves*

Rather than proceeding directly to the estimation of a stratified proportional hazards model, it will be useful to first consider the empirical survival distribution curves for default that are implied by the sample data. To this end, we have constructed smoothed estimates of the empirical survival function using the method of Kaplan and Meier (1958.) Figures 3 and 4 show the empirical, or non-parametric, estimated default survival curves for insured and non-insured (with piggyback) mortgage loans, computed for subsamples defined by whether the loans were of fixed rate or adjustable rate type.

---

[2] Estimation of Cox Proportional Hazards and other survival models is discussed in Kiefer (1988).

[3] Promontory has obtained similar results with alternative randomly selected samples of a similar size.

These curves, as do all the estimates presented in this section, focus exclusively on the risk of default, and treat the competing risk of payoff as a censoring event. This approach is a conventional and meaningful way to present results for a risk of interest (here, default) when competing risks are present.

**Figure 3. Empirical Survival Curve Estimate, Fixed Rate Loans**

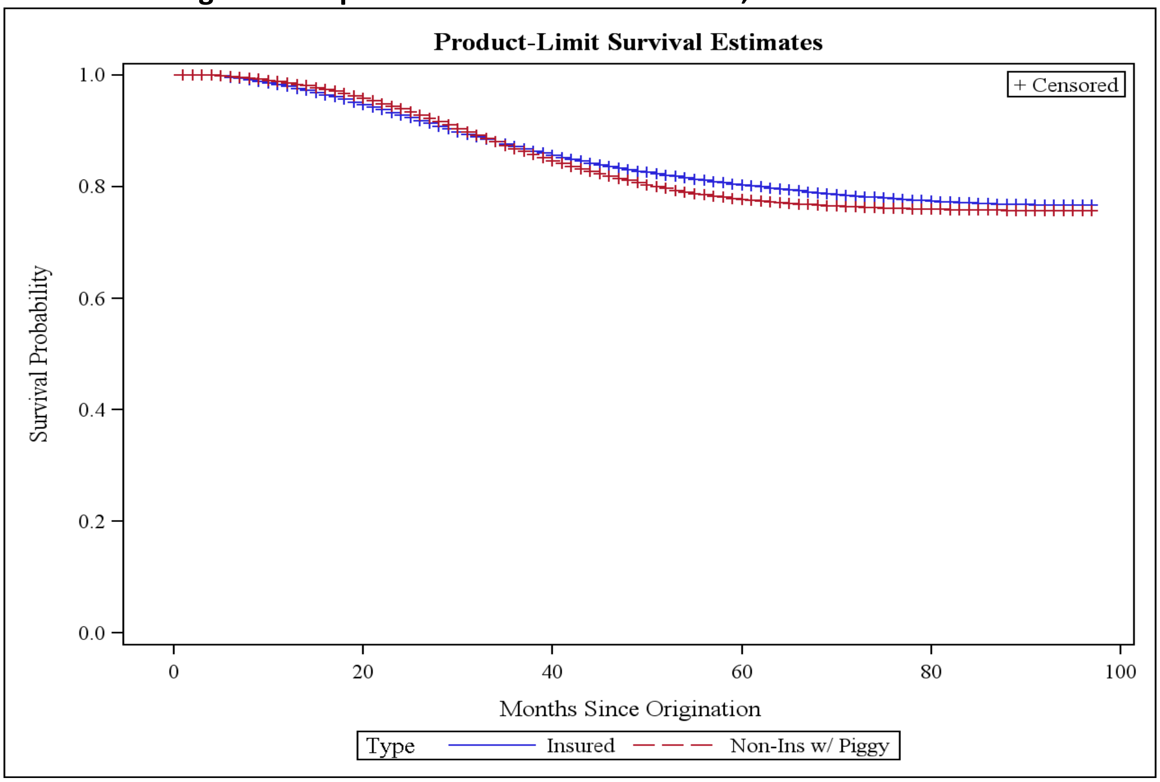
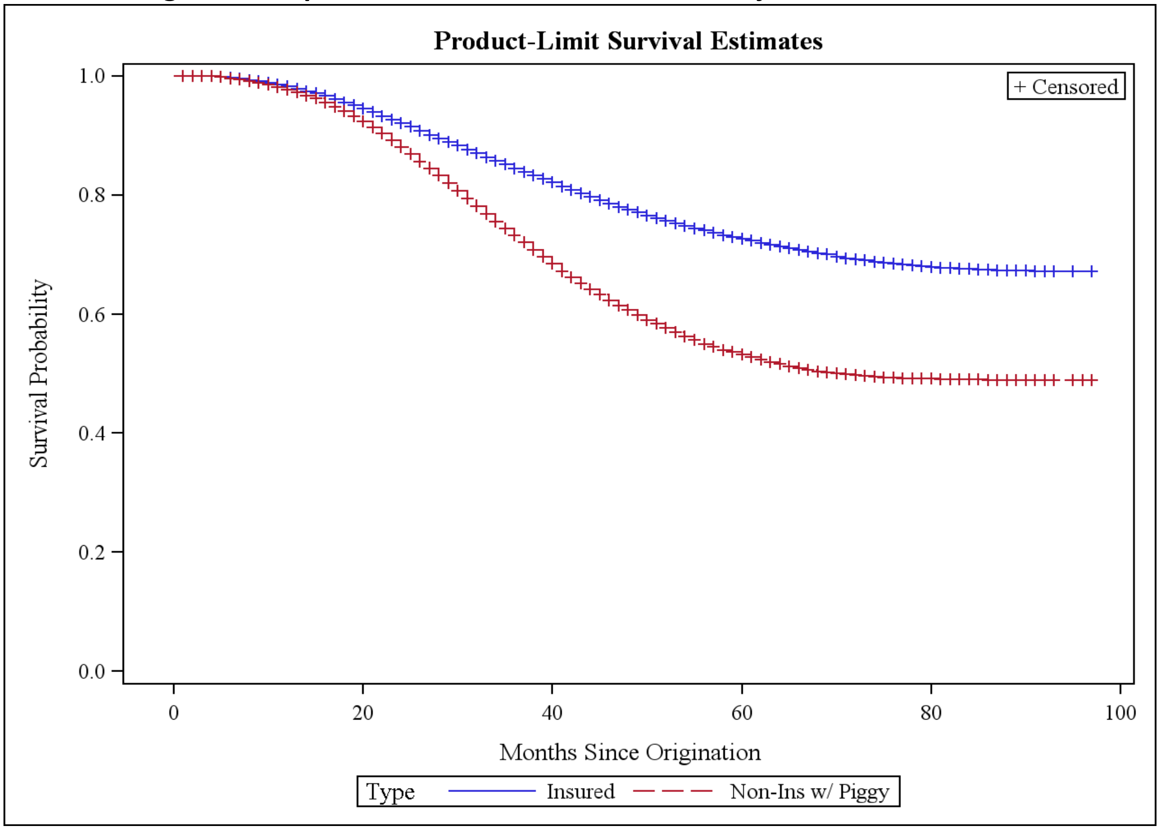


**Figure 4. Empirical Survival Curve Estimate, Adjustable Rate Loans**




Note that even in the empirical survival curves, the long-term higher default risk associated with non-insured loans having piggyback second liens is easy to identify. This is particularly true for the adjustable rate loans, where the survival proportion for the uninsured mortgages ultimately drops well below that of the insured loans.

### *Estimation of a Stratified Proportional Hazards Model*

We are now ready to turn to the estimation of the stratified Cox proportional hazards model. As suggested earlier, we have chosen to specify a model in which we include additional covariates and in which we estimate separate stratified models for subsets of our sample, with loans grouped by rate type. Part of the rationale for estimating different models for different rate types (fixed vs. adjustable) is that borrower behavior in response to changes in economic conditions is likely to be very different across these products. Furthermore, differences in mortgage product types or borrower underwriting practices may exist that are unobservable in our data, but which may result in different magnitudes of the estimated covariate coefficients or in different baseline hazard and survival estimates.

### Covariates

The covariates in our model include several zero-one categorical (or dummy) variables. For each of these variables, a case that has one of the characteristics is coded as a one, and cases without the characteristic are coded as a zero. These variables include the following

- Documentation level (low or full documentation, with full documentation = 1);
- Loan purpose (purchase or refinance, with purchase = 1), and
- Occupancy status (Owner-occupied or not, with owner-occupied = 1).

The model also includes four continuous variables measured at the time of loan origination:

- Combined Loan-to-Value;
- FICO score at origination;
- Original Interest Rate, and
- Original Payment, a constructed variable equal to Original Loan Balance X Initial Interest Rate.

Finally, the model includes four time-varying covariates:

- Interest Rate Differential( t) = Original Interest Rate - Market Interest Rate(t)
- Change in Payment(t) = [Original Interest Rate - Market Interest Rate(t) ] x Original Balance
- Change in Value(t) = (Original Value) x [%Change in Case-Shiller Index(t)], and
- Unemployment Rate(t)

The seasonally adjusted civilian unemployment rate and Case-Shiller Index data were matched to each loan based upon MSA/CBSA if available; otherwise a state or national level measure was used, respectively. The market interest rate data was obtained from Freddie Mac, and it was matched based upon the rate type of the loan. Fixed rate loans were matched to the monthly average of the average weekly 30-year rate; adjustable rate loans were matched to the monthly average of the average weekly 1-year rate.

## Parameter Estimates

Table 8 presents estimation results for the fixed rate and adjustable rate loan group models. Recall that each estimated rate type model has been stratified across insured and non-insured mortgage classes. As a result, we have two sets of parameter estimates, with a given parameter set applying equally to both strata within a given rate group.

The estimated coefficients have signs that are consistent with expectations (recall that due to the proportional hazard specification, a positive parameter indicates that the hazard of default is increasing with the covariate value).

Table 8: Cox Stratified Proportional Hazards Model Parameter Estimates

| Loan Type | Fixed Rate | Adjustable Rate |
|---|---|---|
| Documentation Level (1=Low) | 0.37310 | 0.76391 |
| Loan Purpose (1=Purchase) | -0.05802 | -0.22628 |
| Occupancy Status (1=Owner-Occupied) | -0.14402 | -0.38135 |
| Combined LTV at Origination | 0.02400 | 0.03127 |
| FICO Score at Origination | -0.00880 | -0.00589 |
| Original Interest Rate | 0.21298 | -0.12347 |
| Original Payment (Original Int. Rate*Original Balance) | -0.00478 | 0.01213 |
| Rate Differential (Original Int. Rate - Market Int. Rate) | 0.15648 | 0.09901 |
| Change in Payment (Original Int. Rate - Market Int. Rate)*Original Balance | 0.04650 | -0.00108** |
| Change in Value (Original Value)*(%Change in Case Shiller Index) | 0.04439 | 0.02643 |
| Unemployment Rate | 0.16021 | 0.18988 |

*Note: **Estimate not significantly different from zero. All other estimates are significant at the 0.0001 level.*

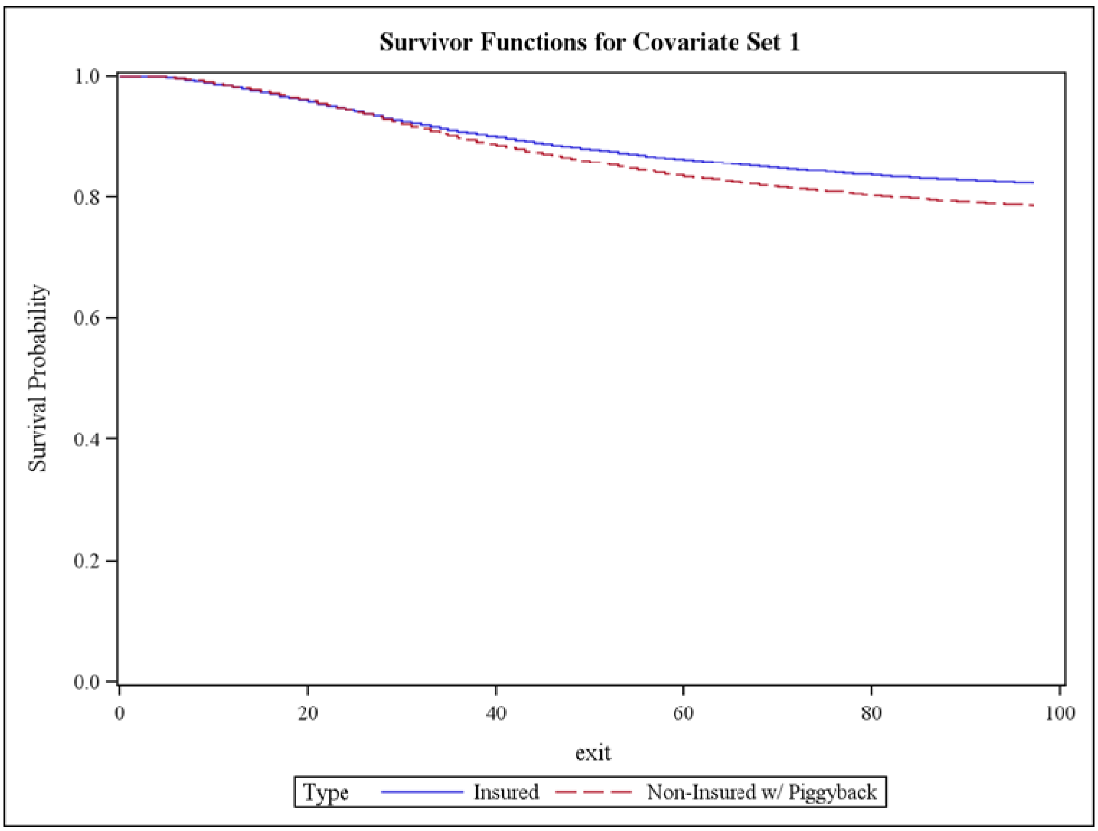
Low documentation, non owner-occupied, high CLTV, and low FICO loans are of greater default risk than loans with the opposite characteristics. Somewhat surprisingly, loans supporting refinancing are of greater risk than loans supporting a new purchase – a result seen in the simple descriptive statistics for this period. The coefficients on the time varying covariates measuring the rate differential between original and current market rates, the change in payment and the change in value are also positive. The greater the difference between the original interest rate and the current market rate, or the greater the different between the original home value and the current implied market value (i.e., the absolute value of potential equity loss), the greater the default risk. Similarly, the higher the current level of unemployment in the MSA or state when the property is located, the higher the default risk. All these impacts are similar across both fixed rate and adjustable rate mortgage groups.

In contrast, when we consider the impact of the level of the original interest rate or the level of the original payment, the signs of the coefficient estimates are reversed between fixed and adjustable rate groups. However, the sign differences make sense: for fixed rate loans, holding original balance constant, higher original interest rates mean higher fixed payments and higher default risk. For adjustable rate loans, the higher original rate probably implies that the risk of a payment shock when the original rate adjusts to market rates is lowered, along with default risk.

## Baseline Survival Curve Estimates

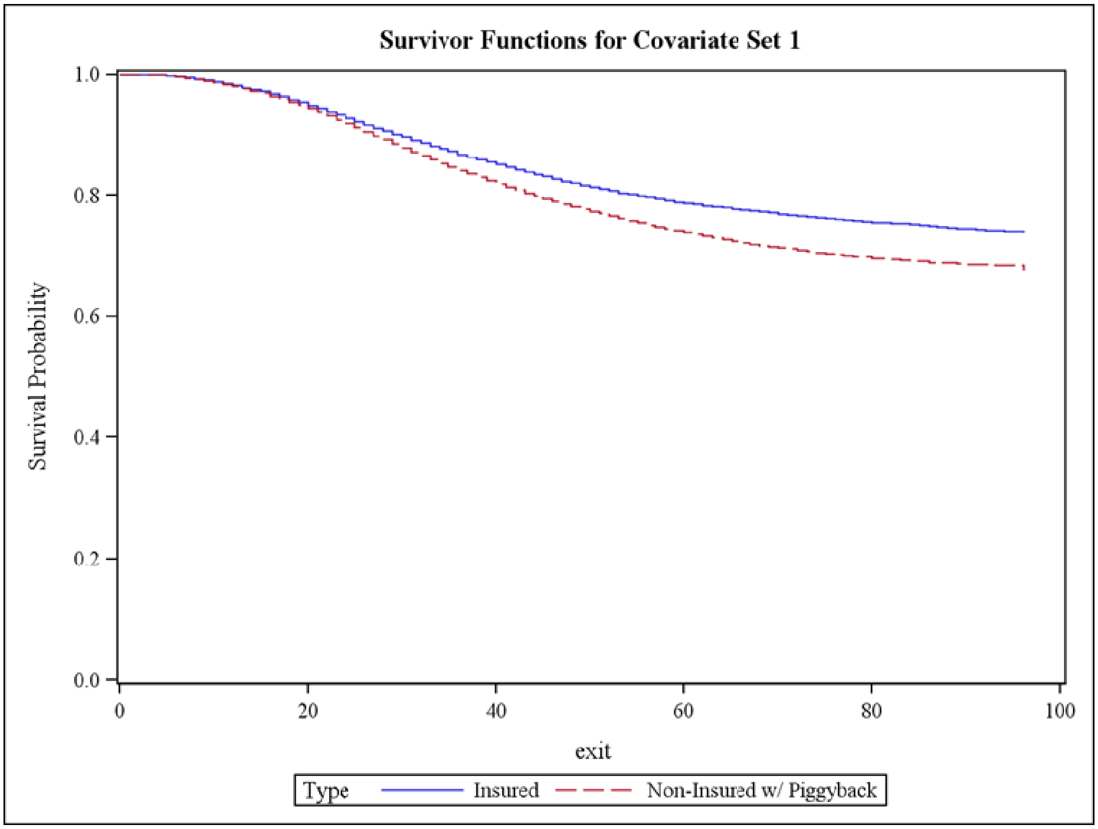
To illustrate the differences between insured and non-insured loans, it is useful to compare the implied baseline survivor functions for the strata corresponding to our estimated set of models[4]. Figures 4 and 5 shows the implied baseline survival curves resulting from our stratified Cox PH model; estimates reflect the survival probability at month t, evaluated at the mean value covariates across the sample population. Effectively, these baseline survival curve estimates illustrate the fundamental differences in performance between insured and non-insured loan groups, controlling simultaneously and equally for all the effects we have been able to attribute to covariates.

**Figure 5. Parametric Baseline Survival Curve Estimates, Fixed Rate Loans**




[4] The baseline hazards and survival functions are estimated as arbitrary functions of time through implementation of a restricted maximum likelihood estimation of the $\alpha_c(t)$ function, in which the covariates for explanatory variables are restricted to their previously estimated values.

Figure 6. Parametric Baseline Survival Curve Estimates, Adjustable Rate Loans




In these curves, the higher default risk associated with the non-insured (with piggyback) loans is very clear – at times even more so than in the empirical survival curves (which did not control for the effect of covariates). For both fixed rate and adjustable rate mortgages, controlling for the impact of covariates results in implied baseline (strata specific) survival curve estimates in which insured loans continue to demonstrate lower extreme delinquency and default risk than non-insured (with piggyback) loans.

Tables 9 and 10 respectively present the estimated numerical baseline survival rates and cumulative default rates, by strata, for selected months-since-origination. Overall, across both fixed and adjustable rate loans, the proportion of non-insured loans surviving to 72 months was .798, compared to .833 for insured loans. Significantly, as shown in Table 10, this difference implies that the baseline cumulative default rate of non-insured loans is 20.98% percent higher than that of insured loans.

## Table 9. Estimated Baseline Survival Rates, S(t)

**Proportion Surviving to Selected Months**

| Rate Type | Type | Months 12 | 24 | 36 | 48 | 60 | 72 |
|---|---|---|---|---|---|---|---|
| All | Insured | 0.983 | 0.943 | 0.903 | 0.873 | 0.851 | 0.833 |
| | Non-Insured w/ Piggyback | 0.983 | 0.942 | 0.890 | 0.851 | 0.820 | 0.798 |
| | Percent Difference (Non-Insured relative to Insured ) | *0.04%* | *-0.13%* | *-1.44%* | *-2.52%* | *-3.65%* | *-4.20%* |
| Fixed Rate | Insured | 0.983 | 0.946 | 0.910 | 0.884 | 0.863 | 0.846 |
| | Non-Insured w/ Piggyback | 0.983 | 0.946 | 0.900 | 0.865 | 0.835 | 0.815 |
| | Percent Difference (Non-Insured relative to Insured ) | *0.08%* | *0.04%* | *-1.13%* | *-2.15%* | *-3.22%* | *-3.66%* |
| Adj. Rate | Insured | 0.983 | 0.930 | 0.869 | 0.820 | 0.788 | 0.767 |
| | Non-Insured w/ Piggyback | 0.981 | 0.920 | 0.841 | 0.782 | 0.740 | 0.710 |
| | Percent Difference (Non-Insured relative to Insured ) | *-0.19%* | *-0.99%* | *-3.16%* | *-4.62%* | *-6.10%* | *-7.32%* |

## Table 10: Estimated Baseline Cumulative Default Rates, F(t)

**Cumulative Proportion Defaulting by Selected Months**

| Rate Type | Type | Months 12 | 24 | 36 | 48 | 60 | 72 |
|---|---|---|---|---|---|---|---|
| All | Insured | 0.017 | 0.057 | 0.097 | 0.127 | 0.149 | 0.167 |
| | Non-Insured w/ Piggyback | 0.017 | 0.058 | 0.110 | 0.149 | 0.180 | 0.202 |
| | Percent Difference (Non-Insured relative to Insured ) | *-2.15%* | *2.09%* | *13.47%* | *17.40%* | *20.79%* | *20.98%* |
| Fixed Rate | Insured | 0.017 | 0.054 | 0.090 | 0.116 | 0.137 | 0.154 |
| | Non-Insured w/ Piggyback | 0.017 | 0.054 | 0.100 | 0.135 | 0.165 | 0.185 |
| | Percent Difference (Non-Insured relative to Insured ) | *-4.60%* | *-0.65%* | *11.38%* | *16.32%* | *20.23%* | *20.10%* |
| Adj. Rate | Insured | 0.017 | 0.070 | 0.131 | 0.180 | 0.212 | 0.233 |
| | Non-Insured w/ Piggyback | 0.019 | 0.080 | 0.159 | 0.218 | 0.260 | 0.290 |
| | Percent Difference (Non-Insured relative to Insured ) | *10.78%* | *13.11%* | *20.99%* | *21.08%* | *22.66%* | *24.02%* |

## c. Diagnostics: Evaluating the Proportional Hazards Assumption

The assumption of the proportional relationship between hazards and covariates that is implied by the Cox model specification should be subjected to an empirical assessment. To perform such an assessment, it is increasingly common to construct residuals along the lines proposed by Schoenfeld (1982). Instead of a single residual for each individual observation, Schoenfeld's method results in constructing separate residuals for each covariate, for each individual loan, using only those loans that defaulted (were not censored.)

Since the Schoenfeld residuals are, in principle, independent of time, a plot that shows a non-random pattern against time is evidence of violation of the proportional hazards assumption. Appendix C provides plots of the estimated, scaled Schoenfeld Residuals against rank time. The minimal departures from a general, random zero-slope pattern vs. time provide reasonable support for the proportional hazards specification used in our analysis.

## 5. Conclusions

The analysis conducted by Promontory generally confirms the results presented in Genworth's 2010 study, and shows that, controlling for various factors, mortgages with piggyback second lien loans have historically experienced higher lifetime rates of severe delinquency than insured mortgages. This conclusion is supported by tabular analysis, graphical vintage curve analysis and by the results from conducting an analysis using statistical methods of survival analysis.

We present the results from estimation from both simple and extended versions of stratified Cox proportional hazards models, the latter estimated across and by US census region. Risk factor parameter estimates are generally in line with expectations as to sign, although variability in the magnitude of estimates exists across regions. We also compare the implied baseline survival curves from the estimated models to smoothed Kaplan-Meier estimates of the empirical survival function. Our modeling approach allows us to produce separate baseline survival estimates for insured and non-insured (with piggyback) mortgages. These baseline curves have been controlled for the impact of risk factors on performance in a way that cannot accomplished by simple tabular or graphical analysis of empirical data

Overall, our analysis supports the assertion that the historical performance of first lien MI-insured loans has been associated with lower rates of extreme delinquency or default, when compared to non-insured first lien loans accompanied by a piggyback second lien, and when controlling for various risk factors.

In closing, it is important to note that the stratified survival analysis regression methodology we deploy does not measure the impact that MI-related underwriting may have on adjusting the factors which are controlled for in the study, such as LTV. Any impact that MI may have on mitigating the risk associated with such factors is likely to be embedded in the model covariates, and would not be reflected in our estimated baseline performance differences between insured and non-insured loans.

The above point should serve to emphasize the importance of the multi-pronged approach that we have taken to consider the impact of MI, and should stimulate further research on this important issue.

## References

Cox ,D.R. (1972) "Regression Models and Life Tables." *Journal of the Royal Statistical Society*, Series B, #34. pp 187-220.

Cox, D.R., D. Oakes (1984), *Analysis of Survival Data*, London, UK: Chapman and Hall.

Hosmer, D.W., Jr.; Lemeshow, S. (1999) *Applied Survival Analysis: Regression Modeling of Time to Event Data*, New York, NY: John Wiley & Sons.

Kalbfleisch, J.D.; Prentice, R.L. (1980), *The Statistical Analysis of Failure Time Data*, New York, NY: John Wiley & Sons.

Kaplan, E.L. ; Meier, P. (1958) "Nonparametric Estimation from Incomplete Observations." *Journal of the American Statistical Association*, 53, pp. 457-481.

Kiefer, N.M. (1988) "Economic Duration Data and Hazard Functions." *Journal of Economic Literature*, 26, pp. 646-679.

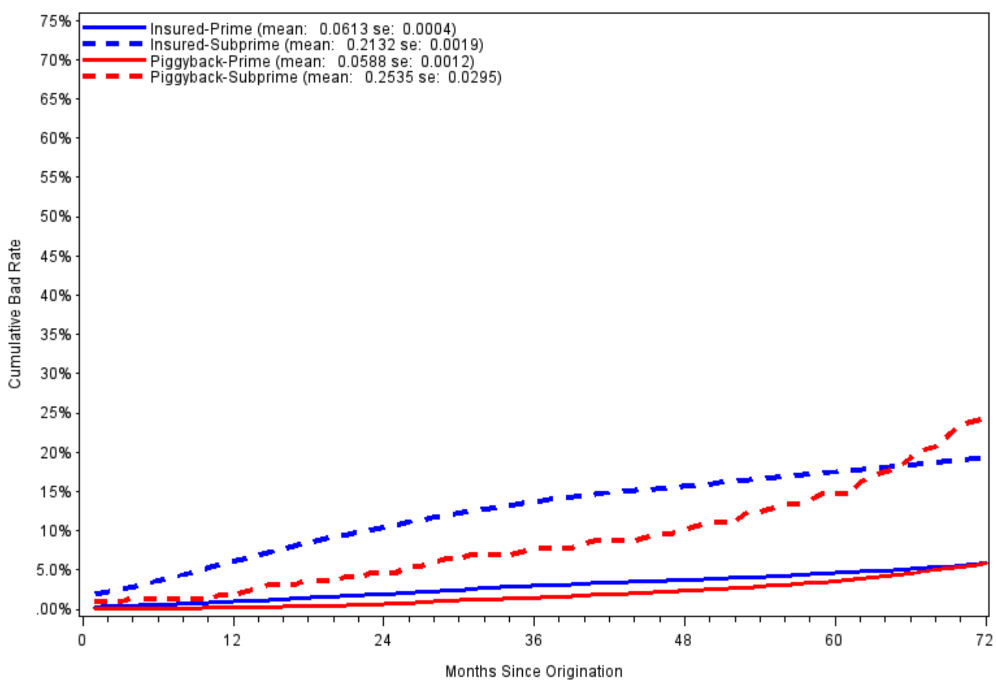
Lawless, J.E. (1982), *Statistical Models and Methods for Lifetime Data*, New York, NY: John Wiley & Sons.

Schoenfeld, D. (1982) "Partial residuals for the proportional hazards regression model." *Biometrika,* 69, pp. 239-241.

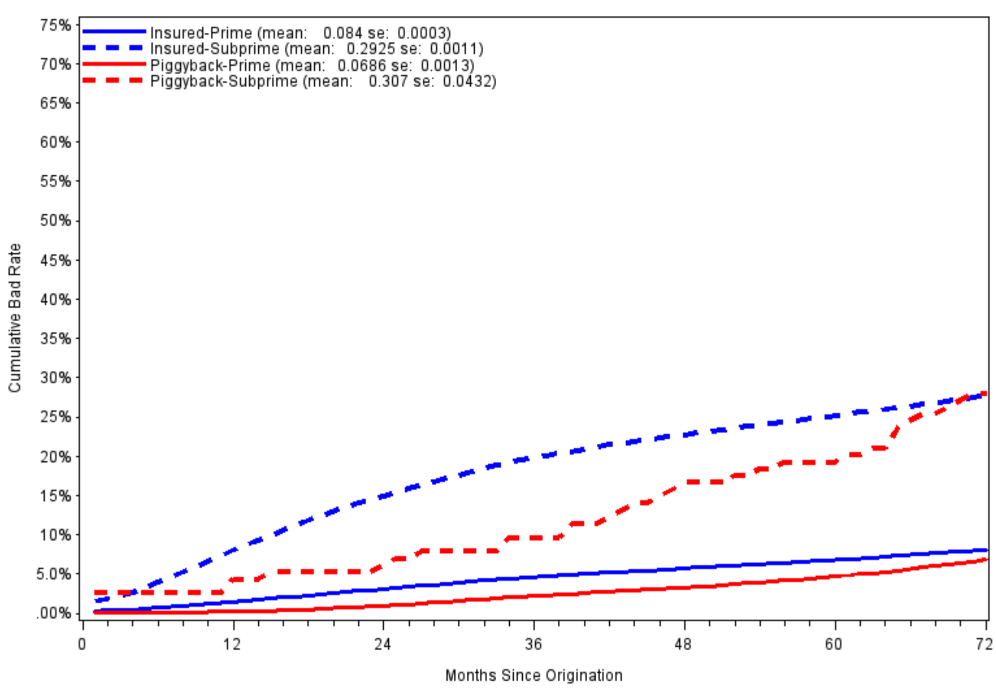
Therneau, T. M.; Grambsch, P. M. (2000) Modeling Survival Data: Extending the Cox Model. New York: Springer-Verlag.

## Appendix A: Vintage Curves

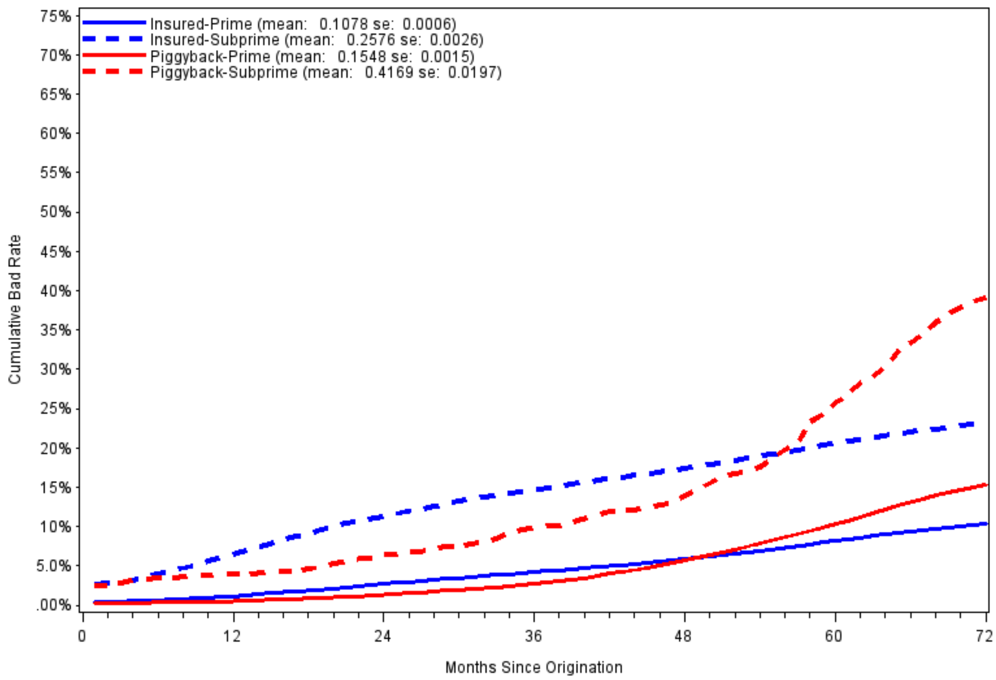
Cumulative Bad Rates for 2003 Vintage and CLTV LE90



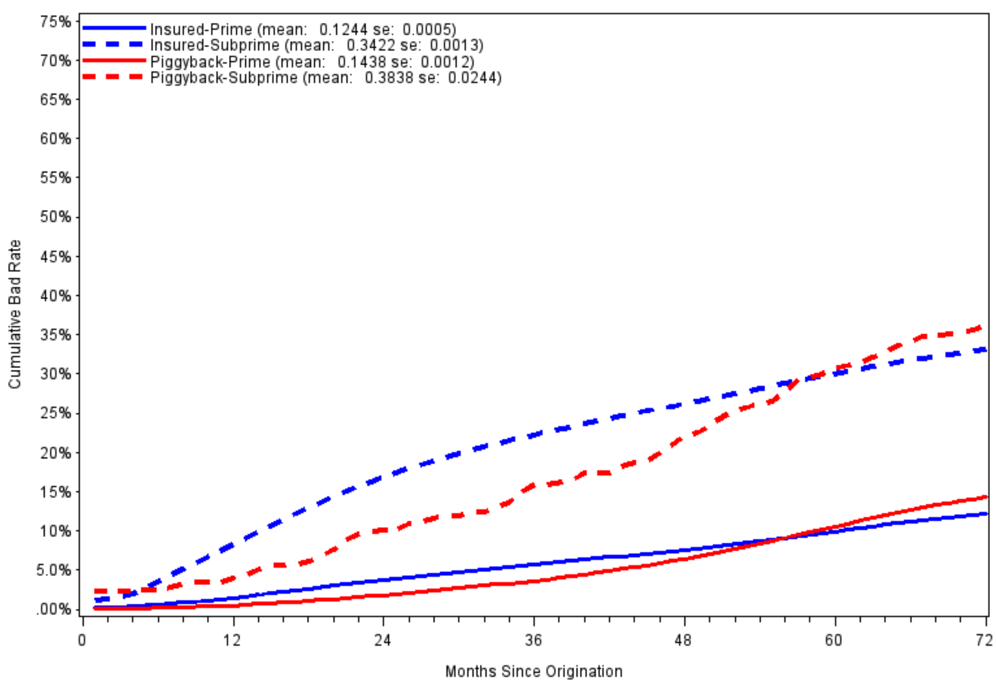

Cumulative Bad Rates for 2003 Vintage and CLTV GT90

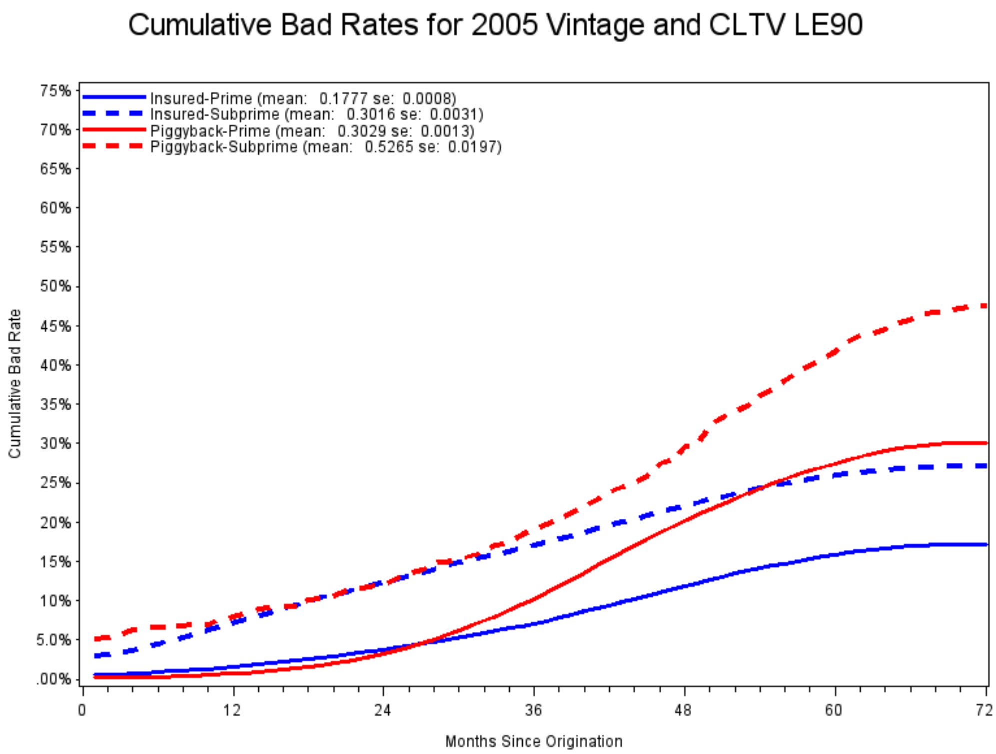


## Cumulative Bad Rates for 2004 Vintage and CLTV LE90

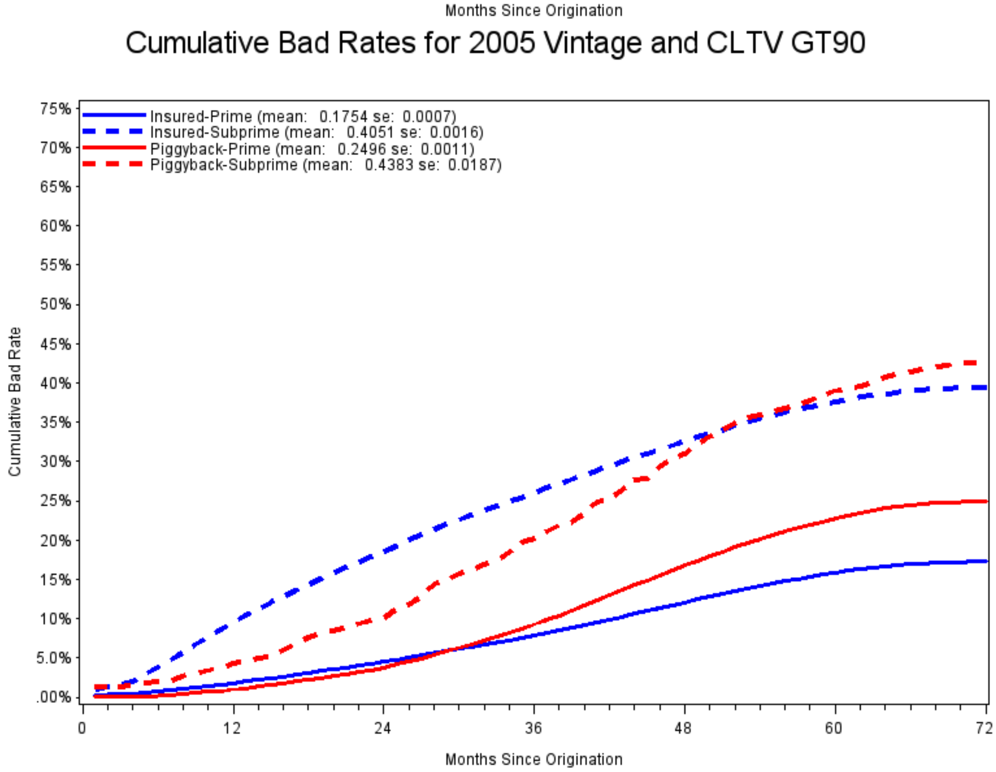



## Cumulative Bad Rates for 2004 Vintage and CLTV GT90



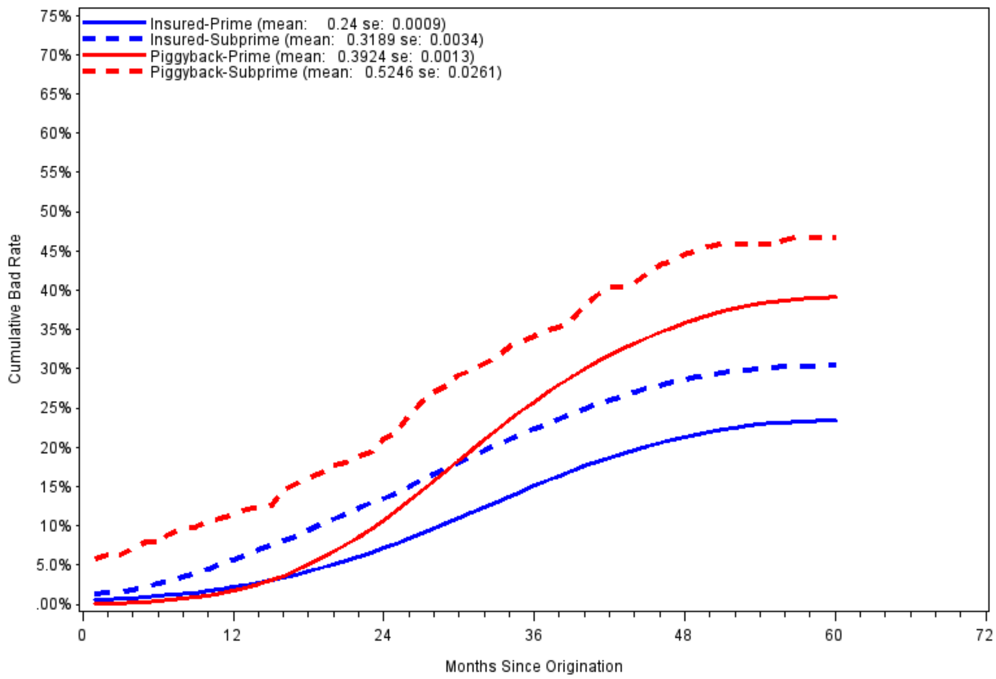


Cumulative Bad Rates for 2005 Vintage and CLTV LE90
Insured-Prime (mean: 0.1777 se: 0.0008)
Insured-Subprime (mean: 0.3016 se: 0.0031)
Piggyback-Prime (mean: 0.3029 se: 0.0013)
Piggyback-Subprime (mean: 0.5265 se: 0.0197)
Cumulative Bad Rate
Months Since Origination

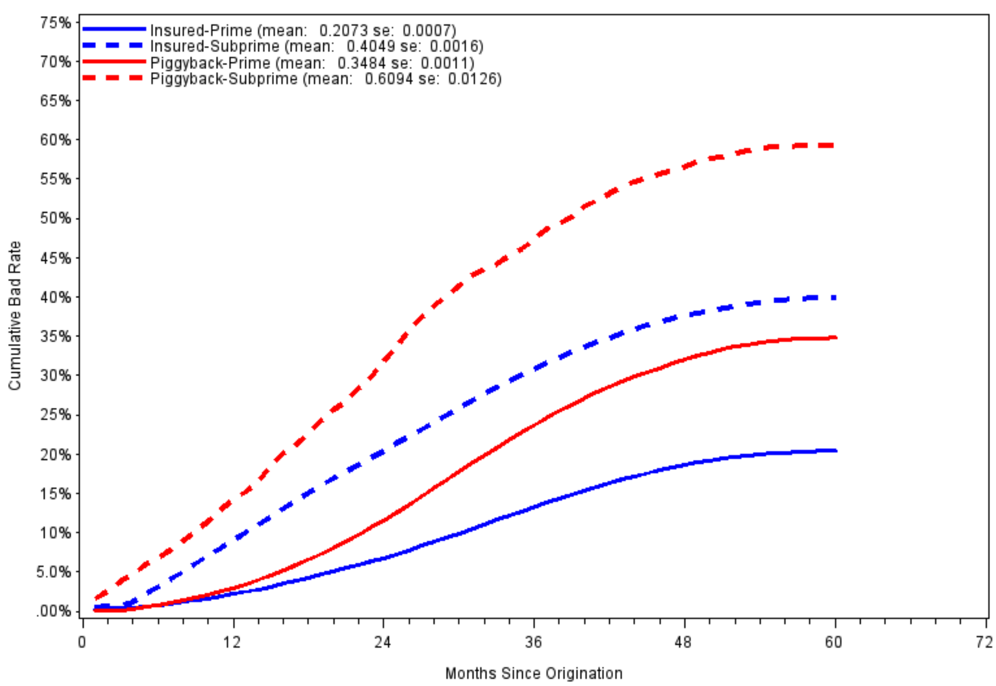



Cumulative Bad Rates for 2005 Vintage and CLTV GT90
Insured-Prime (mean: 0.1754 se: 0.0007)
Insured-Subprime (mean: 0.4051 se: 0.0016)
Piggyback-Prime (mean: 0.2496 se: 0.0011)
Piggyback-Subprime (mean: 0.4383 se: 0.0187)
Cumulative Bad Rate
Months Since Origination

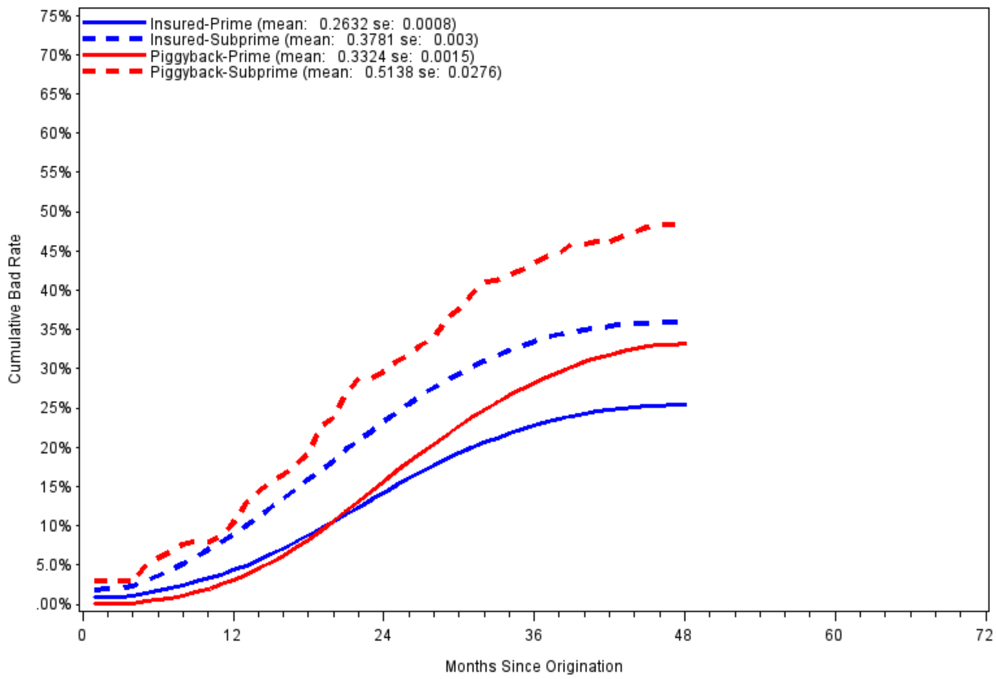
Cumulative Bad Rates for 2006 Vintage and CLTV LE90



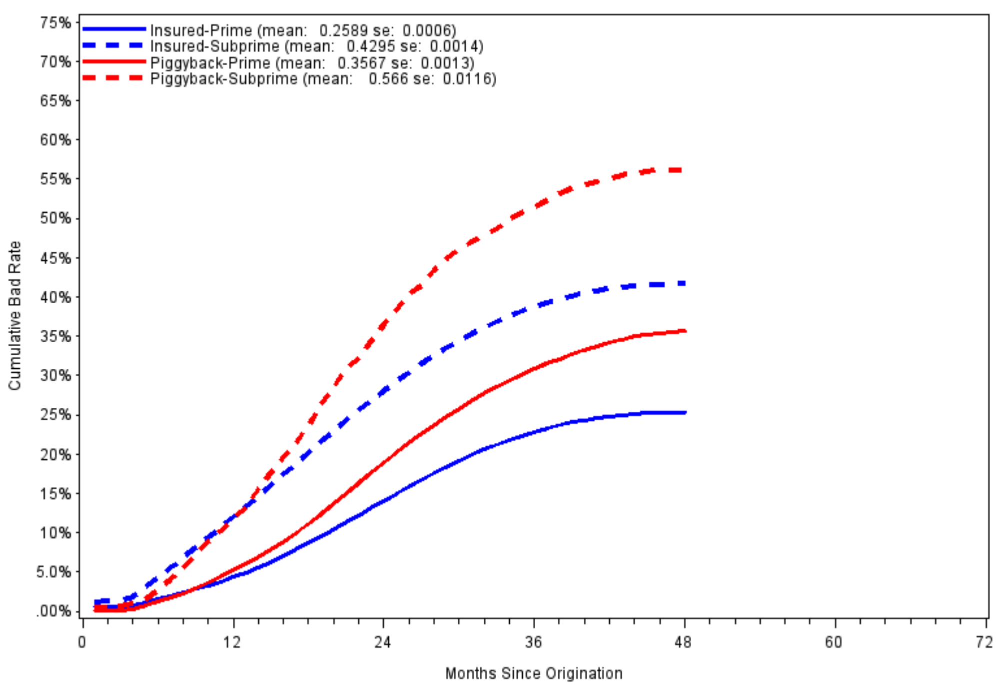

Cumulative Bad Rates for 2006 Vintage and CLTV GT90



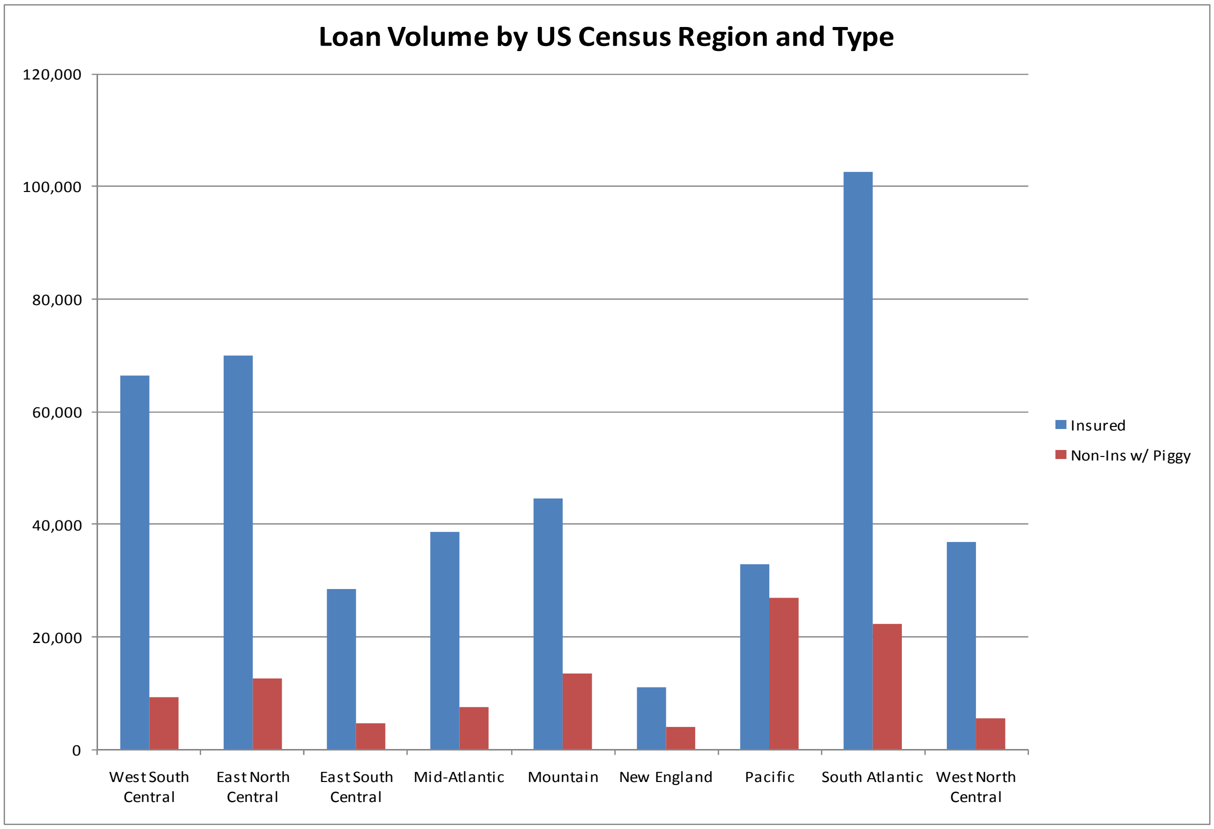
Cumulative Bad Rates for 2007 Vintage and CLTV LE90

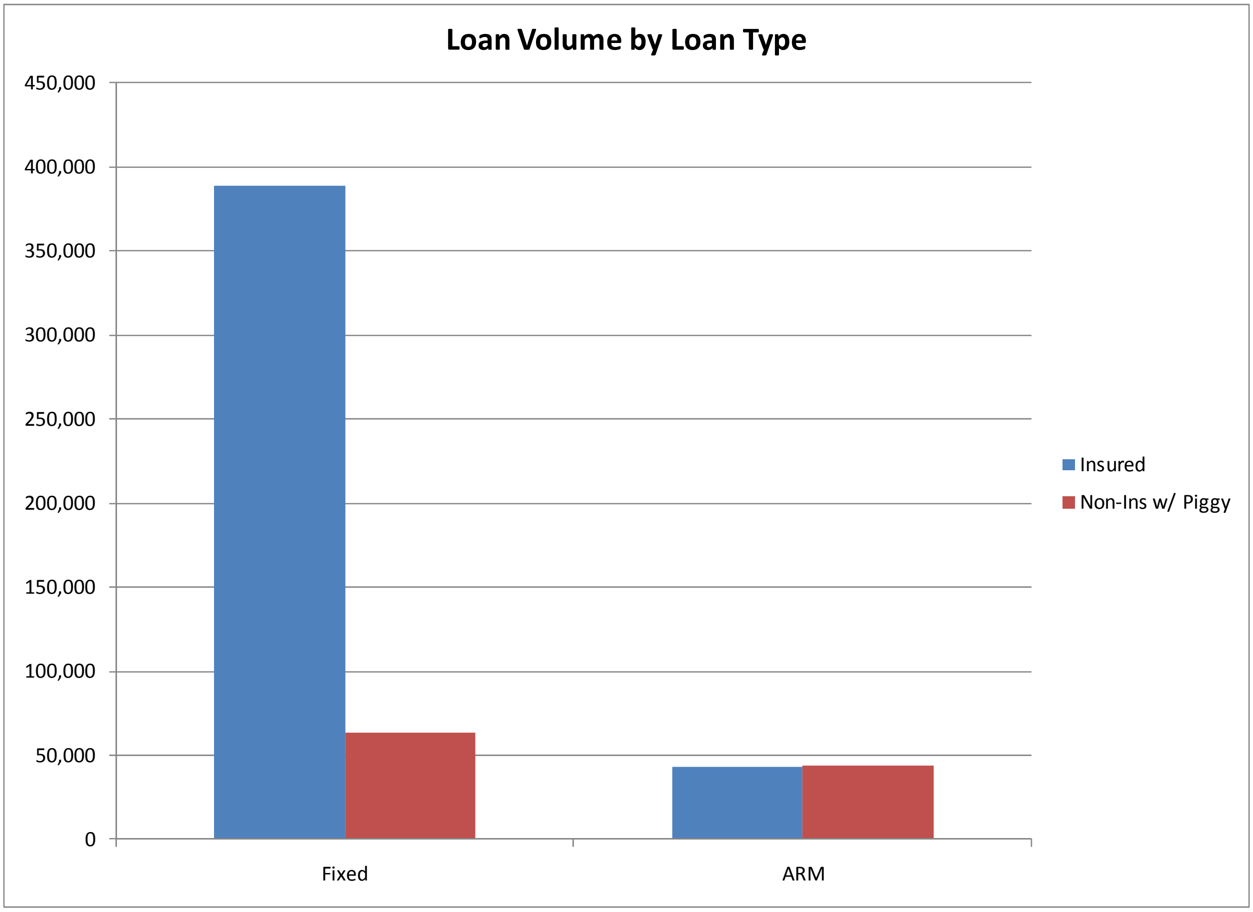


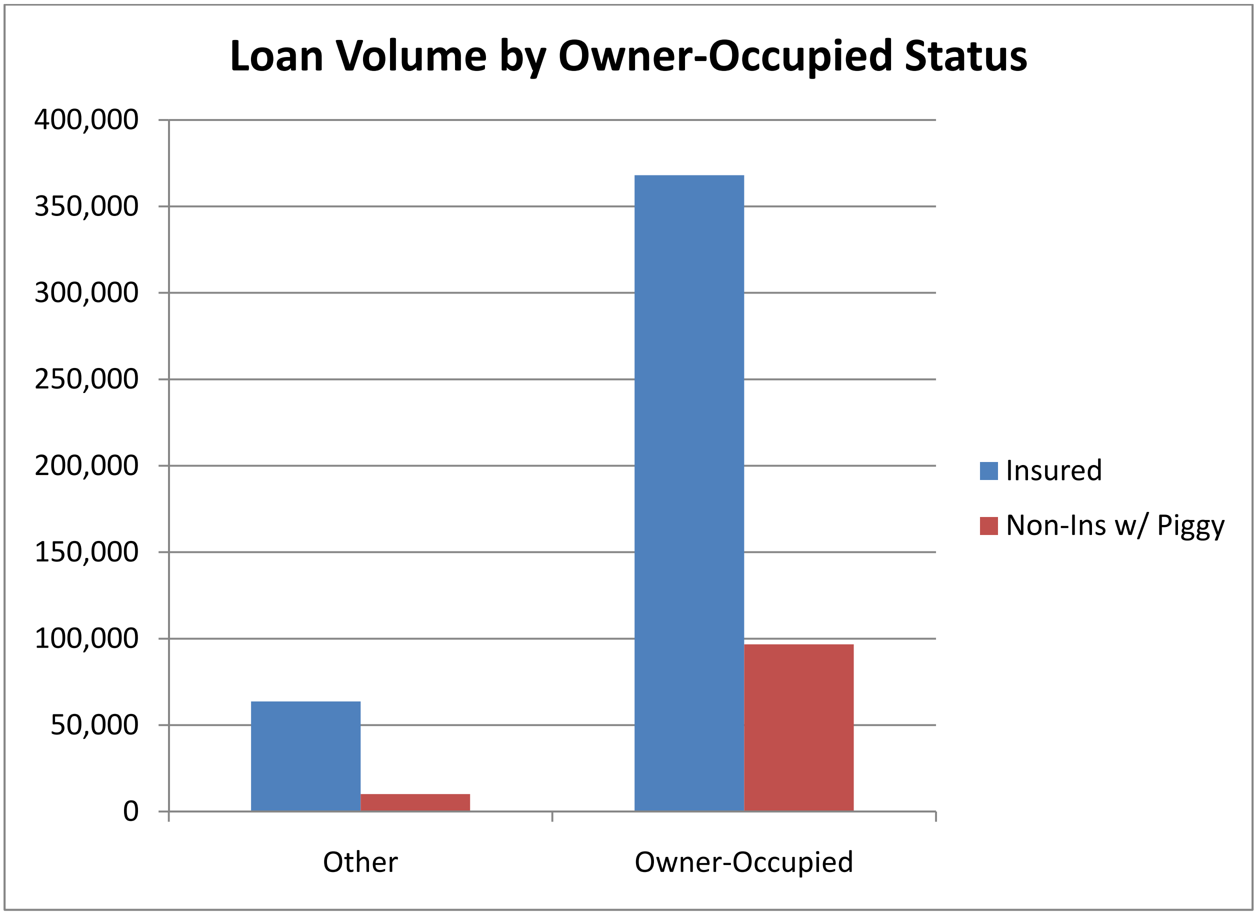

Cumulative Bad Rates for 2007 Vintage and CLTV GT90

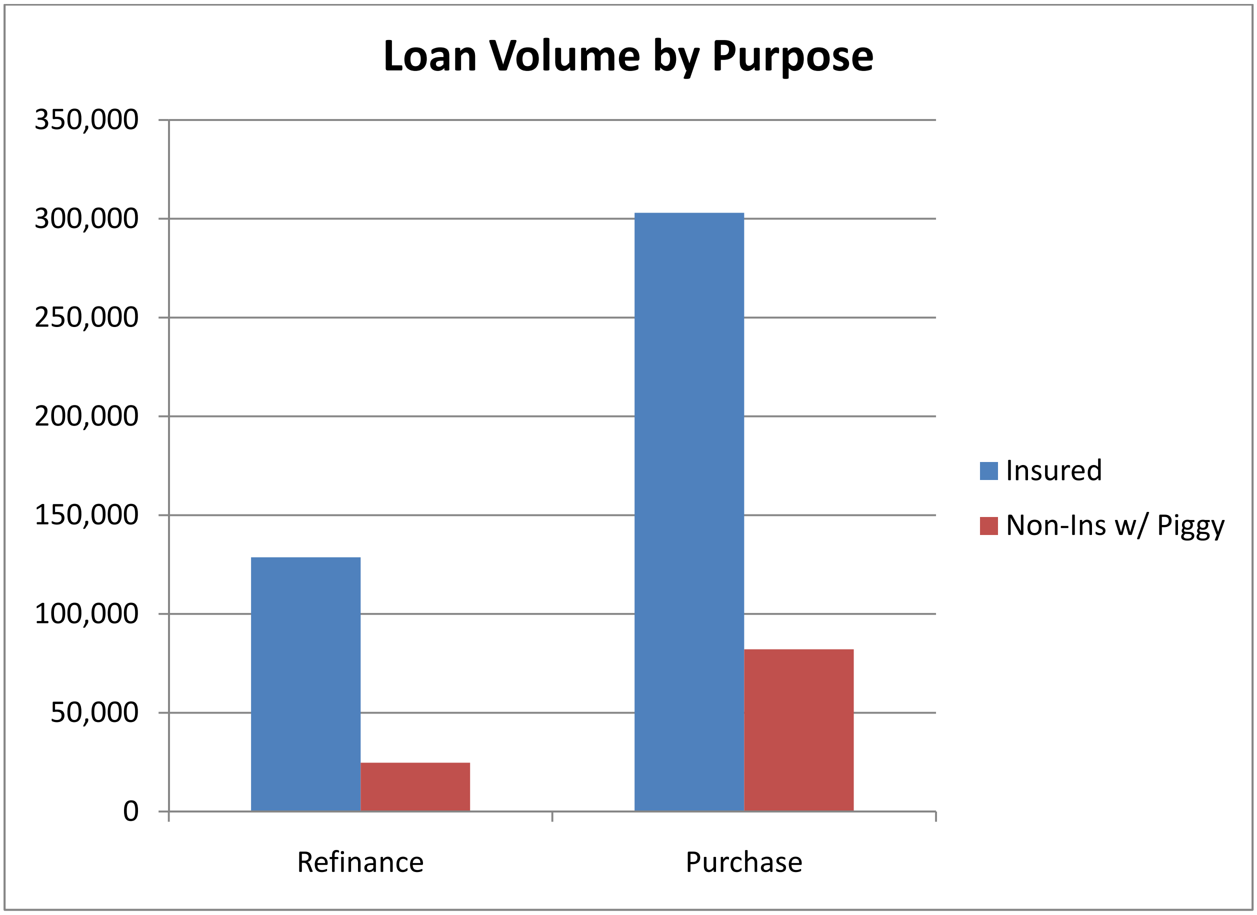


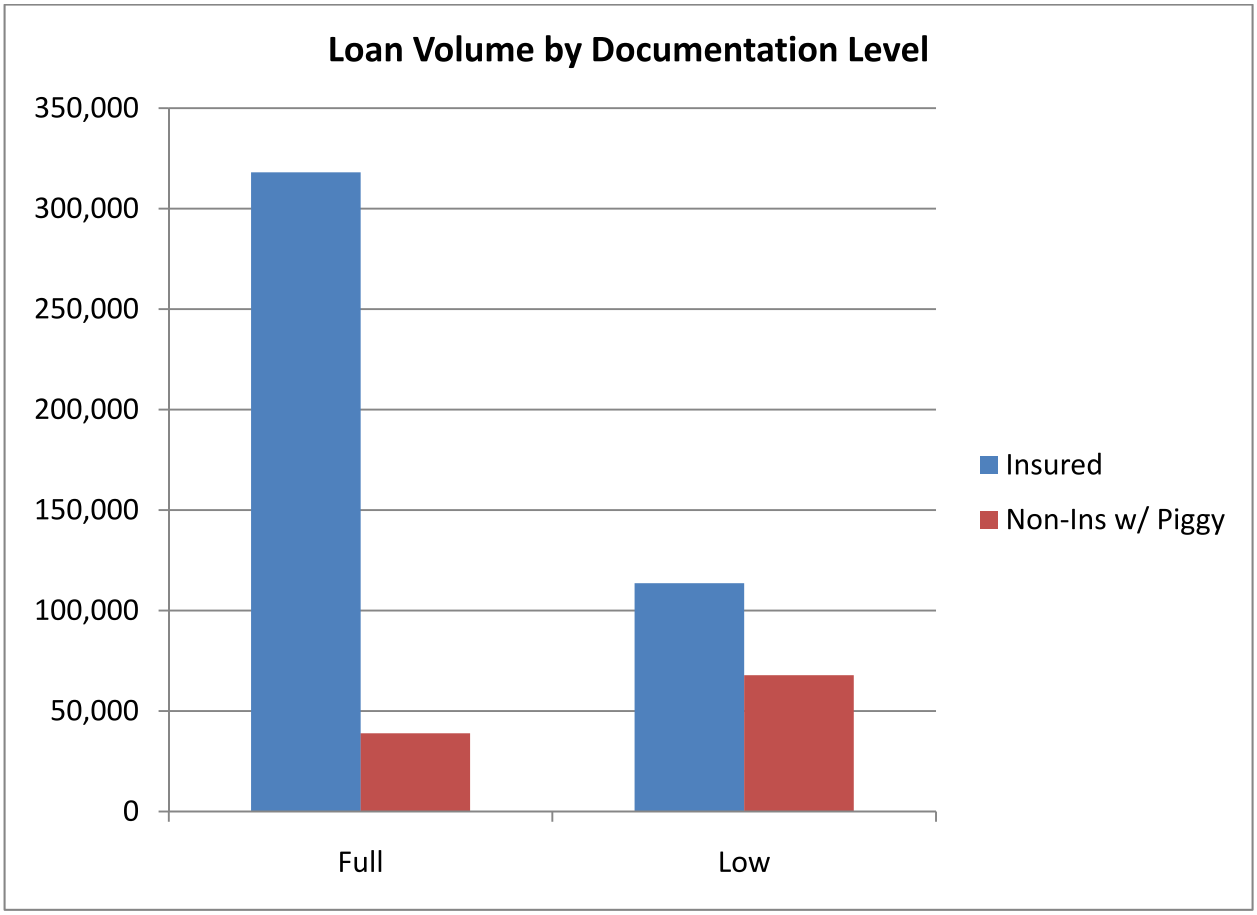
# Appendix B:  Survival Analysis Modeling Dataset Summary

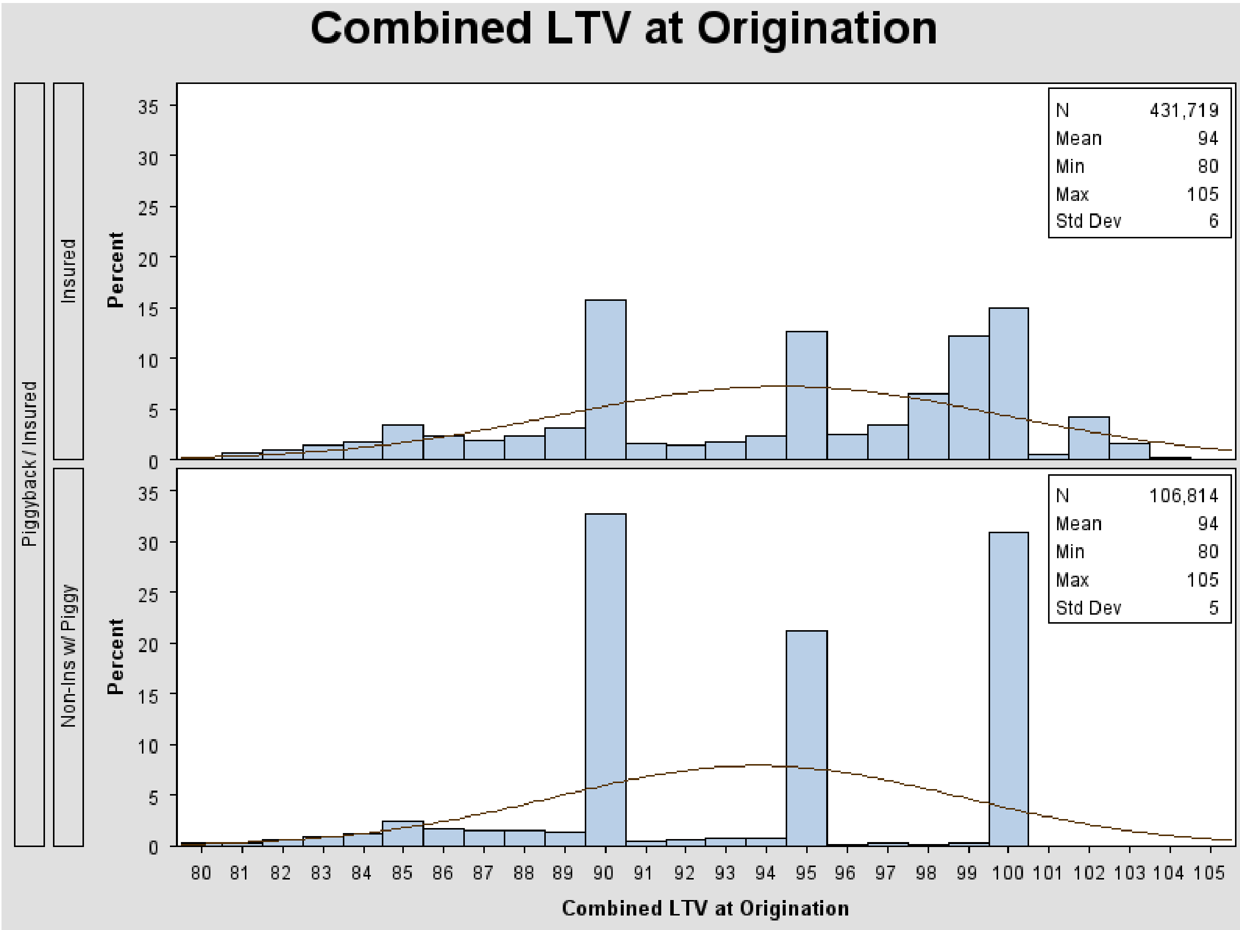



Loan Volume by Loan Type
450,000
400,000
350,000
300,000
250,000
200,000
150,000
100,000
50,000
0
Fixed
ARM
Insured
Non-Ins w/ Piggy

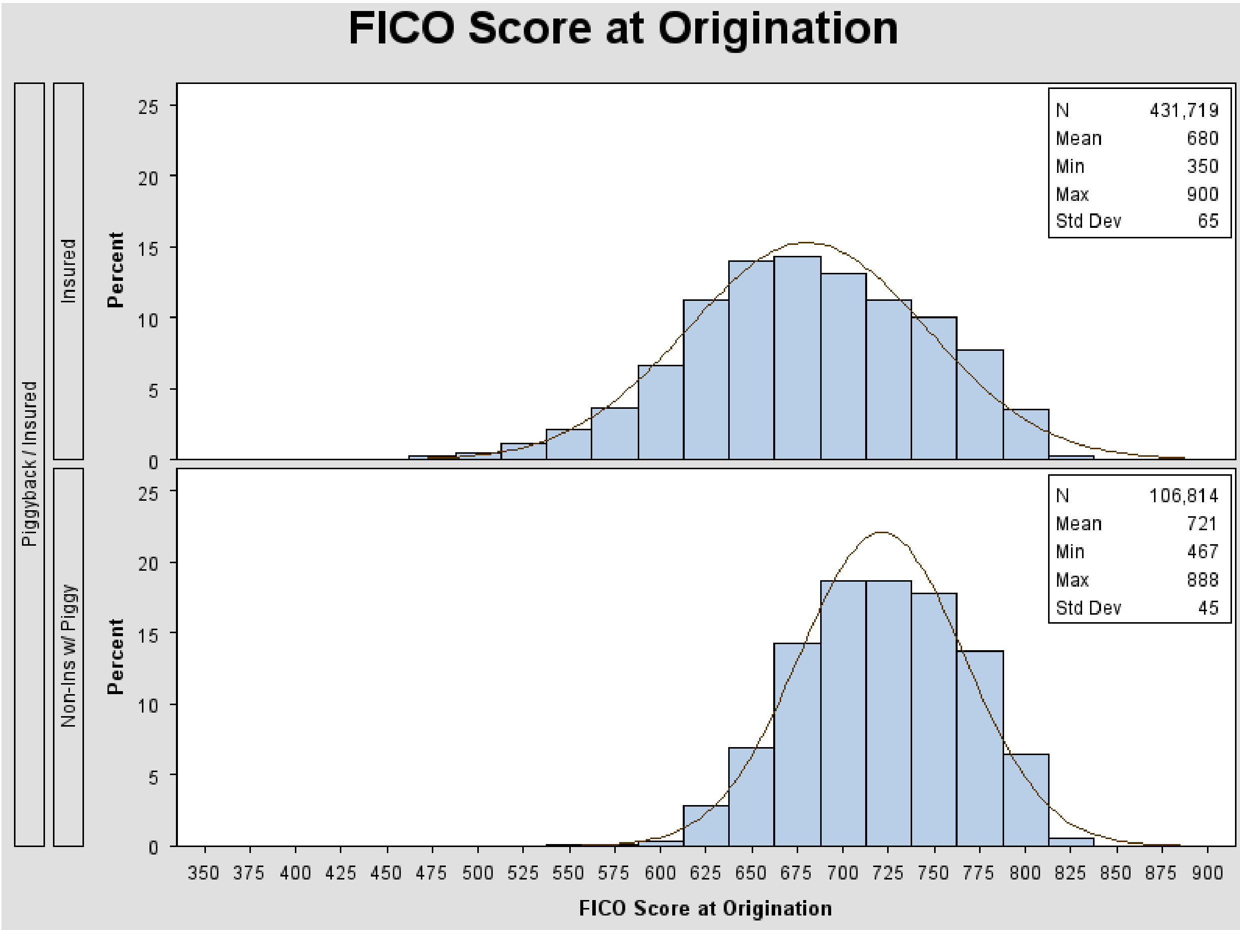

Loan Volume by Owner-Occupied Status
400,000
350,000
300,000
250,000
200,000
150,000
100,000
50,000
0
Other
Owner-Occupied
Insured
Non-Ins w/ Piggy


Loan Volume by Purpose
350,000
300,000
250,000
200,000
150,000
100,000
50,000
0
Refinance
Purchase
Insured
Non-Ins w/ Piggy


Loan Volume by Documentation Level
350,000
300,000
250,000
200,000
150,000
100,000
50,000
0
Full
Low
Insured
Non-Ins w/ Piggy



Combined LTV at Origination
Piggyback / Insured
Insured
Non-Ins w/ Piggy
Percent
N 431,719
Mean 94
Min 80
Max 105
Std Dev 6
N 106,814
Mean 94
Min 80
Max 105
Std Dev 5
0 5 10 15 20 25 30 35
80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105
Combined LTV at Origination


FICO Score at Origination
Piggyback / Insured
Insured
Non-Ins w/ Piggy
Percent
N 431,719
Mean 680
Min 350
Max 900
Std Dev 65
N 106,814
Mean 721
Min 467
Max 888
Std Dev 45
0
5
10
15
20
25
350 375 400 425 450 475 500 525 550 575 600 625 650 675 700 725 750 775 800 825 850 875 900
FICO Score at Origination

# Appendix C:  Scaled Schoenfeld Residual Plots

The Schoenfeld residual, $r_{ik}$ is the covariate value, $X_{ik}$, for the $i^{th}$ loan which actually defaulted at time t, minus the expected value of the covariate for the risk set at time t (i.e., a weighted-average of the covariate, weighted by each loan's likelihood of defaulting at t).

Because they will vary in size and distribution, the Schoenfeld residuals are usually scaled before being analyzed.  The k-dimensional vector of **Scaled Schoenfeld Residuals, SR,** for the $i^{th}$ loan is defined as:

$SR= \boldsymbol{\beta} + D^*Cov(\boldsymbol{\beta})^*\mathbf{r}_i'$

where

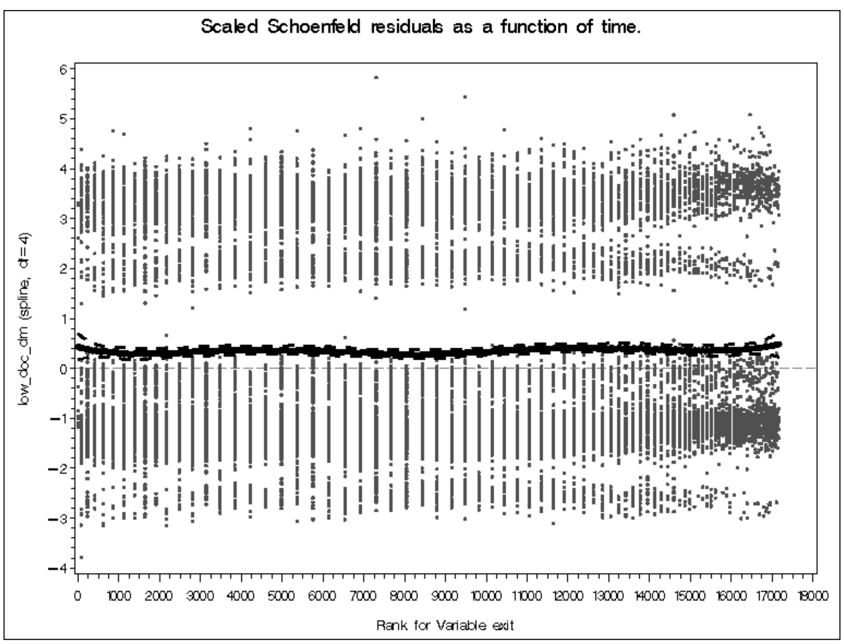
$\boldsymbol{\beta}$=the estimated Cox model coefficient vector

D= the number of loans defaulting, and

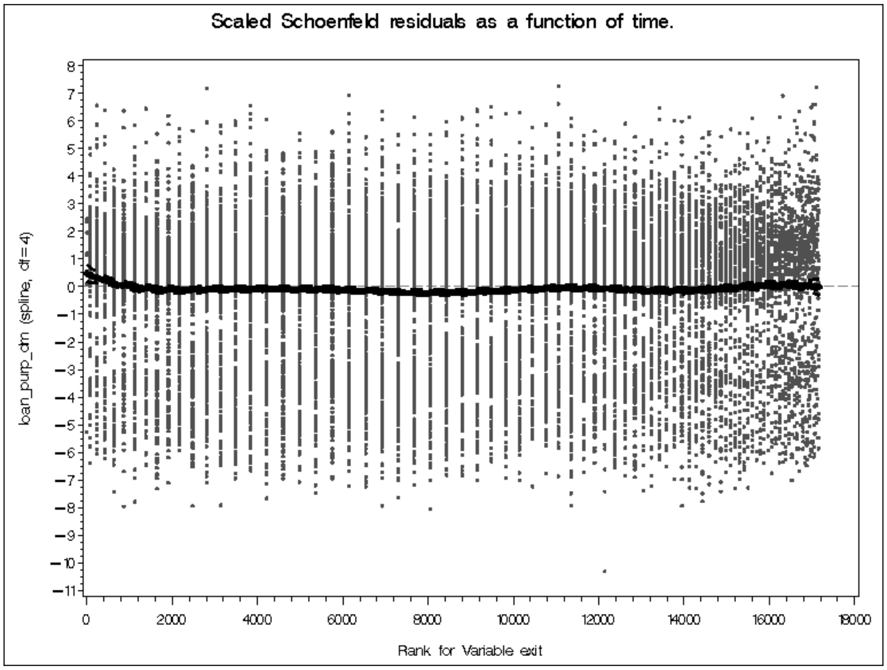
$\mathbf{r}_i$= the vector of Schoenfeld residuals for loan i.

## Plots for Fixed-Rate Loans, by Covariate

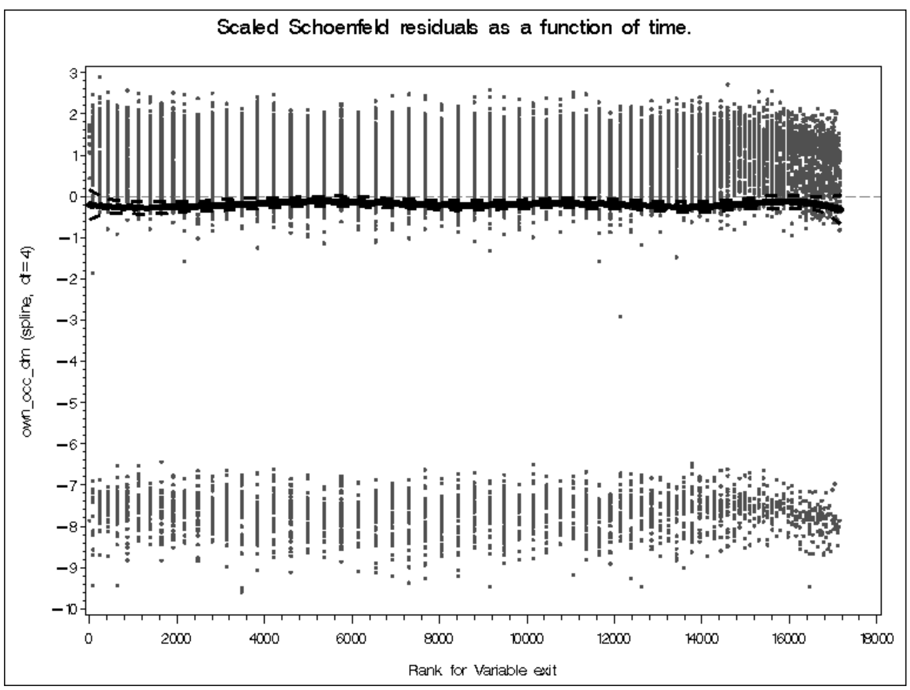
**Documentation Level**



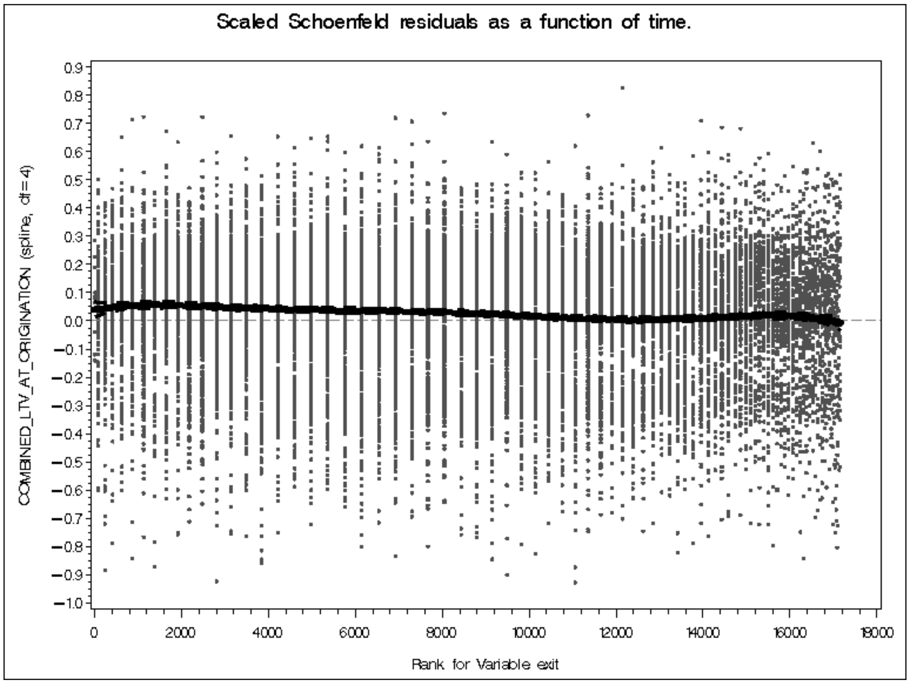
## Loan Purpose



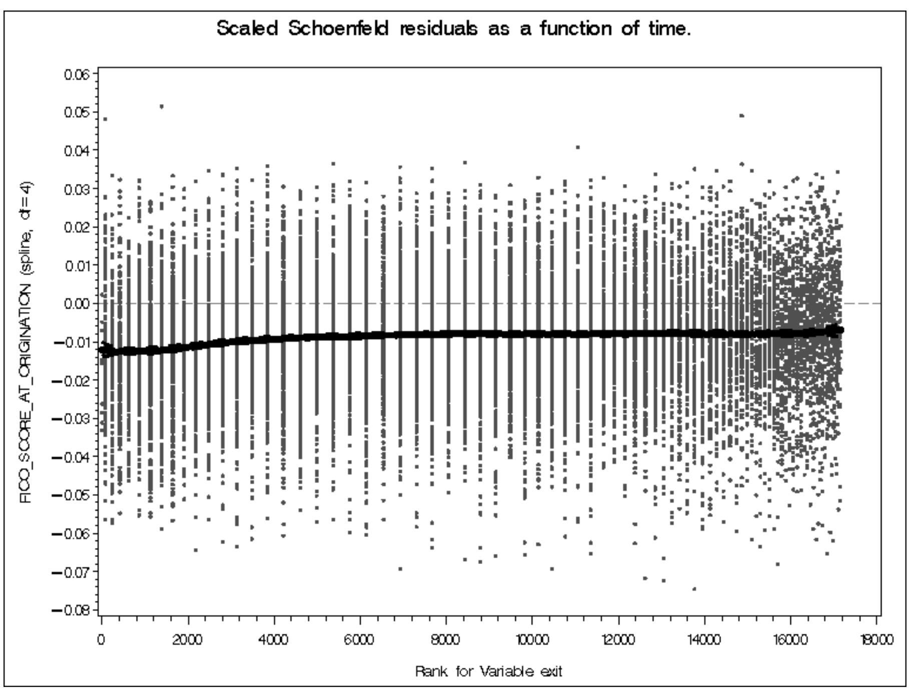

## Occupancy Status



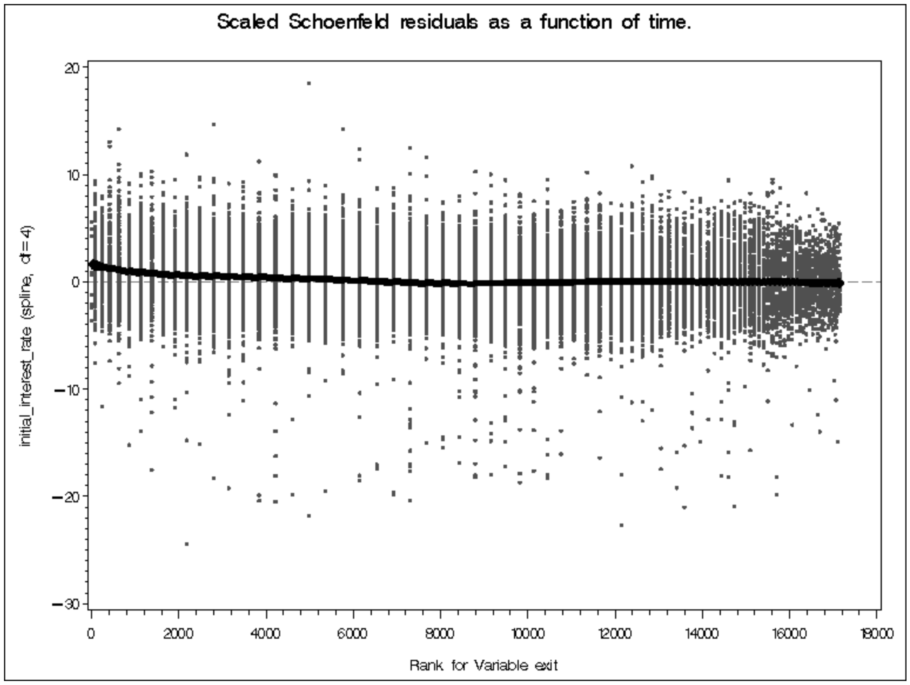
## Combined LTV at Origination



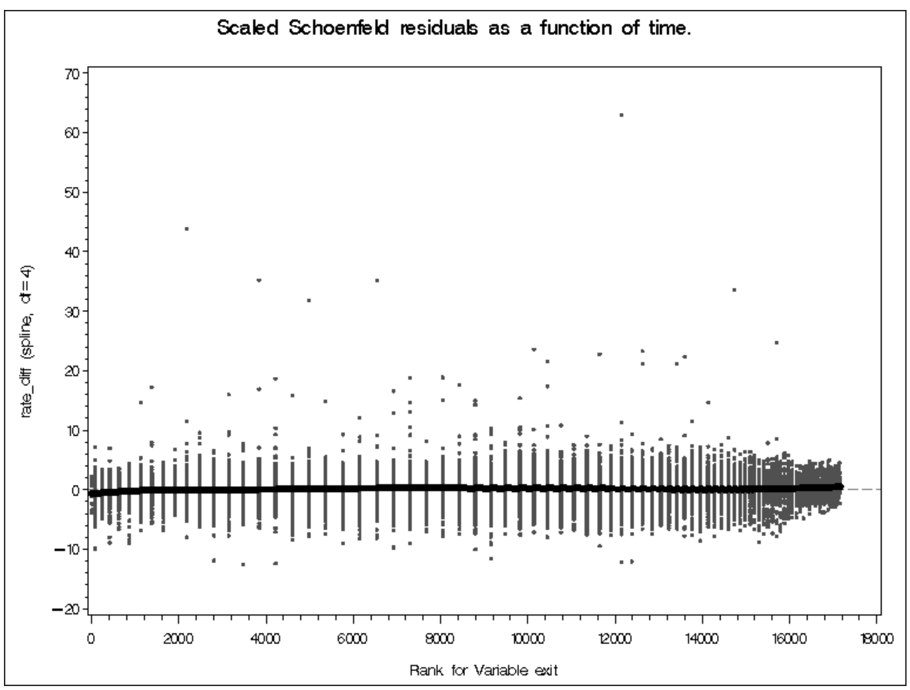

## FICO Score at Origination



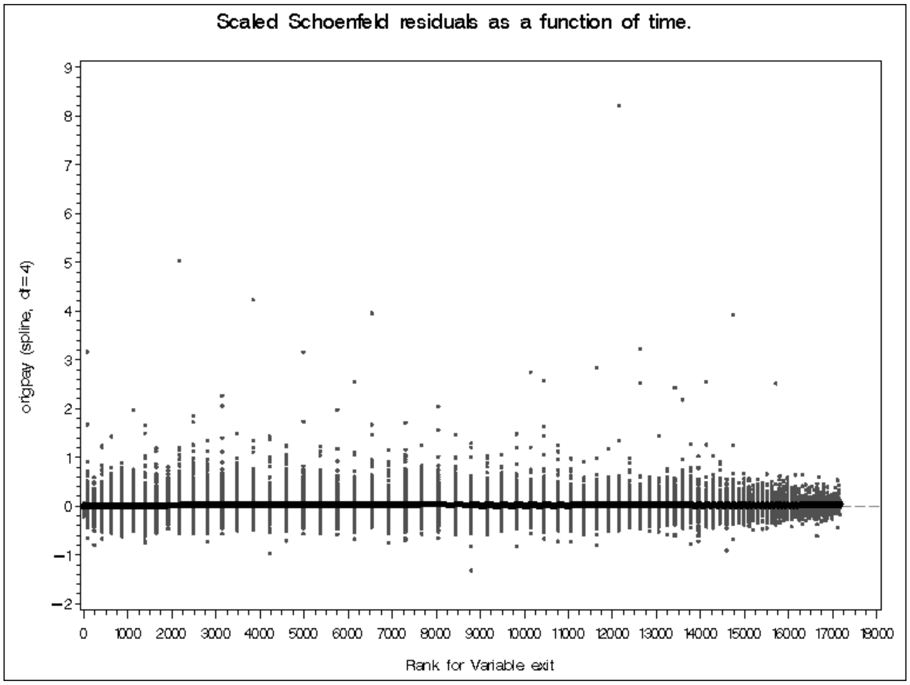
## Original Interest Rate



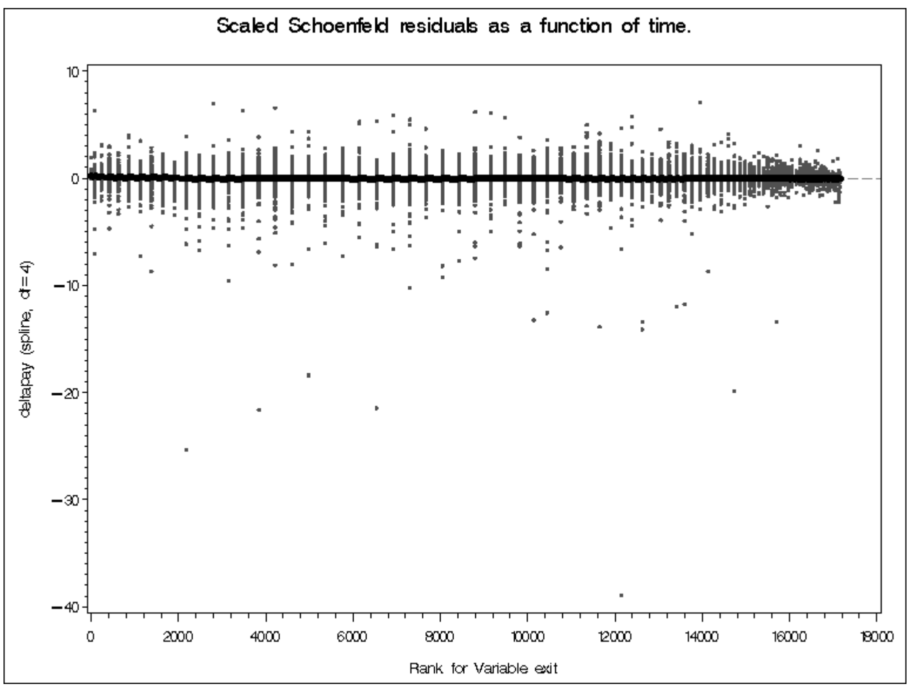

## Rate Differential (t)



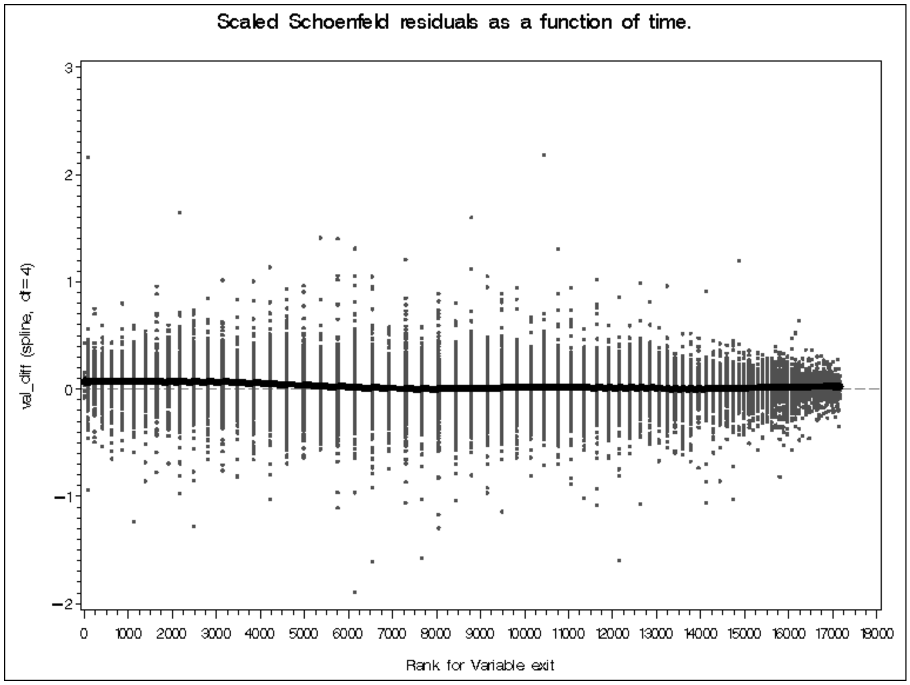
## Original Payment



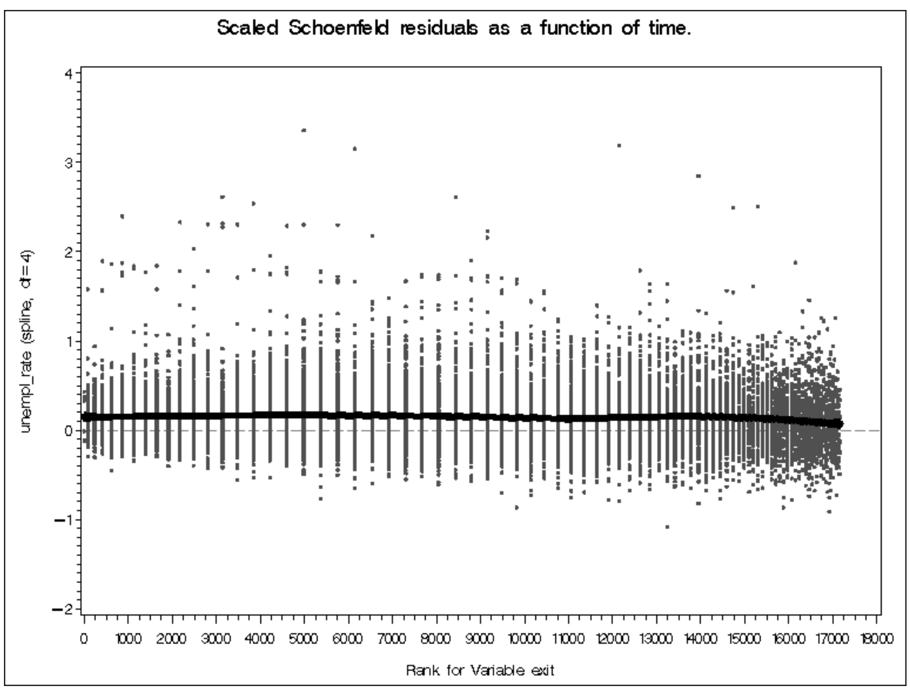

## Change in Payment (t)



## Change in Value (t)

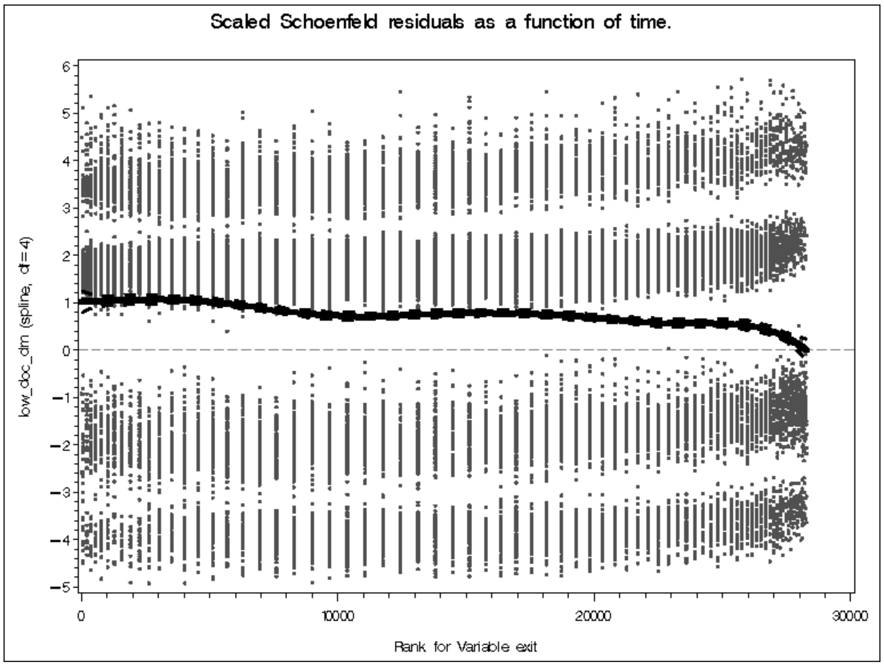



## Unemployment Rate (t)

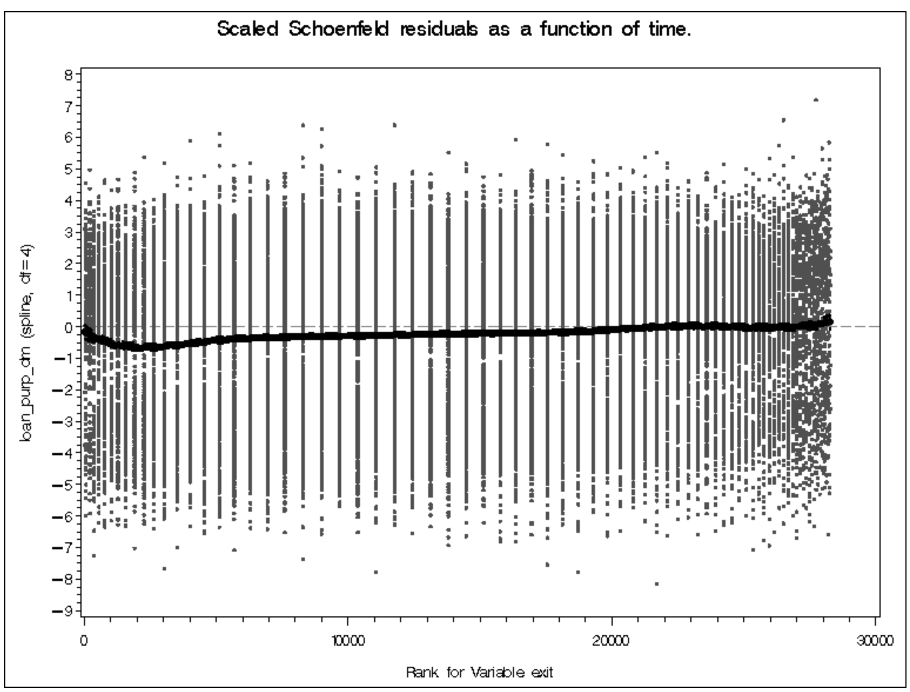


## Plots for Adjustable-Rate Loans, by Covariate

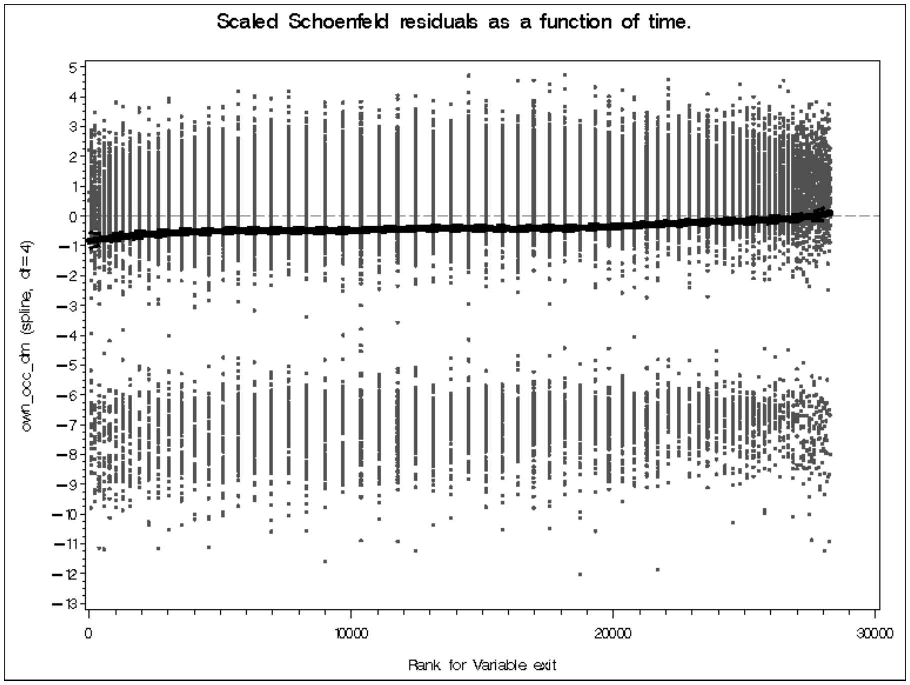
**Documentation Level**



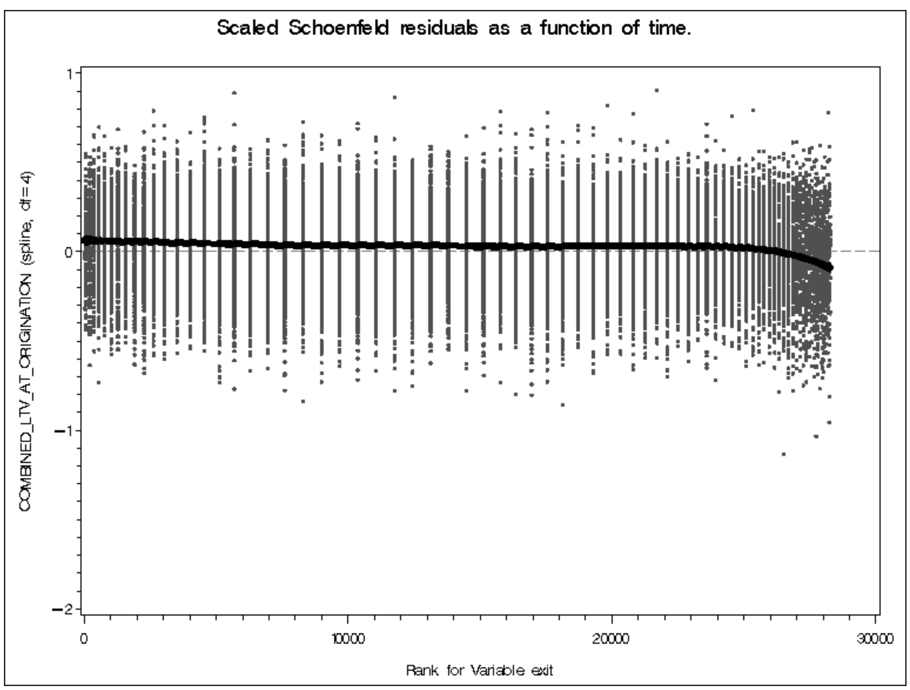

**Loan Purpose**



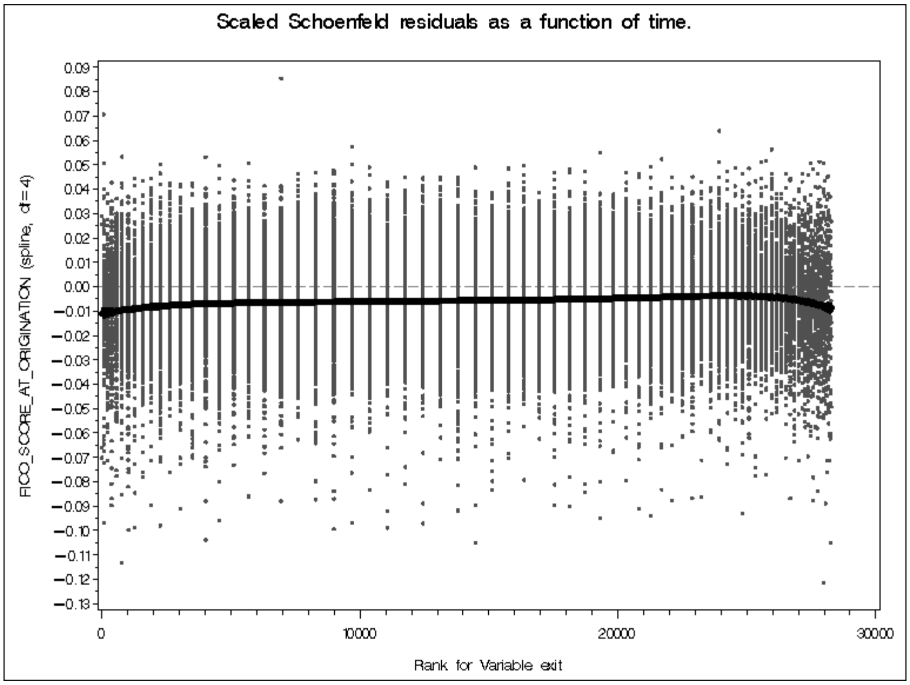
## Occupancy Status



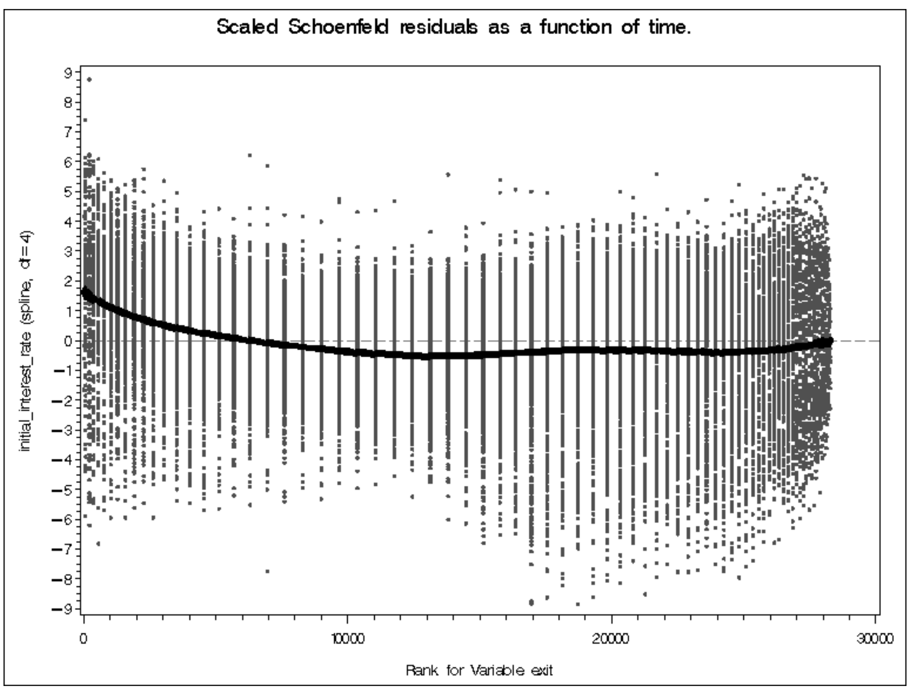

## Combined LTV at Origination



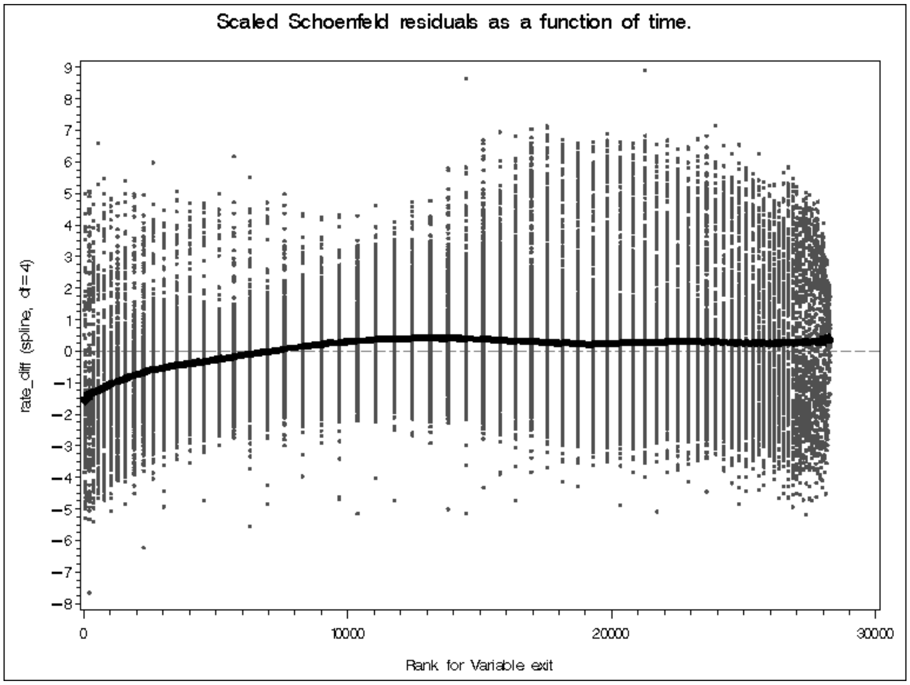
## FICO Score at Origination



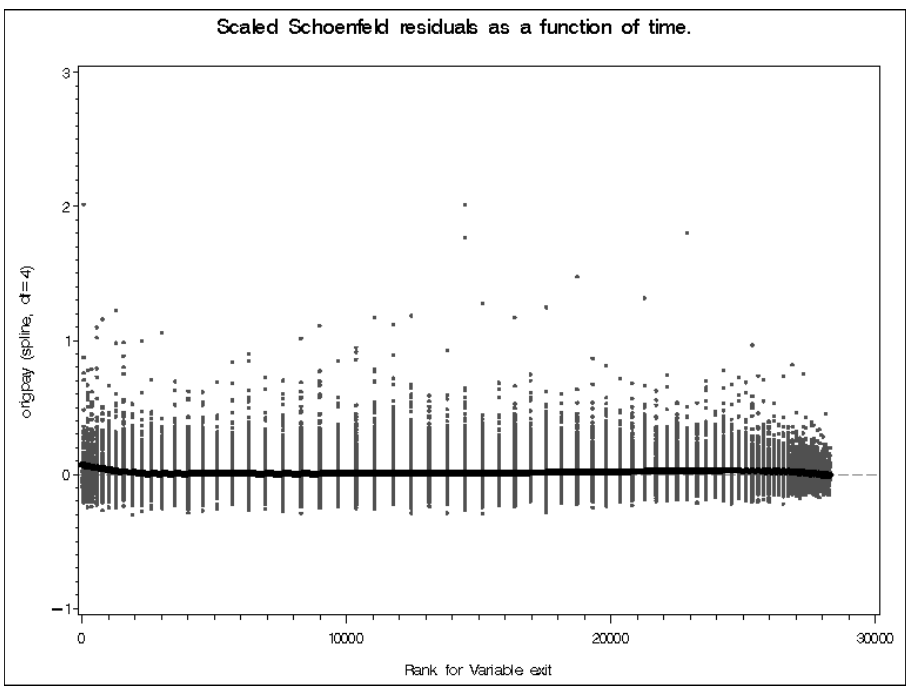

## Original Interest Rate



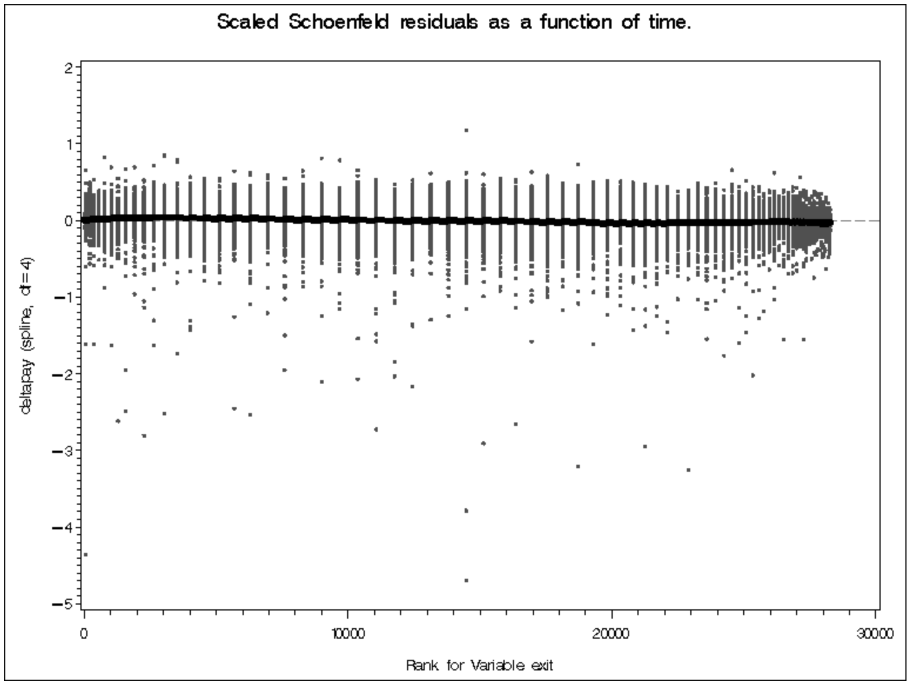
**Rate Differential (t)**



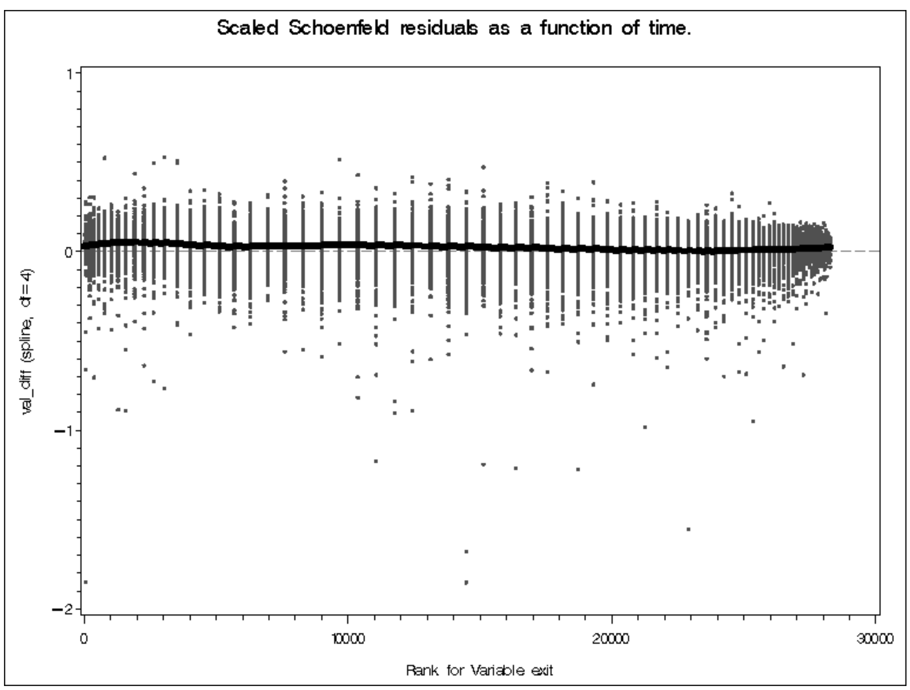

**Original Payment**



### Change in Payment (t)




### Change in Value (t)



## Unemployment Rate (t)