Paper 65-2018

Using Net Lift Measure to Model the Effects of Changes in Interest Rates on Mortgage Prepayment

Alec Zhixiao Lin, Lin's Analytics, Los Angeles, CA
Xiao Hu, University of Southern California, Los Angeles, CA

ABSTRACT

Mortgage loan prepayment is of constant interest to both academia and practitioners. A considerable drop in the market interest rates will trigger a wave of early payoffs and cause losses to investors in mortgage-related derivatives. This paper suggests the use of net lift measure to estimate the effects of changes in interest rates on underlying mortgage prepayment. The loan-level modeling method will help investors with more accurate valuations of mortgage-related assets. Insights from such a study are also useful to banks owning mortgage portfolios for developing retention strategies in the awake of a mortgage rate drop.

INTRODUCTION

The majority of mortgagors have the option to prepay their loans for any reason at any time without penalty. There are three causes for prepayment: turnover, cash-out refinance and rate/term refinance. A dramatic fall in the market rate can trigger an unprecedented wave of prepayment in the following ways:

1. Many borrowers want to take the opportunity to upgrade to bigger or better houses after selling their current ones. The sale of a current dwelling before it reaches its full maturity results in a prepayment.
2. Although cash-out refinance is often associated with increasing house values, lowered rates will also encourage some borrowers to cash-out a portion of their home equity for other uses such as home improvement, vacation, payment for college tuition for their children, etc.
3. Many borrowers will apply for rate/term refinance to lower down their monthly mortgage payment or to shorten their loan term in order to save interest cost.

What is good for borrowers is not necessarily good for investors. A surge in prepayment will shorten the term of the investment and reduce the return. Mortgage servicers will also see the revenue streams dry up as many loans are abruptly marked as Paid In Full before reaching expected maturity.

As interest rates are notoriously difficult to predict, forecasts based on different economic scenarios are often used to generate stress-testing and OAS (option-adjusted spread). To help with asset evaluation and to better prepare investors for any shock by a drop in the market rate, accurately assessing the impact of interest rate changes is critically important. Among the multiple methods suggested for modeling prepayment, PSA (Public Securities Association) model is the most commonly used one but is often considered inadequate by many practitioners. This pool-level framework assumes that prepayments are driven exclusively by loan age and does not assess the impact of market rate movement, which many investors are concerned with. This paper suggests the use of net lift measure to improve accuracy. Findings collected from such studies will lend insights on building more robust and useful prepayment models for business.

NET LIFT MEASURE

Net lift model is also called uplifting model or true lift model. It bins a qualitative or binary outcome (usually expressed as 0 or 1) into a continuous outcome at pool levels. If the data contain enough levels (or bins) for the continuous outcome to show a good variability in values, we can apply linear regression for modeling. Observations of the same level or bin have the same value in the dependent variable.

One critical pre-requisite for applying net lift modeling is an experimental design that contains a control group and an experimental group. Let's use mortgage prepayment to illustrate the use of this methodology in our study.

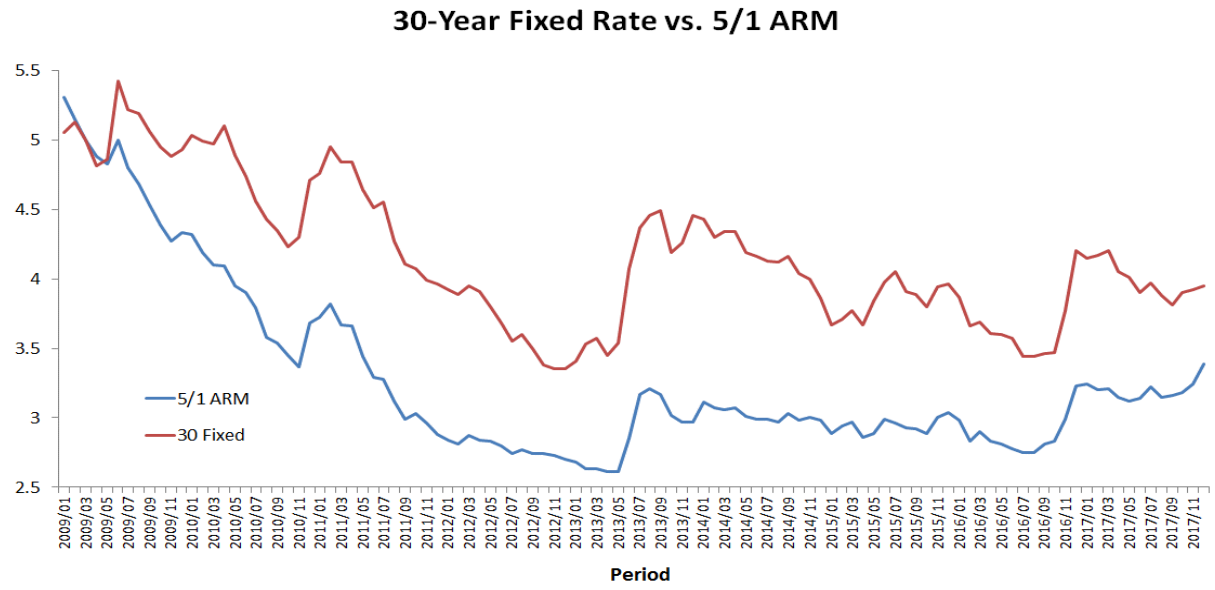

Figure 1 – Market Interest Rate by Month

A period of higher interest rate and a period of lower interest rate are usually a good pair for comparison. Based on Figure 1, we were able to discern four periods that can be used to form two pair of analysis (see Table 1 below). As refinancing usually incurs some cost, the new rates need to be sufficiently lower – usually more than 1% - in order to motivate borrowers. The market rates in the first pair exhibit a bigger difference and hence a better pair for a comparison.

Period	Average Market Rate	Incentive for Refinance
2009/01 to 2010/03	5.04	low
2012/07 to 2013/06	3.47	high
2013/07 to 2014/06	4.33	low
2015/01 to 2016/09	3.75	high

Table 1 – Periods selected for comparing early payoffs

Here is the basic framework for creating a net lift measure in our study:

1. A period of higher interest is considered as a control group. Property turnover always occurs regardless of the prevailing interest rates because trigger events such as marriage, child birth, divorce or retirement are not dependent upon market interest rates. Cash-out refinance occurs more often in the areas experiencing increasing house values. Rate/term refinance occurs infrequently.
2. With a dramatic drop in the interest rate, turnover increases. Cash-out refinance also becomes more common as a lower rate reduces the cost of borrowing. More importantly, rate/term refinance kicks in prominently as many borrowers want to take the opportunity to reduce their monthly mortgage payments. Therefore, incremental prepayment is expected in such a period. Our study considers a period of lowered interest rate as an experimental group.
3. The difference between above two steps is the net lift measure for incremental prepayment triggered by a change in the market rate. Table 2 shows that the voluntary prepayment is much higher in the period of 2012/07 to 2013/06 in all FICO bins when the market experienced a significant drop in interest rate. A linear regression can be used to estimate the impact of interest rate change by using the last column as the dependent variable. Observations in the same FICO bins have the same value for the dependent variable.

FICO	% Loans Prepaid	Prepayment 2009/01-2010/03	Prepayment 2012/07-2013/06	Prepayment triggered by Rate Change
< 600	3.9%	7.6%	13.4%	5.8%
600-619	3.9%	8.6%	14.5%	5.9%
620-638	6.2%	9.1%	16.6%	7.5%
640-659	8.1%	9.6%	18.6%	9.0%
661-679	9.8%	10.7%	21.2%	10.5%
680-699	10.7%	11.9%	22.2%	10.3%
700-719	11.2%	12.7%	23.5%	10.7%
720-739	12.6%	14.7%	24.3%	9.6%
740-759	14.2%	15.8%	25.9%	10.1%
760-779	13.3%	16.1%	25.8%	9.7%
78-799	5.8%	13.6%	22.8%	9.2%
800+	0.3%	10.6%	18.3%	7.7%

Table 2 – Deriving a net lift measure

The net lift measure used in our study does not differentiate between turnover, cash-out refinance and rate/term refinance. This greatly simplifies how the outcome variable is defined and how the data should be collected and cleaned.

FINDING A DEPEDENT VARIABLE WITH A GOOD VARIABILITY

In Figure 2, the prepayment in the control group is denoted by blue area at the bottom. A drop in the interest rate will trigger additional prepayment, which is denoted by the red area above. That is,

$$\%\text{Incremental Prepayment}_{\textit{triggered by Rate Drop}} = \%\text{Prepayment}_{\textit{Low-Rate Period}} - \%\text{Prepayement}_{\textit{High-Rate Period}}$$

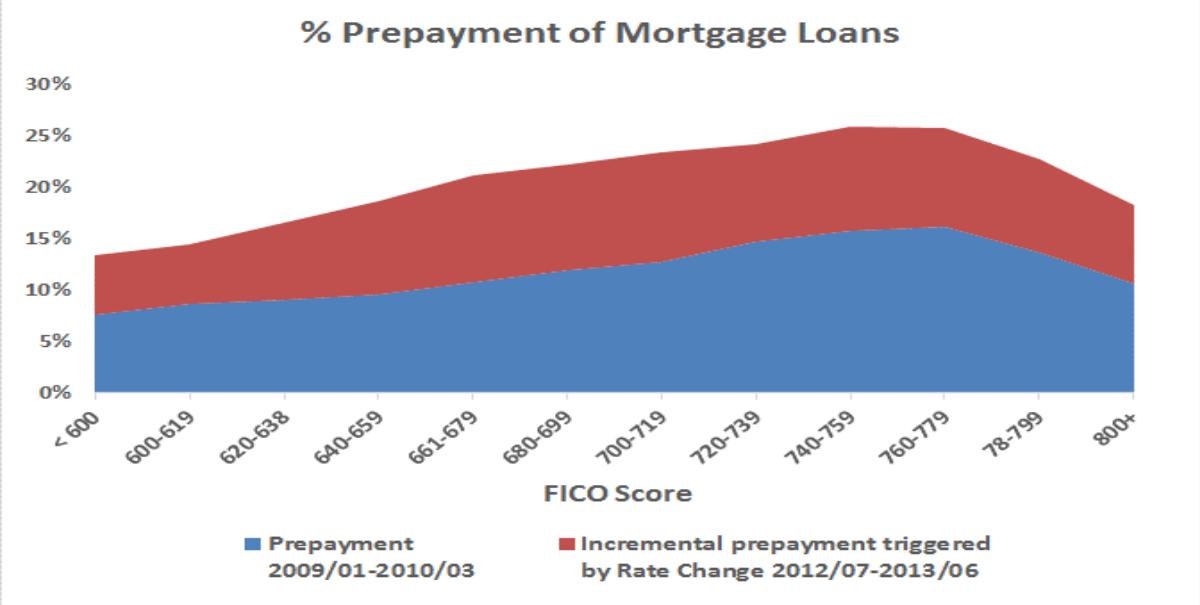

Figure 2 – Additional prepayment triggered by a rate drop

If we can find enough pool-level variability in the outcome, we can use it as the dependent variable to model on loan-level characteristics to estimate the impact of rate drop on voluntary mortgage prepayment. The net lift measure for additional prepayment in the Table 2 and Figure 2 ranges from 5.8% to 10.7%. After experimenting with different combinations of variables, we found that the deviation

between borrowers' coupon rates and prevailing market rate can give a much wider spread from 1.6% to 16.0%, with a close-to-even distribution. We can expect a stronger model based on this treatment.

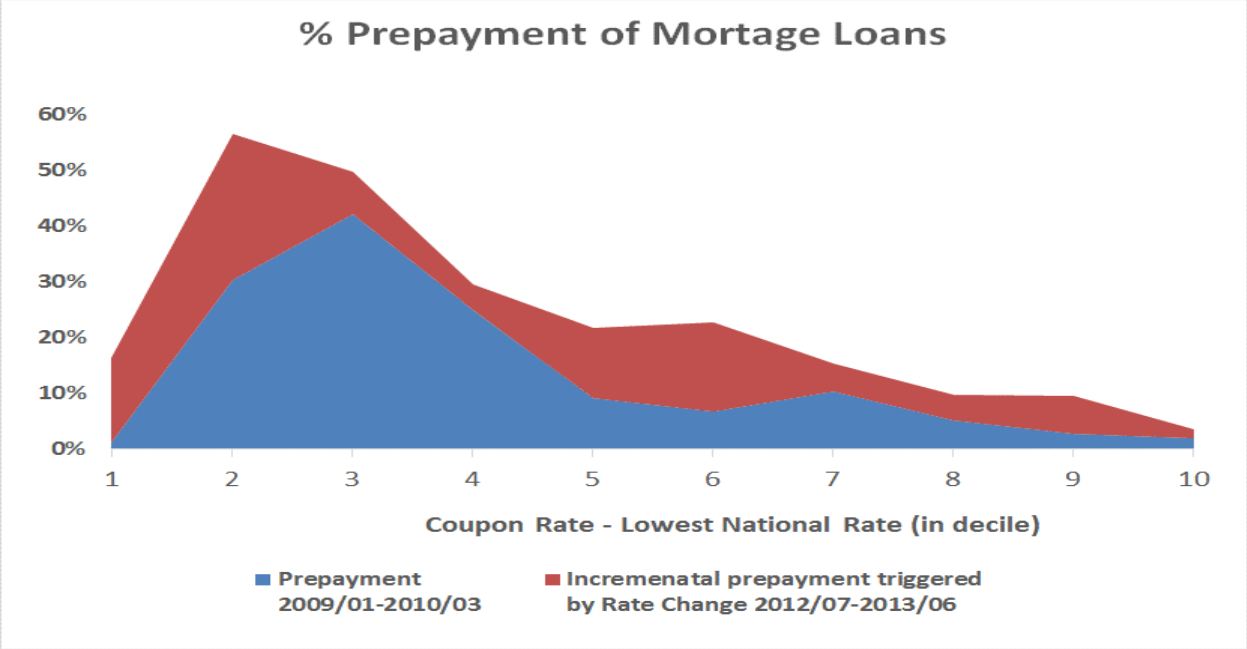

Figure 3 – A better net lift measure for modeling

We would like to point out one thing. The values of outcome is limited to the aggregated averages by segment, so we do not expect the R^2 for linear regression to be as high as one might expect for most linear models.

THE DATA

Freddie Mac publishes loan-level performance data of fully amortizing fixed-rate mortgages that the company purchased or guaranteed from 1999 and on[1]. Terminated loans show flags for voluntary prepayment, foreclosure, and repossession (REO). Customer characteristics such as FICO and loan-to-value ratio (LTV) at origination are also available. We take the following measures to make a sample of loans for our study:

- Loans with 30-year fixed rates only.
- Single-family houses only.
- Loans from three states: Massachusetts, Tennessee and Georgia. Connecticut, Massachusetts and New York have the highest home price in the US, with median price for new single-family home prices all above $410,000 as of first quarter of 2014. Mississippi, Kentucky and West Virginia have registered the lowest medians for new home prices, all under $200,000[2]. Georgia is selected because it is said to be closest to the overall population of homeowners in the US[3].
- Loans with more than 120 remaining months to maturity[4].
- FICO ≥ 660 to exclude sub-prime loans.

To more accurately define voluntary early payoff, we take the following steps:

[1] Fannie Mae publishes similar data.
[2] National Association of Home Builders: http://eyeonhousing.org/2014/07/new-home-prices-by-metro-area-and-state/
[3] The Washington Post: https://www.washingtonpost.com/news/the-fix/wp/2015/09/30/if-being-representative-is-your-goal-illinois-should-be-the-first-primary-state/?utm_term=.ba8117153e77
[4] Borrowers with short terms remaining in loan term are more likely to pay off the remaining balance by using personal savings.

- Current Actual Unpaid Balance becomes zero.
- Ending status of a loan shows a voluntary payoff.
- No delinquency for more than 90 days in performance history.
- No modification for the loan term.
- No recoveries. Recoveries are usually associated with foreclosure or bankruptcy.
- No Foreclosure or repossession (REO).

EXPLANATORY VARIABLES

There is a general agreement that three factors appear to drive prepayment: 1) the prevailing mortgage rates relative to the coupon; 2) the characteristics of the mortgage (e.g., loan size, loan age, location of the property), and 3) time of the year. The decision of early payoff is driven by a combination of these factors.

Our study is restricted by the data available to us. For modeling, we consider using the following explanatory variables:

- FICO Score at origination: Fico score of the borrower at origination (when the application is received)[5].
- Original LTV (loan-to-value) ratio.
- Original CLTV (combined loan-to-value) ratio.
- DTI (debt-to-income) ratio at origination.
- Original loan amount.
- Prepayment penalty flag: mortgages with prepayment penalty are hypothesized to have less incentive for refinancing.
- Loan purpose at origination (turnover, cash-out refinance or rate/term refinance)
- Loan age
- Current loan amount, expressed as current actual UPB (unpaid balance).
- Change in LTV, which is the difference between current LTV at the time of prepayment and LTV at origination.
- Burnout as a refinancing incentive, defined as follows.

$$\sum_{t=1}^{j} max(log(C/R_t), 0)$$

where C = coupon (or contract) rate of a mortgage

R_t= interest rate available for refinancing at period t

However, the above formula is too simplistic. It suggests that the higher the coupon rate on a mortgage, the higher incentive the borrower has for refinance. Although this makes psychological sense, borrowers with very higher coupon rates due to low FICO scores and high DTI are unlikely to be eligible for a much lower prevailing market rate. For example, a borrower with a coupon rate of 6.75% for a mortgage originated six months ago is unlikely to qualify for 4% for refinancing.

After repeated examinations on and experiments with data from different states and across different periods, we designed two variables associated with burnout. The following is the first derived variable:

$$\text{burnout\_raw} = \sum_{t=1}^{j} max(log(C/(1.1 \times (R_t + LLPA_{adj}))), 0)$$

[5] The data show the FICO score of the loan at the time of acquisition by Freddie Mac. It is usually several months later than origination. We consider it as a good substitute for FICO at the origination.

where $LLPA_{adj}$ =loan-level pricing adjustment as stipulated by Fanny Mae and Freddie Mac for adjusting interest rate with borrowers' FICO score and LTV[6].

We further make adjustment to account for incentive and eligibility as follows to account for eligibility[7]:

$$\text{burnount\_factor}=-\sqrt{|\text{burnout}_{\text{raw}}-0.3|}$$

After observing preliminary modeling results, we decided to add the following second variable for eligibility, similar to a second-degree measure:

$$\text{rate\_eligibility}=|C-5.875|$$

It suggests that loans with coupon rates above 5.875% are even less likely to get refinanced.

Our data do not have the following attributes:

- Cumulative home price appreciation.
- Current DTI at the time of refinancing. Although this information is collected by loan originators at the time of refinancing, it is usually not available for existing loans unless complete credit information is pulled from credit bureaus.
- WALA (remaining weighted average loan term). This is loan ages weighted by UPB to the total UPB of the pool. Since some loans are paid off faster, the weight will be also changing by months.

RESULTS

We ran three separate models for three states. Similar results were obtained as follows:

Variable	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	0.00336	0.00413	0.81	0.4165	0
FICO	0.00001084	9.30E-07	11.65	<.0001	1.00696
DTI	-0.00048408	2.68E-05	-18.07	<.0001	1.01938
original_unpaid_balance	-1.07E-07	2.10E-08	-5.08	<.0001	1.04309
remaining_months_till_maturity	-0.00004271	1.29E-05	-3.32	0.0009	1.13015
rate_eligibility_degree2	-0.09371	0.000725	-129.25	<.0001	1.00997
LTV	0.00325	2E-05	162.53	<.0001	1.08556
burnout_factor	0.07878	0.000629	125.3	<.0001	1.02608

Table 3 – Modeling summary for Georgia, R^2=0.1923

Variable	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	-0.13918	0.00271	-51.38	<.0001	0
FICO	0.00001077	1.93E-06	5.59	<.0001	1.00703
CLTV	1.90E-07	5.93E-08	3.21	0.0013	1.00842
DTI	-0.00010331	3.44E-05	-3.01	0.0026	1.02842
original_unpaid_balance	-2.24E-07	1.42E-08	-15.71	<.0001	1.06014
rate_eligibility_degree2	-0.06089	0.000727	-83.74	<.0001	1.06316
LTV	0.00351	0.000021	166.97	<.0001	1.02157
burnout_factor	0.07607	0.000937	81.18	<.0001	1.02405

Table 4 – Modeling summary for Massachusetts, R^2=0.2492

[6] http://www.freddiemac.com/singlefamily/pdf/ex19.pdf. Fannie Mae uses the same sheet for adjusting interest rates.

[7] The peak of 0.3 is partially determined upon the length of rate drop cycle.

Variable	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	-0.10472	0.00665	-15.76	<.0001	0
FICO	0.0000037	1.85E-06	2	0.0457	1.00365
original_unpaid_balance	-3.55E-07	3.21E-08	-11.06	<.0001	1.03095
remaining_months_till_maturity	-0.00015252	2.02E-05	-7.55	<.0001	1.11204
rate_eligibility_degree2	-0.08185	0.00103	-79.49	<.0001	1.01738
LTV	0.00355	2.87E-05	123.86	<.0001	1.07432
burnout_factor	0.02858	0.000986	28.98	<.0001	1.02184

Table 5 – Modeling summary for Tennessee, R^2=0.208

We can collect some common insights from three models:

- Original LTV is the most important factor.
- Burnout factor and eligibility measure are good predictor. The two burnout-associated measures do not have a strong correlation.
- FICO, DTI and original unpaid balance have marginal impact on refinancing.

Other suggested variables were not retained by the regression process.

CONCLUSION

This paper uses net lift measure to capture the effects of interest rate drop on mortgage prepayment. The same framework can also be used for analyzing the impact of a rise in interest rate. The loan-level modeling can produce a better differentiation in prepayment speeds among borrowers than what a pool-level analysis will show. The framework suggested by this paper can also be used for analytics and modeling related to personal loans, auto loans, etc.

REFERENCES

Radcliffe, N.J. (2007). “Using Control Groups to Target on Predicted Lift: Building and Assessing Uplift Models”. Direct Marketing Analytics Journal. An annual Publication from Direct Marketing Association Analytics Council, Pages 14-21.
Medvedev, Y. (2018). “Personal Lending: What Is Customer Price/Credit Optimization? Is Experimental Design Inevitable to Optimize Price/Credit?” SAS Global Forum 2018, Paper 1774-2018.
Kubiak, R. (2012). “Net Lift Model for Effective Direct Marketing Campaigns”, SAS Global Forum 2012, Paper 108-2012.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at
Alec Zhixiao Lin
Email: alecindc@gmail.com
Web: www.linkedin.com/pub/alec-zhixiao-lin/25/708/261/

APPENDIX

The following is the SAS code used for running the study.

```
** Part I: Extracting and merging data **
```

```
** merge origination data with monthly performance data;
%macro shellyr;
%do year=1999 %to 2016;
%macro extract(qtr);
data morg.svcg_&qtr.&year;
infile "\\downloaded_data\historical_data1_time_&qtr.&year..txt"  dlm='|'
MISSOVER DSD lrecl=32767 firstobs=1 ;
input
ID_loan         :  $12.
Period          :   8.
Act_endg_upb    :   8.
delq_sts        :   $8.
loan_age        :   8.
mths_remng      :   8.
repch_flag      :   $1.
flag_mod        :   $1.
CD_Zero_BAL     :   $3.
Dt_zero_BAL     :   8.
New_Int_rt      :   8.
Amt_Non_Int_Brng_Upb : 12.
Dt_Lst_Pi          :  6.
MI_Recoveries   : 12.
Net_Sale_Proceeds : $14.
Non_MI_Recoveries : 12.
Expenses         : 12.
legal_costs      : 12.
maint_pres_costs : 12.
taxes_ins_costs  : 12.
misc_costs       : 12.
actual_loss      : 12.
modcost            : 12.
stepmod_ind  : $1.
;

data morg.orig_&qtr.&year;
infile "\\downloanded_data\historical_data1_&qtr.&year..txt" dlm='|' MISSOVER
DSD lrecl=32767 firstobs=1 ;

input
fico           : 8.
dt_first_pi    : 8.
flag_fthb      : $1.
dt_matr        : 8.
cd_msa         : 8.
mi_pct         : 8.
cnt_units      : 8.
occpy_sts      : $1.
cltv           : 8.
dti            : 8.
orig_upb       : 8.
ltv            : 8.
int_rt         : 8.
channel        : $1.
ppmt_pnlty     : $1.
prod_type      : $5.
st             : $2.
prop_type      : $2.
```

```
zipcode        : $5.
id_loan        : $16.
loan_purpose   : $5.
orig_loan_term : 8.
cnt_borr       : $2.
seller_name    : $30.
servicer_name  : $30.
flag_sc          : $1.
; run;

data morg.orig_&qtr.&year; set morg.orig_&qtr.&year;run;

proc sort data=morg.orig_&qtr.&year; by id_loan; run;
proc sort data=morg.svcg_&qtr.&year; by id_loan; run;

** Select single-family homes with 30-year fixed rates only;
data morg.orig_svcg_&year._&qtr;
merge morg.orig_&qtr.&year(in=s)
      morg.svcg_&qtr.&year(in=t);
by id_loan;
if t and s;

if orig_loan_term=360 and prop_type='SF';
run;
%mend;
%extract(Q1);
%extract(Q2);
%extract(Q3);
%extract(Q4);
%end;
%mend shellyr;

%shellyr;


** Stack all merged files;
data morg.performanceall;
set morg.orig_svcg:; run;

** Select single family homes with 30-year fixed rates in MA, TN and GA for
** chosen periods for our study;
%macro selst(state);
data morg.performanceall_&state._200901_201003
     morg.performanceall_&state._201207_201306
     morg.performanceall_&state._201307_201406
     morg.performanceall_&state._201501_201609;
set morg.performanceall;
where st="&state";

if 200901 le period le 201003 then output
morg.performanceall_&state._200901_201003;
if 201207 le period le 201306 then output
morg.performanceall_&state._201207_201306;
if 201307 le period le 201406 then output
morg.performanceall_&state._201307_201406;
if 201501 le period le 201609 then output
morg.performanceall_&state._201501_201609; run;
```

```
%mend;
%selst(GA);
%selst(MA);
%selst(TN);


** Part II: Obtain net lift measure **
** The following is to get net lift measure, not for modeling yet;
%macro chkperiod2(state, period1, period2, multiplier);
data two_files;
set morg.performanceall_&state._&period1(in=t)
    morg.performanceall_&state._&period2(in=s);

if t then prevailing_rate='High';
if s then prevailing_rate='Low'; run;

proc sort data=two_files; by period; run;
proc sort data=morg.market_rate_fixed30; by period; run;

data burnout;
merge two_files(in=t)
      morg.market_rate_fixed30;
by period;
if t;

if dti < 100 and cnt_units=1 and mi_pct=0;

burnout_temp1=max(Int_rt-market_rate, 0);
burnout_temp2=max(log(Int_rt/market_rate), 0);
burnout_temp3=max((Int_rt-market_rate)/market_rate, 0);

cltv_ltv_diff=cltv-ltv_orig_ltv;

ltv=(Act_endg_upb/orig_upb)*100;

** Add LLPA adjustments for interest rates;
if LTV le 60 then do;
   If fico < 660 then llpa_adj=0.5;
   else llpa_adj=0;
end;

else if LTV le 70 then do;
   if fico < 640 then llpa_adj=1.5;
   else if fico < 660 then llpa_adj=1.25;
   else if fico < 680 then llpa_adj=1;
   else if fico < 720 then llpa_adj=0.5;
   else llpa_adj=0.25;
end;

else if LTV le 75 then do;
   if fico < 640 then llpa_adj=3;
   else if fico < 660 then llpa_adj=2.75;
   else if fico < 680 then llpa_adj=2.25;
   else if fico < 700 then llpa_adj=1.25;
   else if fico < 720 then llpa_adj=1;
   else if fico < 740 then llpa_adj=0.5;
   else llpa_adj=0.25;
```

```
end;

else if LTV le 80 then do;
  if fico < 660 then llpa_adj=3;
  else if fico < 680 then llpa_adj=2.75;
  else if fico < 700 then llpa_adj=1.75;
  else if fico < 720 then llpa_adj=1.25;
  else if fico < 740 then llpa_adj=0.75;
  else llpa_adj=0.5;
end;

else if LTV le 85 then do;
  if fico < 660 then llpa_adj=3.25;
  else if fico < 680 then llpa_adj=2.75;
  else if fico < 700 then llpa_adj=1.5;
  else if fico < 720 then llpa_adj=1;
  else if fico < 740 then llpa_adj=0.5;
  else llpa_adj=0.25;
end;

else if LTV le 90 then do;
  if fico < 640 then llpa_adj=3.25;
  else if fico < 660 then llpa_adj=2.75;
  else if fico < 680 then llpa_adj=2.25;
  else if fico < 700 then llpa_adj=1.25;
  else if fico < 720 then llpa_adj=1;
  else if fico < 740 then llpa_adj=0.5;
  else llpa_adj=0.25;
end;

else if LTV le 95 then do;
  if fico < 640 then llpa_adj=3.25;
  else if fico < 660 then llpa_adj=2.75;
  else if fico < 680 then llpa_adj=2.25;
  else if fico < 700 then llpa_adj=1.25;
  else if fico < 720 then llpa_adj=1;
  else if fico < 740 then llpa_adj=0.5;
  else llpa_adj=0.25;
end;

else if LTV le 97 then do;
  if fico < 620 then llpa_adj=3.75;
  else if fico < 640 then llpa_adj=3.5;
  else if fico < 660 then llpa_adj=2.75;
  else if fico < 680 then llpa_adj=2.25;
  else if fico < 720 then llpa_adj=1.5;
  else if fico < 740 then llpa_adj=1;
  else llpa_adj=0.75;
end;

if llpa_adj=0 then eligi_tier=0;
else if llpa_adj le 0.5 then eligi_tier=1;
else if llpa_adj le 1 then eligi_tier=2;
else if llpa_adj le 2 then eligi_tier=3;
else eligi_tier=4;

current_burnout=(Int_rt/sum(market_rate, llpa_adj));
```

```
current_burnout1=max(log(Int_rt/(1.1*(market_rate+llpa_adj))), 0);

diff_ltv=ltv-orig_ltv;
run;

proc sql;
create table morg.bt_&state._&period1._period2 as select
id_loan,
prevailing_rate,
sum(burnout_temp1) as burnout1,
sum(burnout_temp2) as burnout2,
sum(burnout_temp3) as burnout3,
sum(current_burnout) as burnout4,
sum(current_burnout1) as burnout5,
avg(eligi_tier) as eligi_tier,
mean(ltv) as ltv,
mean(diff_ltv) as diff_ltv,
mean(cltv_ltv_diff) as cltv_ltv_diff,
max(loan_age) as loan_age
from burnout
group by id_loan, prevailing_rate; quit;

data morg.bt_&state._&period1._period2;
set morg.bt_&state._&period1._period2;

if prevailing_rate='Low' then do;
burnout1=burnout1*&multiplier;
burnout2=burnout2*&multiplier;
end; run;

proc sql;
create table shortmaturity_temp as select
id_loan, min(mths_remng) as mths_till_maturity
from two_files
group by id_loan; quit;

data shortmaturity;
set shortmaturity_temp;
if mths_till_maturity < 120; run;

** excluding loans ever reaching 90+ days in delinquency;
** Termination of these loans could due to bad performance, not due to
** voluntary prepayment;
proc sql;
create table delq_loans as select
distinct id_loan from two_files
where input(delq_sts, 10.) > 90 or flag_mod='Y'
      or sum(Non_MI_Recoveries, MI_Recoveries) gt 0
      or delq_sts='R' or CD_Zero_BAL in ('03', '06', '09'); quit;

data exclude_temp;
set shortmaturity
    delq_loans; run;

proc sort data=exclude_temp out=uniq_excl nodupkey; by id_loan; run;

** Select paid_in_full loans with good status;
```

```
proc sql;
create table paid_in_full_temp as select
distinct id_loan
from two_files
where CD_Zero_BAL='01' or Act_endg_upb=0
order by id_loan; quit;

data paid_in_full;
merge paid_in_full_temp(in=t)
      uniq_excl(in=s keep=id_loan);
by id_loan;
if t and not s; run;

proc sort data=two_files nodupkey out=uniq_loan; by id_loan; run;
proc sort data=paid_in_full nodupkey out=uniq_paid_in_full; by id_loan; run;

data morg.&state._&period1._&period2;
merge uniq_loan(in=t)
      uniq_paid_in_full(in=s)
      morg.bt_&state._&period1._period2;
by id_loan;
if t;

if fico > 850 then delete;

if s then prepaid=1; else prepaid=0;

if fico < 600 then fico_cap=600;
else if fico > 820 then fico_cap=820;
else fico_cap=fico;
fico_band=int(fico_cap/20)*20; run;

proc rank data=morg.&state._&period1._&period2 groups=15 out=check_burnout;
var burnout1 burnout2;
ranks rank_burnout1 rank_burnout2; run;

proc means data=check_burnout mean nway;
class prevailing_rate rank_burnout1;
var prepaid;
output out=compare_bt1(drop=_type_)
      mean=prepaid; run;

proc means data=check_burnout mean nway;
class prevailing_rate rank_burnout2;
var prepaid;
output out=compare_bt2(drop=_type_)
      mean=prepaid; run;

data compare_bt1_high compare_bt1_low;
set compare_bt1;

if prevailing_rate='High' then output compare_bt1_high;
else output compare_bt1_low; run;

data compare_bt2_high compare_bt2_low;
set compare_bt2;
```

```
if prevailing_rate='High' then output compare_bt2_high;
else output compare_bt2_low; run;

proc sort data=compare_bt1_high; by rank_burnout1; run;
proc sort data=compare_bt1_low; by rank_burnout1; run;
proc sort data=compare_bt2_high; by rank_burnout2; run;
proc sort data=compare_bt2_low; by rank_burnout2; run;

data net_lift_bt1;
merge compare_bt1_high(rename=(prepaid=prepaid_bt1_high))
      compare_bt1_low(rename=(prepaid=prepaid_bt1_low));
by rank_burnout1;
net_lift_measure1=prepaid_bt1_low-prepaid_bt1_high;
run;

data net_lift_bt2;
merge compare_bt2_high(rename=(prepaid=prepaid_bt2_high))
      compare_bt2_low(rename=(prepaid=prepaid_bt2_low));
by rank_burnout2;
net_lift_measure2=prepaid_bt2_low-prepaid_bt2_high;
run;

proc print data=net_lift_bt1; run;
proc print data=net_lift_bt2; run;

proc sort data=net_lift_bt1; by rank_burnout1; run;
proc sort data=net_lift_bt2; by rank_burnout2; run;

proc sql;
create table morg.merge_bt_ranks_&state
as select a.*,
          b.net_lift_measure1,
          c.net_lift_measure2
from check_burnout a
join net_lift_bt1 b
  on a.rank_burnout1=b.rank_burnout1
join net_lift_bt2 c
  on a.rank_burnout2=c.rank_burnout2;
quit;
%mend;
%chkperiod2(GA, 200901_201003, 201207_201306, 15/12);
%chkperiod2(MA, 200901_201003, 201207_201306, 15/12);
%chkperiod2(TN, 200901_201003, 201207_201306, 15/12);


** Part III: Regression **
** More data cleaning is also done here;
%macro model_st(state);
data net_lift_modeling_&state;
set morg.merge_bt_ranks_&state;
where ppmt_pnlty='N';  /* select loans with no prepayment penalty */
burnout=-sqrt(abs(burnout5-0.3));
orig_upb_trf=((orig_upb-112000)/1000)**2;

** No differentiation in performance for loans with FICO<690;

if fico < 690 then fico_trf=690;
```

```
else fico_trf=fico;

cltv_sq=cltv**2;

** Loans with < 2 months in loan age have little incentive for refinance;
if mths_remng > 358 then very_young_mortgage=1;
else very_young_mortgage=0;

if very_young_mortgage=0;
rate_eligibility=abs(int_rt-5.875);
run;

proc means data=net_lift_modeling_&state;
var fico
fico_trf
cltv
dti
orig_upb
ltv
loan_age
Int_rt
burnout
diff_ltv
orig_upb_trf
rate_eligibility
cltv_sq
prepaid; run;

** Final regression;
** Attributes with no or very low predictive power have been deleted;
proc reg data= net_lift_modeling_&state outest=estout;
model prepaid=
fico_trf
cltv_sq
dti
orig_upb_trf
mths_remng
rate_eligibility
ltv
burnout
/selection=stepwise sls=0.05 sle=0.05;
run;
%mend;
%model_st(GA);
%model_st(MA);
%model_st(TN);
```