# Mortgage refinancing as individual decision under uncertainty: how to build a reliable prepayment model

Dr. Vladimir Fishman and Dr. Anatoly Reynberg

Experian Decision Analytics

Abstract:

Prediction of mortgage loan prepayment on individual level can significantly improve valuation of mortgage-backed securities. Prepayment events can be triggered by different reasons, the most important of which are refinancing, relocations, and foreclosure. These events have different patterns and should be modeled separately. Usually, there is no explicit knowledge of the prepayment reason. However, prepayment events due to exogenous reasons can be identified indirectly.

Models with standard characteristics often exhibit insufficient predictive performance both out-of-sample and especially out-of-period. Our results will show how this performance can be improved by using a new approach which combines loan attributes with historical econometric data.

Since mortgage prepayment is an example of individual decision under uncertainty, it is one of the most suitable areas where ideas of mental accounting and cumulative prospect theory can be tested. We will show that in accordance with prospect theory individual decision to re-finance in most cases differs significantly from what can be considered as rational decision. Behaviour pattern is consistent with S-shape value function and demonstrates significant loss aversion effect. We will discuss how applying mental accounting ideas can increase predictive power and reliability of prepayment models.

We will show that that prospect theory can be used as a common ground for out-of-sample and out-of-period prediction and greatly improve prediction quality

Keywords: loan level prepayment prediction, mental accounting, cumulative prospect theory.

## Predictive models for individual mortgage prepayment: business objectives

Prepayment prediction on individual level can be an important tool for different financial applications. It can be used by mortgage originators for more accurate pricing, and can also help to build pools of mortgages with precisely tailored characteristics of interest rate sensitivity. From score building methodology perspective it is important for such models to predict individual financial decisions under uncertainty, and among other things mental accounting angle can be taken under consideration. In this paper we describe our attempt to enrich the traditional model building practices by application of the cumulative prospect theory by Tversky and Kahneman [1,2] to improve overall model quality.

## Data preparation for modeling process

For our experiments we used a dataset of actual individual mortgage data for a period of about 4.5 years. The last 1 year we put aside and used it for testing out-of- period prediction. The remaining data we randomly divided into two parts to analyze out-of-sample performance.

**Development sample**
June,1994 – Nov.,1997, inclusively
Training : 43,147 records
Validation : 18,463 records
**Out-of-period sample**
Dec., 1997 – Nov,. 1998, inclusively
8,364 records

Prepayment can be driven by a number of different reasons; most important of them are refinancing and relocation. During recent economic downturn mortgage defaults became important contributor as well. Our data obviously contains both types and there is no explicit knowledge of prepayment reason. These two events have different patterns and are influenced by different predictive characteristics. We will concentrate on refinancing events, for which mental accounting considerations can be applied. In order to filter out the relocation driven events we removed the cases for which prepayment is most likely due to relocation rather than refinancing, for example when original rate of the mortgage is lower than current mortgage rate offered by market. To make this filter more efficient we used so called NPV break-even threshold. If savings due to refinancing are equal to the refinancing cost, $c(r, r_0, t) = 0$, corresponding rate is called NPV break-even level. Decision to re-finance at break-even level effectively means that pre-payment option is given away for free. Therefore all prepayments above this level should be considered exogenous (primarily due to relocation).

## Modeling experiments and prediction quality evaluation

Our modeling exercises can be summarized as follows:

- Building predictive models using individual application level attributes
- Building predictive models with additional macroeconomic attributes
- Building predictive models using additional econometric history, and aggregated attributes responsible for rate variance, burnout factor, and seasonality

For all cases we conducted performance tests for out-of-sample and out-of-period prediction quality

Any model predicting binary event can be evaluated from the two different points of view- as a classifier, and as a predictor. In order to measure model performance as a classifier we used Gini index and Kolmogorov-Smirnov (KS) criterion.
In order to estimate model prediction quality the probabilities produced by the model should be measured against the actual rate, so the model should be properly calibrated.
Since behavior pattern may change as economic environment conditions evolve, building a model with reliable out-of-period prediction quality is especially tricky as we saw for many business cases.

The simplest approach to building a model for refinancing at the individual level is to use individual mortgage attributes and current mortgage rate as predictors.

**Mortgage attributes**

- Origination rate
- Mortgage amount
- Loan-to-value ratio
- Loan age
- Loan balance
- Total payment amt
- Escrow amt
- Curtailment ratio
- Cumulative curtail.
- Average pay pattern

However, as we will show, model quality can be significantly improved if econometric attributes (like inflation, unemployment, etc.) are taken into account as well.

**Macroeconomics**

- 10 yr rate
- 1 yr rate
- Mortgage rate
- Unemployment
- Inflation
- Home price index

It is logical to assume that individual decision of refinancing should be based on probabilities of outcomes. In case of prepayment it is probabilities of future mortgage rate variations. The actual probabilities however are unknown and according to the mental accounting paradigm we should assume that decision is made based on subjective estimate of the probabilities. And these subjective probabilities are influenced by prior mortgage rate behavior. Therefore history of mortgage rates and other econometric attributes and their volatility could be significant attributes in prediction of prepayment events. Also, it would be reasonable to expect that one of important factors is whether a borrower had similar or better opportunity to prepay in the past. If so, this fact should produce considerable burnout effect. So, minimal rate during life of the mortgage can be used as one of the attributes for modeling. To evaluate burnout effect we used a standard expression [3]:

$$Burnout_t = \sum_{\tau=\tau_{orig}}^{t} \max(\log(c / r_\tau),0),$$

where $c$ is coupon rate, and $r_\tau$ is refinancing rate available in period $\tau$.

We used the following aggregated measures based on common sense considerations.

**Aggregated attributes**

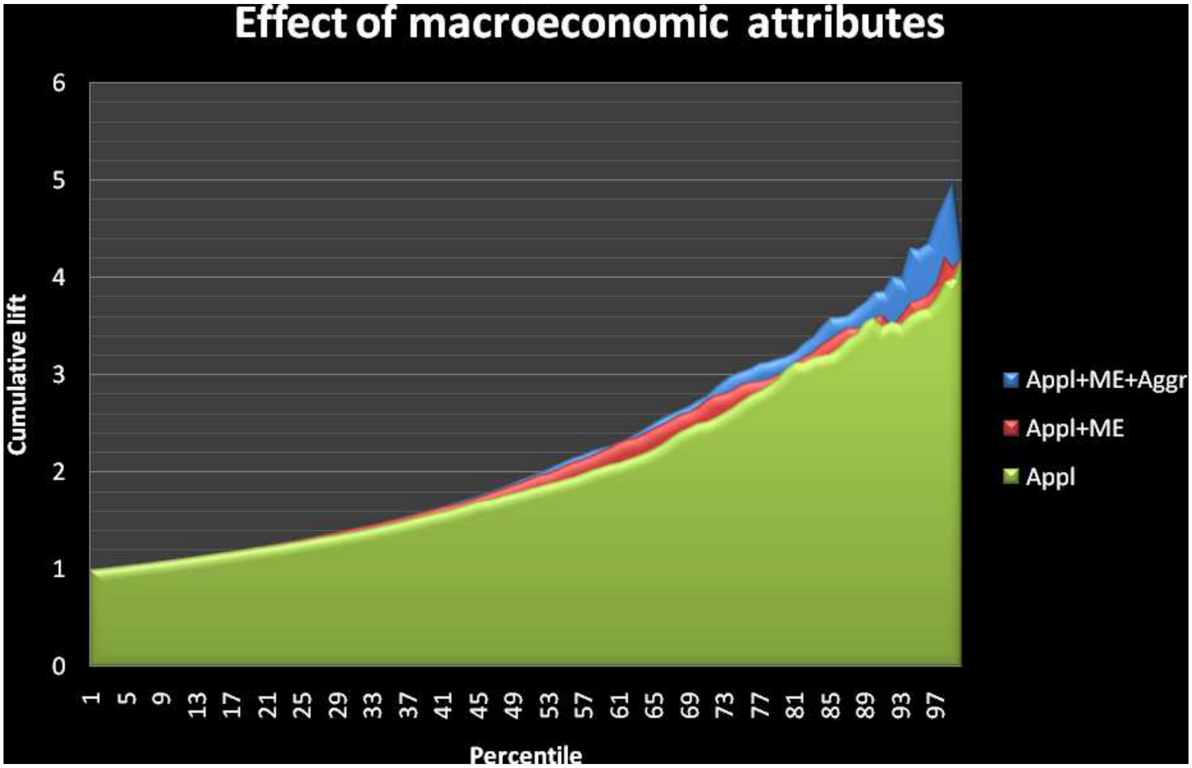
- 3m mtg. rate spread
- 1yr mtg. rate spread
- 1yr max rate
- 1yr min rate
- Lifetime min mtg. rate
- Current/lifetime spread
- Burnout
- Mtg. rate volatility
- Seasonality

Our experiments show that these characteristics indeed have high predictive power, besides, we noticed that slope of mortgage rate curve is an important predictor. For mortgage borrower it means that prepayment option is losing or gaining value and in accordance with so called disposition effect, subjective feeling of winning or losing influences the decision to exercise the option. We see all this as an encouraging sign that applying mental accounting approach could be promising.

To use economic history some sort of aggregation should be applied. Instead of using common sense to decide how to aggregate historical data we may use a neural network transactional modeling approach [4] and leave for the training algorithm to decide which form of aggregation should be used to maximize predictive power. The idea of transactional modeling approach is that the contribution of transactional history of different attributes like mortgage rate for example is controlled by a number of parameters, which effectively define the form of aggregation. These parameters (transactional weights) are unknown at first and will be determined along with other neural network weights within the same training procedure. In this particular case we tried both approaches. The results are similar and main outcome is that taking transactional history into account produces noticeable improvement.

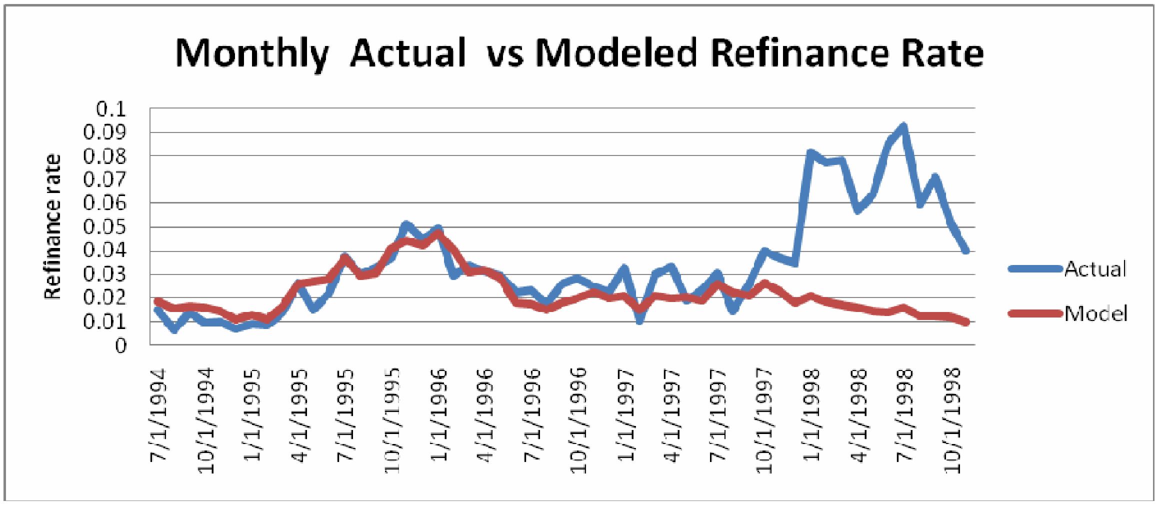
The following graph demonstrates how adding macroeconomic attributes in straightforward and aggregated forms affect the predictive power of prepayment model.

Figure 1




Still model quality is not good enough, especially out-of-period. We see that the model out-of-period prediction is not properly calibrated.

Figure 1




In the next section we will show how mental accounting consideration can be used to make the models more robust.

## Using mental accounting ideas to improve of prepayment models

Since mortgage prepayment is an example of individual decision under uncertainty, it is one of the most suitable areas where ideas of mental accounting and cumulative prospect theory can be tested. The area is even more suitable than stock trading strategy, because in stock trading financial institutions play more important role than individuals.

When a personal decision, which can cause future economic profit or loss, is made, the seemingly simplest choice is the one to maximize expected value of profit.

The fact that this is not a valid approach became clear more than two centuries ago. Bernoulli was the first who demonstrated by so called St. Petersburg paradox that even when expected wealth associated with a gamble is infinitely large the gamble still can be rejected because of risk.

Next fundamental theory was associated with names of von Neumann and Morgenstern. Their assumption was that individuals behave rationally in order to maximize subjective utility function which combines risk and return consideration. Since individuals are assumed to behave rationally the theory is supposed to be both descriptive and normative.

The major result in mental accounting which was awarded Nobel Price in recent years is cumulative prospect theory by Tversky and Kahneman.

They showed that descriptive theory and normative theory cannot be combined. No matter how individual utility function is chosen, people systematically violate the expected utility principle. And Tversky an Kahneman developed a convincing descriptive behavior theory which is consistent with human psychological nature.

Most important finding of prospect theory:

a. People evaluate the value of outcome with respect to a reference point, incremental value rather than absolute wealth is important (framing effect)

b. People are less willing to gamble with profit than with loss. Value function of prospect theory, which replaces classical expected utility function, is steeper for losses than for gains.
c. There is an asymmetry in how people perceive gains and losses (loss aversion effect).
d. Possible outcomes contribution to value function is proportional to decision weights which are not equal to probabilities. Based on prospect theory people tend to overweight small probability and underweight high probability.

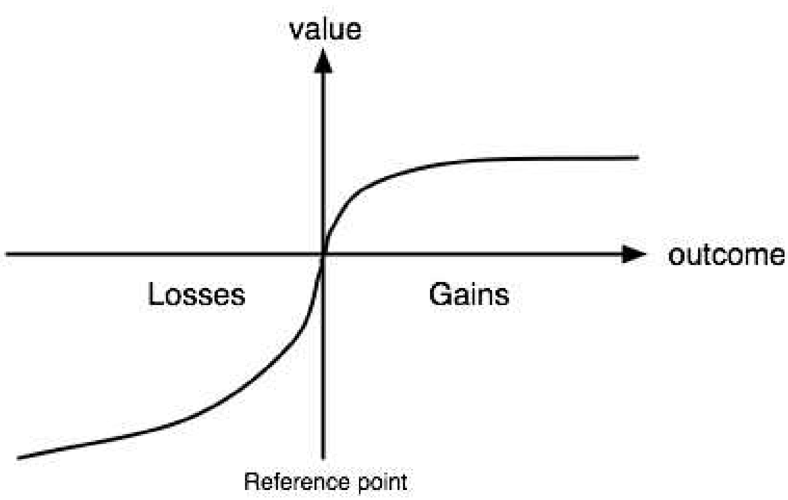
**Prospect theory value function formulation:**

In accordance with prospect theory the value function $V(W)$ for each outcome with wealth $W$ and reference level $R$ can be expressed as one of following S – shaped functions:

$$V(W) = \frac{(W-R)^{1-\gamma}}{1-\gamma} \qquad \text{if} \quad (W-R) \geq 0$$

$$V(W) = -\lambda \frac{(R-W)^{1-\gamma}}{1-\gamma} \qquad \text{if} \quad (W-R) < 0$$

Figure 3




If gamble can be subjectively evaluated as paying

$x_{-m} < x_{-m+1} < ... < x_{-1} < x_0 = 0 < x_1 < ... < x_n$
with corresponding probabilities
$p_{-m}, p_{-m+1}, ..., p_{-1}, p_0, p_1, ..., p_n$ ,
The value function for the gamble will have the following form:

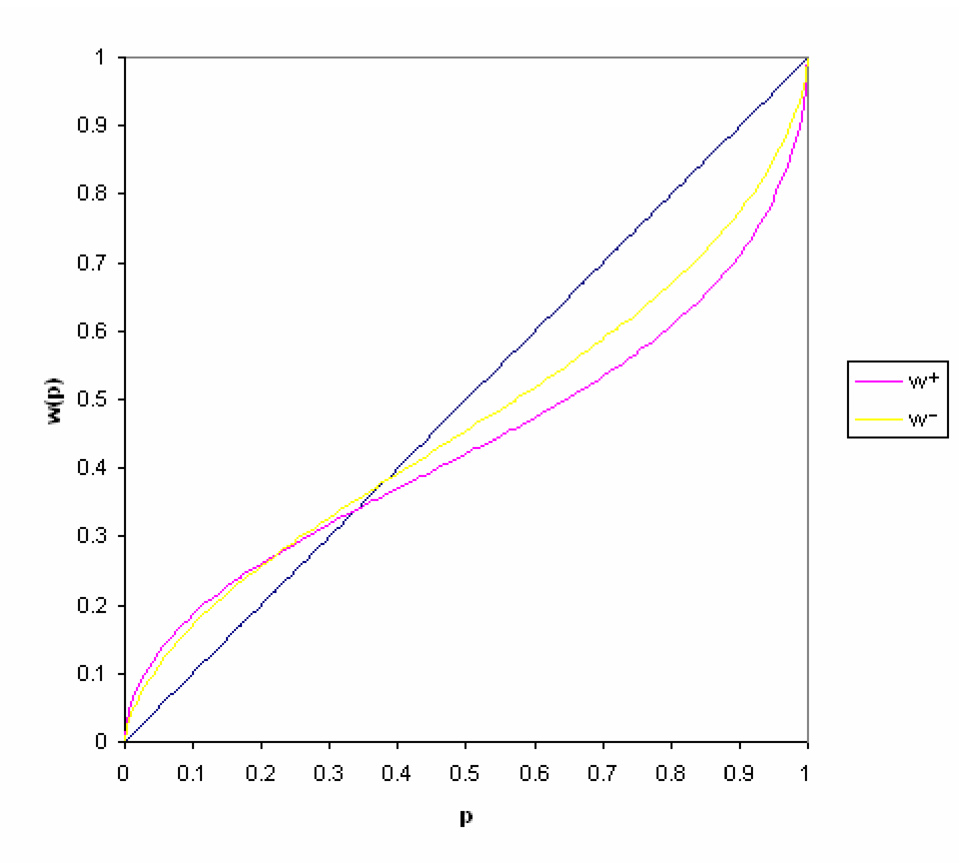
$U = \sum_{i=-m}^{n} \pi_i V(x_i)$, where $V(x_i)$ is S – shaped function as above and

$$\pi_i = w^+(p_i + ... + p_n) - w^+(p_{i+1} + ... + p_n) \quad \text{if} \quad 0 \le i \le n$$
$$\pi_i = w^-(p_{-m} + ... + p_i) - w^-(p_{-m} + ... + p_{i-1}) \quad \text{if} \quad -m \le i \le 0$$

Probability weighting functions $w^-(p)$ and $w^+(p)$ capture subjective perception of probability and can be expressed as following:

$$w^+(p) = \frac{p^{\alpha^+}}{(p^{\alpha^+} + (1-p)^{\alpha^+})^{1/\alpha^+}}$$
$$w^-(p) = \frac{p^{\alpha^-}}{(p^{\alpha^-} + (1-p)^{\alpha^-})^{1/\alpha^-}}$$
$$0 < \alpha^+ \le 1, \; 0 < \alpha^- \le 1$$

Figure 4



## Using prospect theory to improve out-of-period prediction

From cumulative theory perspective the gamble can be described in a following simple way (see Appendix A for mortgage accounting formulae and notations):

A mortgage holder faces the following alternative:

1. Do nothing which we consider a reference point
2. Refinance. Then if rates change, the resulting gain/loss is:

$$D = c(r, r_0, t) - \max(c(r + dr, r_0, t + dt) - \sum_{i=1}^{i=i_{\max}} (w - v)(1 - \rho) B_{t_i}, 0)$$

The sum expression is the amount the borrower would save by refinancing because of paying less interest regardless of outcome.
If rates go up the result is a gain, otherwise it can be a loss. For simplicity we considered only two outcomes: rates up and rates down. The magnitude of the change was evaluated based on observed mortgage rate volatility.

In accordance with prospect theory the value function for each outcome with wealth $W$ and reference level $R$ can be expressed as one of following S – shaped functions:

$$V(D) = 1 - e^{-\gamma(W-R)} \qquad \text{if} \quad (W - R) \geq 0$$
$$V(D) = \lambda(e^{\gamma(W-R)} - 1) \qquad \text{if} \quad (W - R) < 0$$

or

$$V(D) = 1 - \frac{D^{1-\gamma}}{1-\gamma} \qquad \text{if} \quad D \geq 0$$

$$V(D) = -\lambda \frac{(-D)^{1-\gamma}}{1-\gamma} \qquad \text{if} \quad D \geq 0$$

Commonly used probability weighting function:

$$W^+(p) = W^-(p) = \frac{p^\alpha}{(p^\alpha + (1-p)^\alpha)^{1/\alpha}} \qquad 0 < \alpha \leq 1$$

The resulting value of the game is:

$$U = V^+ W^+(p^+) + V^- W^-(p^-)$$

Before evaluating parameters of prospect theory we applied (a non-supervised) segmentation. The idea was that people with different level of wealth may have different

patterns of prepayment behavior. We did not have individual wealth among available attributes, so we used home value and loan amount as a proxy. We built 5 segments using tree segmentation technique (CHAID).
Parameters used in prospect theory, primarily $\lambda$, (the results show low sensitivity to $\alpha$ and $\gamma$) can be evaluated for each segment of dataset.
Table 1

| **Segm. #** | **# of records** | **Condition** | **Avg. appraisal value** | **Avg. Mtg.Amt** | **Avg. Home Value** | **Avg fam income** | **Prosp. Theory λ** |
|---|---|---|---|---|---|---|---|
| 1 | 10,927 | M<50000 & V<138400 | 31,159 | 32,805 | 71,500 | 37,640 | **3.9** |
| 2 | 9,618 | M<87100 & V>=138400 | 100,075 | 55,640 | 169,000 | 50,400 | **2.25** |
| 3 | 13,261 | 50000<=M< 87100 & V<138400 | 77,857 | 67,436 | 84,870 | 42,860 | **1.5** |
| 4 | 21,522 | M>=87100 & V<253500 | 175,511 | 132,933 | 144,720 | 54,990 | **1.18** |
| 5 | 6,265 | M>=87100 & V>=253500 | 262,727 | 172,473 | 273,760 | 62,920 | **1.33** |

M- Mortgage Amount (from application)
V- Area Average Home Value (from associated demographic data)

Overall the strategy of applying prospect theory to improve model quality can be summarized as following:

- Build population segments for development sample using tree segmentation technique (CHAID).
- Evaluate coefficients of value function for each population segment
- Build in-sample model
- Apply the model for out-of-period data and evaluate predicted response rate for each segment

- Use value function coefficients to evaluate out-of-period probability of prepayment for each population segment
- Apply scaling coefficients to make model output consistent with prospect theory prediction

Then we can compare out-of-period degradation for these two cases. As we see, for out-of-period sample the model works reasonably well as a classifier without using prospect theory approach, but it fails to predict actual probabilities.
The results show that if prospect theory is used as a common ground for out-of-sample and out-of-period prediction, the predictive power can be drastically improved.

Figure 5

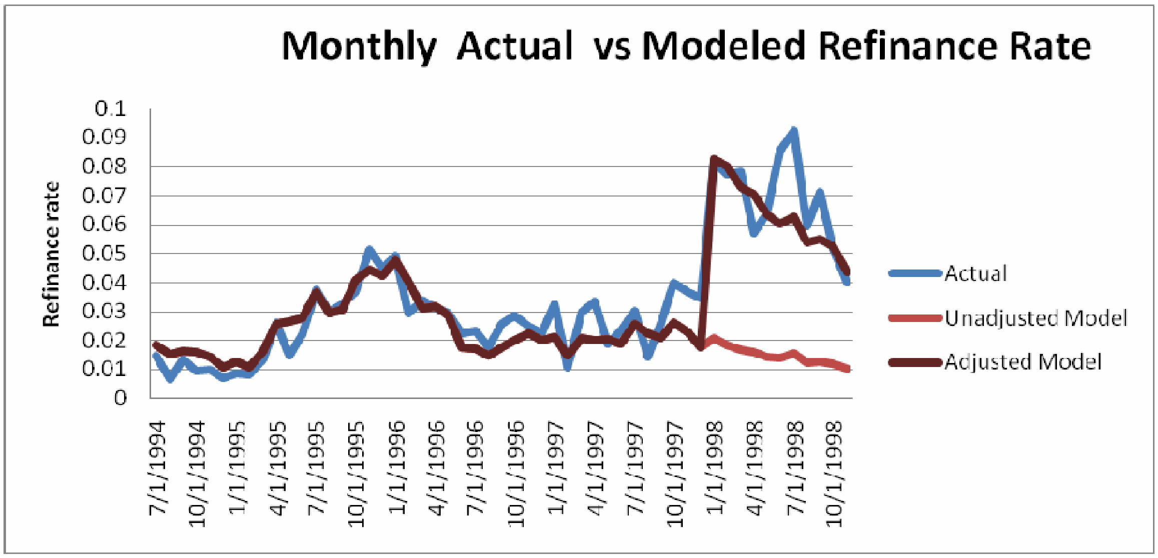



Table 2

| | **Training subset** | | **Out-of-sample validation** | | | **Out-of-period validation** | |
|---|---|---|---|---|---|---|---|
| | Actual | Model | Actual | Model | Actual | Model | Adjusted |
| Sample size | 43,147 | | 18,463 | | | 8,364 | |
| Refinance rate | 0.0242 | 0.0241 | 0.0247 | 0.0234 | **0.0635** | **0.0132** | **0.0607** |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Gini (%) | | 68.1 | | 67.2 | | **54.1** | **53.6** |
| KS (%) | | 57.0 | | 56.3 | | **44.8** | **43.4** |

## Rational prepayment versus actual prepayment

It is interesting to analyze to what extend actual prepayment differs from what can be considered as rational. To determine if it is rational to refinance is a complex mathematical problem which is equivalent to pricing of a call option. The value of option is based on assumptions regarding underlying interest rate and mortgage rate stochastic processes and risk adjustment coefficients. A standard way to solve the problem is using lemma Ito to convert the problem to a set of partial differential equations, which can be solved using appropriate boundary conditions (see for example [5]). The process is computationally intensive. The results can differ depending on choice of underlying processes and risk adjustment parameters but not very significantly, and it is interesting to analyze if refinancing strategy based on these results is consistent with actual behavior. We used some recent results which allow for some simple underlying process to express the rational prepayment condition in closed form [6].

$$r - r_0 \le x^* \equiv \frac{1}{\psi}(\phi + W(-\exp(-\phi)))$$

Where:

$$\psi = \frac{\sqrt{2(i+\lambda)}}{\sigma}$$

$$\phi = 1 + \psi(i+\lambda)\frac{C}{B_t(1-\tau)}$$

$$\lambda = p_{reloc} + \frac{r_0}{\exp(r_0 T) - 1} + i_{\inf}$$

Simplified version:

$$x^* = -\sqrt{\sqrt{2(i+\lambda)}\,\frac{\sigma C}{B_t(1-\tau)}}$$

Here $W(.)$ is the Lambert W-function,
the inverse function of the function $f(w) = we^{w}$

$$z = W(z)e^{W(z)}$$

As a crude estimate we can use:

Standard deviation of mortgage rate: z (or evaluate based on actual data)
Tax rate: $\tau = 0.28$

Refinancing cost: $C = 0.01B_t + \$2000$
Exogenous rate: $p_{reloc} = 0.15$
Inflation: $i_{\inf} = 0.03$

Figure 6

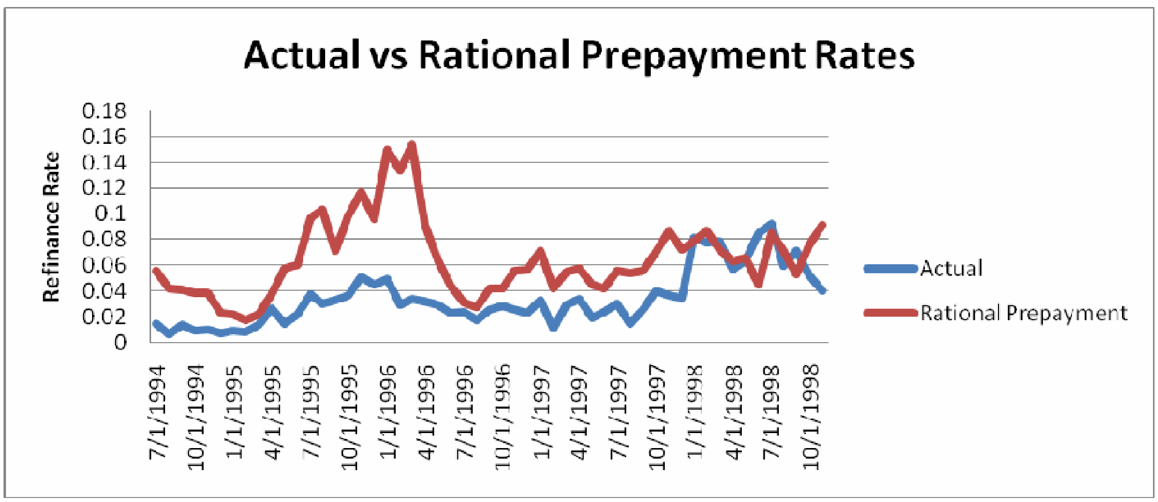



Table 3

| | | Rational | prepayment | | | |
|---|---|---|---|---|---|---|
| | # of | Threshold | Prediction | Prediction | Actual | Actual |
| | records | (basis points) | # | rate | # | rate |

| | | | | | | |
|---|---|---|---|---|---|---|
| In-period | 61,593 | 251 | 3,908 | **0.063** | 1,486 | **0.024** |
| Out-of-period | 8,352 | 195 | 584 | **0.070** | 519 | **0.062** |

Our results show that pattern of actual prepayment differs significantly from rational decisions.

## Conclusion and future research

We can summarize our conclusions as follows:

1. Econometric history is important to predict prepayment behavior
2. Still models often do not produce reliable prediction for out-of-period prepayment rate
3. Using prospect theory can be used as a common ground for out-of-sample and out-of-period prediction and can significantly improve prediction quality
4. The results show that actual behavior pattern differs significantly from the results of applying a normative theory, which can be considered as rational behavior

It would be interesting in the future to apply this approach to the recent data, and also to analyze "optional default" events which are common for recent housing market downturn, as well as to use the method in other areas of predictive modeling.

Decision to refinance is a decision to exercise an American option. If rate goes up the option value is increasing and option holder is gaining money, if rate is down holder is losing money. In case of stock trading behavior there is noticeable disposition effect: loss aversion is higher when investor is losing money. If mental accounting in mortgage prepayment case follows a similar pattern, the value of $\lambda$ when rates are down should be lower than when rates are up. If a dataset includes periods of rising and falling mortgage rates it would be possible to analyze this effect.

**References:**

[1] Amos Tversky, and Daniel Kahneman, 1992. Advances in Prospect Theory: Cumulative Representation of Uncertainty. Journal of Risk and Uncertainty, **5**(4), 297–323.

[2] Daniel Kahneman, and Amos Tversky, 2000. *Choices, Values, and Frames*. Cambridge: Cambridge University Press.

[3] Michael La-Cour-Little, Michael Marschoun and Clark L. Maxam 2002

Improving Parametric Mortgage Prepayment Model with Non-parametric Kernel Regresssion

JRER Vol. 24 No. 3 299-327

[4] Vladimir Fishman; Yuri Galperin; Anatoly Reynberg 2005
Method of building predictive models on transactional data. 2005-03-29: US Patent 6873979

[5] Chen, Yong; Connolly, Michael; Tang, Wenjin; Tang; Su, Tie;
The Value of Mortgage Prepayment and Default Options
The Journal of Futures Markets, Vol. 29, No. 9, 840–861 (2009)

[6] Sumit Agarwal, John C. Riscoll, David I. Laibson 2009
Optimal Mortgage Refinancing: A closed Form Solution
http://www.economics.harvard.edu/faculty/laibson/files/Mortgage%20refinancing.pdf

## Appendix A

### Summary of mortgage accounting

We consider fixed rate mortgages with
Maturity: $T$
Rate: $r_0$
Current age: $t$
Current market rate: $r$
Outcome of exercising prepayment option: $c$
Current monthly payment: $w$
Monthly payment after re-financing: $v$
Outstanding balance at time $t$ : $B_t$
Closing cost: $C = eB_t + C_0$

Current monthly payment

$$w = \frac{r_0}{1-(1+r_0)^{-T}}$$

Monthly payment after re-financing

$$v = \frac{r}{1-(1+r)^{-T}}$$

Outstanding principal amount

$$B_t = \frac{1-(1+r_0)^{-(T-t)}}{1-(1+r_0)^{-T}}$$

Outcome of exercising prepayment option

$$c(r, r_0, t) = [w/v - (1+e)]B_t - C_0$$

Gain due to lower rate can be corrected to include tax considerations and limited expected life of loan until exogenous prepayment (relocation, etc)

$$c(r, r_0, t) = [(w/v - 1)(1-\rho) + e)]B_t - C_0$$

Where $\rho$ is taxation rate