Order Number 9108143

A comparative analysis of empirical mortgage prepayment models: The GNMA experience

Lucy, Robert P., Ph.D.

City University of New York, 1990

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

A

A COMPARATIVE ANALYSIS OF EMPIRICAL MORTGAGE PREPAYMENT MODELS: THE GNMA EXPERIENCE

by

ROBERT P. LUCY

A dissertation submitted to the Graduate Faculty in Economics in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York.

1990

This manuscript has been read and accepted for the Graduate Faculty in Economics in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

6/25/90
Date

Chair of Examining Committee

6/25/90
Date

Michael Grossman
Executive Officer

Michael Grossman

Ronald Anderson

Salih Neftci
Supervisory Committee

The City University of New York

Abstract

A COMPARATIVE ANALYSIS OF EMPIRICAL MORTGAGE PREPAYMENT MODELS: THE GNMA EXPERIENCE

by

Robert P. Lucy

Adviser: Professor Ronald W. Anderson

Most mortgage debt carries with it the right of the mortgagor to pay off the loan in whole or in part prior to maturity, sometimes without penalty. It is this prepayment option of the mortgagor that makes the valuation of mortgages and mortgage-backed securities so difficult since rational pricing must explicitly account for any expected prepayment behavior. This thesis has as its main objective the estimation, interpretation, and comparison of prepayment probabilities from two distinct types of empirical models, the proportional hazard model and the aggregate logit model. Both address the following specific question: what is the conditional probability of a mortgagor prepaying his mortgage given the current set of exogenous factors that are thought to influence such behavior? The models are estimated using monthly observations of outstanding principal on GNMA pools. The proportional hazard model is found to fit the GNMA prepayment data quite poorly whereas the aggregate logit model is found to provide a more than adequate fit. The estimated aggregate logit models predict that the probability of prepayment is directly related to the financial incentive to refinance a mortgage by either of two measures although the responsiveness to either measure is not constant over time. In addition, prepayments are shown to react in a nonsymmetrical way to changes in the market interest rate above and below the contracted rate, they exhibit a non-monotonic pure-aging effect, there appears to be seasonality in prepayments, and prepayments seem to be counter-cyclical. In addition, prepayments are estimated to increase in a volatile interest rate environment. This last result is contrary to the predictions of standard option pricing theory.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF PLOTS

CHAPTER I

INTRODUCTION

"Is this security properly valued?" This is the single most important question that needs to be answered by participants in the markets for residential mortgages and mortgage-backed securities. Residential mortgages have traditionally constituted the largest single component of private indebtedness in the United States. In 1984, total residential mortgage debt outstanding was nearly three times as great as total corporate debt outstanding.[1] Most of this debt carries with it the right of the mortgagor to pay off the loan in whole or in part prior to maturity, sometimes without penalty. It is this prepayment option of the mortgagor that makes the valuation of mortgages so difficult since rational pricing must explicitly account for expected prepayment behavior.

This thesis has as its main objective the estimation, interpretation, and comparison of prepayment probabilities from two distinct types of empirical models, the proportional hazard model and the aggregate logit model. These models will address the following specific question: what is the conditional probability of a mortgagor prepaying

---

[1]See Fabozzi [1985].

his mortgage given the current state of the economy as reflected by current interest rates and other relevant variables? An answer to this question will greatly enhance our ability to assess value in the multi-billion dollar mortgage market.

It is apparent that the soundness of the banking system depends in part on the ability of banks and thrifts to manage the risks associated with the possible prepayment of a large part of their mortgage assets. A variety of derivative securities have come into existence that allow the reallocation of this prepayment risk away from the depository institutions and onto a broader range of potential investors. The most important of these securities is the mortgage-backed pass-through. A pass-through security is issued when a mortgage banker or other financial institution creates a pool of mortgages and then sells securities that give the investor a pro-rata share in the cash flow generated by the pool. The issuer essentially "passes through" to the investors the scheduled monthly principal and interest payments it collects on the underlying mortgages as well as any principal prepayments and proceeds from foreclosed loans. Note that the issuer of the pass-throughs treats the sale of these securities as a sale of assets; the underlying debt obligation remains that of the underlying mortgagors collectively. The issuer may or may not be the originator of the underlying mortgages but the issuer typically does assume the servicing of the underlying loans. As such, a servicing fee is usually recovered by the issuer before the principal, interest and prepayments are passed on to the investors.

The most important type of pass-through in the mortgage-backed securities market in terms of dollar value outstanding is the Government National Mortgage Association (GNMA) pass-through or "Ginnie Mae". The Ginnie Mae pass-through is also of primary interest to this thesis since the empirical estimation of mortgage prepayment probabilities will be conducted using monthly observations of outstanding principal on these pools. Further specific characteristics of Ginnie Maes and the underlying pools of mortgages will be discussed in Chapter III.

Ginnie Mae pass-throughs are traded in a market that consists of a primary market centered around GNMA auctions and a secondary market similar to the secondary markets for Treasuries and corporate bonds. Dealers will either quote prices or yields for GNMA pass-throughs. Yields are of course more convenient in that they allow investors to make easier comparisons across pools which differ with respect to principal outstanding[2] and to the contracted rate of interest on the underlying loans. However, such yield calculations are often based on simplifying assumptions about prepayments that are almost never met by any given Ginnie Mae pool in practice. One standard yield calculation

---

[2]Ginnie Mae pass-throughs are, like the underlying mortgages, fully amortizing securities.

is based on the assumption that the pool is always newly originated and that the principal will be repaid on schedule for twelve years after which time the pool will be fully prepaid. This twelve year average life assumption is based on past Federal Housing Administration (FHA) experience with mortgage terminations.[3] As would be expected, actual experience with most Ginnie Maes has been quite different than the FHA experience, especially for those pools issued in the early 1980's when interest rates were very high. Many, but not all, of the mortgages in these pools were prepaid when interest rates began to fall in the mid 1980's as mortgagors refinanced their loans at lower rates of interest.

Within the industry another method of taking prepayments into account when calculating the yield is to assume that for every month until maturity a constant proportion of outstanding mortgages will prepay. This constant conditional prepayment rate (CPR) can be calculated using last month's prepayment rate, a moving average of past prepayment rates, or some other transformation thereof. In this approach, the yield of the pass-through is obtained by modifying standard formulae for amortizing assets to allow for the projected CPR. This method also suffers from several defects. First, the CPR is often calculated on a "generic" basis; when calculating the CPR, pools are often

---

[3]Historical FHA data shows that a 30 year mortgage "terminates" (defaults,prepays), on average, after 12 years.

grouped into same-interest rate cohorts where differences in term to maturity, among other things, are ignored.[4] Second, there is some reasonably strong evidence that monthly prepayment rates for individual pools are not "consistent".[5] That is, the historical prepayment behavior of a GNMA pool is a poor indication of its future prepayment behavior. Fast-prepaying pools of the past are not necessarily the fast-prepaying pools of the future. Third, this simple method of extrapolating the past into the future is insensitive to the possible future state of the economy such as represented by the level of interest rates.

In recent years, dealers have begun to experiment with yield quotes based on various assumptions about the future path of prepayments and the relationship of this path to economic variables such as interest rates. It is obvious that such yield quotes must be interpreted with caution until prepayment behavior is better understood.

There are other important reasons to study prepayments, also from the perspective of the decisions of investors and dealers in the mortgage market. The investor does not know the exact proportion of a security that will be redeemed by any given date. Having a clearer understanding of prepayment behavior could help in formulating asset portfolios better designed to match the investor's desired holding period. For dealers a basic problem with Ginnie

---

[4]See Kidder Peabody [1985] for an example of this.

[5]See Senft [1985], page 540.

Maes is that it is difficult to determine the appropriate hedge with which he can isolate himself from major movements in interest rates. A side benefit of the more accurate valuation that comes from understanding prepayments would be to give the appropriate Treasury security hedge ratio for Ginnie Maes.

Academic interest in the issue of mortgage prepayments had been limited until relatively recently.[6] This perhaps reflected a lack of complete understanding of the fact that Ginnie Maes and in fact all mortgages and mortgage-backed securities differ significantly from other fixed income securities such as Treasuries. It is useful to compare Ginnie Maes with Treasury bonds in order to understand the former's distinctive features and consequent valuation difficulties. Both Ginnie Maes and T-bonds are long-term interest bearing securities which trade at market prices that vary inversely with the level of interest rates. For both, the interest and principal are guaranteed by the U.S. government and they are considered default risk-free.[7] The two differ in that Ginnie Maes pay interest and principal monthly whereas T-bonds pay a semi-annual coupon only. A much more important difference is, of course, that the Ginnie Mae may be prepaid fully or in part at any time prior to its maturity. In contrast, a T-bond is callable at most

---

[6]Relevant literature will be reviewed in Chapter II.

[7]See Chapter III for a full discussion of the default risk-free nature of Ginnie Maes.

one time prior to its maturity. In addition, the rate of prepayment of a Ginnie Mae depends upon the rate of prepayment of the mortgages in the pool it is based upon. Consequently, the prepayment experience can differ from one pool to another even if they carry the same coupon interest rate and the same term to maturity. Even though they have often been regarded as fixed income securities, Ginnie Maes are in fact titles to random nominal cash payments which may be highly differentiated across pools.

The Ginnie Mae pass-through can be viewed by the investor as a portfolio consisting of long positions in fully amortizing default risk-free bonds plus short positions in implicit call options. A Ginnie Mae pool is particularly sensitive to the exercise of the call options since the FHA mortgages which back it up carry no prepayment penalty.* An approach to valuing such securities in modern finance would be to value the option components explicitly while leaving the bond components to be priced off the Treasury yield curve. An understanding of prepayment behavior will significantly enhance the ability to value the call options.

The remainder of this thesis is organized as follows. Chapter II summarizes the academic literature that has been concerned with mortgage prepayments, both the theoretical and empirical research. Chapter III provides a complete description of the GNMA data that will be used in the

---

*See Chapter III.

empirical analysis. Chapter IV describes the variables that can influence prepayment behavior. Chapter V presents the first of the two general classes of empirical models that will be estimated and compared in this thesis, the proportional hazard model. Chapter VI presents the other, the aggregate logit. Chapter VII will present the results of the empirical analysis. Chapter VIII will offer conclusions.

CHAPTER II

LITERATURE REVIEW

The starting point for theoretical research on prepayments and the valuation of mortgage-backed securities falls within the general framework for valuing callable bonds.[1] In a world with perfect mortgage markets and without transactions costs and default risk and in which mortgagors strictly maximize their wealth, a fixed-rate[2] mortgage is prepaid if and only if the current market interest rate falls below the contracted rate. The fundamental problem with this result is that in the real world, mortgage prepayments (i) sometimes occur when market rates are above contracted rates and (ii) often do not occur when market rates are below contracted rates.

The rationalization of (i) is straightforward. When a mortgagor sells his home and moves to a new residence, he sometimes prepays the existing mortgage. It may be that he has a contractual obligation to prepay since some mortgages have a "due on sale" clause built in. These clauses give

---

[1]See Brennan and Schwartz [1977] for a general model for valuing callable bonds.

[2]Hereinafter, mortgages are assumed to be level payment, fixed-rate mortgages. GNMA pools contain fixed-rate mortgages. See Chapter III.

the lender the option of collecting the face value of the mortgage upon the sale of the residence. Of course this option will be exercised only when the market rate is above the contract rate. When there is no binding due on sale clause, the seller will still end up prepaying if the mortgage is not assumed by the buyer. There is no guarantee that the buyer will assume a below-market rate mortgage in a world where there are transactions costs and imperfect markets for second mortgages.[3] The main point is that prepayments are not solely caused by mortgagors who refinance at lower rates of interest. Prepayments also occur when households move and mortgages are not assumed. Such prepayments are sometimes referred to as "uneconomic" although it is often perfectly rational for the mortgagor to move and for the buyer to forego assumption of the existing below-market rate mortgage.

The rationalization of (ii) is just as direct. Prepayments do not always occur when current rates fall below contract rates because of the presence of significant transactions costs associated with refinancing. These costs include the points payable up front in obtaining a new mortgage loan and any other fees, charges and incidental costs, including prepayment penalties when relevant.

The impact of transactions costs on the interest rate differential needed to justify refinancing is analyzed by

---

[3]Hendershott, Hu and Villani [1983] discuss the determinants of the buyer's decision to assume an existing mortgage under perfect certainty.

Follain and Tzang [1988] under conditions where future interest rates are known with certainty. Their basic conclusion is that refinancing should occur whenever the present value of interest savings is greater than total transactions costs. In this situation, the "net benefit" to refinancing immediately is positive. For any given interest rate differential, they show that the net benefit will be larger the longer the expected tenure of the mortgagor in the residence.

In a world where future interest rates are uncertain, prepayment may not occur when the net benefit is positive because the mortgagor may feel that it will pay to wait for further declines in interest rates. This "value of waiting" before refinancing is discussed theoretically by Siegel [1984] and Follain, Scott and Yang [1988]. They both conclude that, all else equal, the value of waiting is positively related to the volatility of interest rates. This conclusion is similar to a basic proposition of standard option pricing theory where the value of a call option is directly related to interest rate volatility.[4] Both studies also argue that, all else equal, the value of waiting is positively related to the mortgagor's expected tenure in the residence. Note, however, that when the mortgagor's expected tenure is relatively long, the net benefit to refinancing <u>immediately</u> is also relatively great. There are two opposing forces at work in this situation.

---

[4]See, for example, Black and Scholes [1973].

Further complications introduced by interest rate uncertainty are such that no simple closed-form formula for calculating the critical interest rate differential is readily available. One solution to this problem has been to use a form of the binomial option pricing model to value the mortgagor's prepayment option directly. This type of research has been undertaken by Hall [1985] and Follain, Scott and Yang [1988].

The pioneering research on pricing mortgage-backed securities is that of Dunn and McConnell [1981a,b]. They present a model for valuing Ginnie Maes. Specifically, they introduce a constant probability of "uneconomic" prepayment each month. Their approach is not wholly satisfactory, however, in that "economic" prepayment still is immediate whenever market rates drop below the contract rate. Furthermore, even when the market rate exceeds the contract rate, the probability of prepayment is not likely to be a constant as they assume but will vary with a variety of economic and demographic considerations.

Dunn and Singleton [1983] develop an alternative method of pricing Ginnie Maes based on a general consumption based asset pricing model. They point out that a complete pricing model would require an explicit formulation of the stochastic process for prepayments. Rather than attempt this they employ a rational expectations methodology which does not need to specifically model prepayments.

Schwartz and Torous [1989] present a continuous time interest-contingent claim model for pricing mortgage-backed securities that explicitly integrates an empirically estimated prepayment function into the valuation framework. Their method of treating prepayments is an alternative to the optimal call condition imposed by Dunn and McConnell. It emphasizes the importance of understanding the empirical determinants of prepayment behavior.

Empirical studies of mortgage prepayment behavior have been conducted by Curley and Guttentag [1974], Peters, Pinkus, and Askin [1984], Arak and Goodman [1985], Navratil [1985], Green and Shoven [1986], Quigley [1987], Jacob, Lord, and Tilley [1987], Giliberto and Thibodeau [1988], and Schwartz and Torous [1989]. Most of these studies use mortgage pools or aggregate data of some kind to estimate the conditional probability of prepayment given the variables thought to influence such behavior. Most concentrate in one way or another on the relationship between interest rate movements and prepayments.

The first reported attempt to empirically model the prepayment process appears to be that of Curley and Guttentag. They used ordinary least squares (OLS) to regress the logarithm of FHA annual termination rates against the average annual interest rate differential, the "points" that had prevailed on average during the year, and a variable that measured the age of the mortgages relative to their original term to maturity. The estimated parameters for the interest rate differential and points were highly significant in the expected direction. The also found that terminations increased with the age of the mortgages. The model was able to explain 87 percent of the variability in the dependent variable. The authors used their estimated model to calculate mortgage termination probabilities under a few simple scenarios for future interest rates. These probabilities were then used to calculate the expected yield on a mortgage.

Peters, Pinkus, and Askin grouped 503,000 conventional mortgages into cohorts based upon year of origination, geographic region, and half point interest rate intervals. They then calculated the prepayment rates for each cohort and regressed these rates using OLS on a number of explanatory variables including the interest rate differential, macroeconomic variables such as growth in GNP, and regional dummy variables. They find a highly significant relationship between the interest rate differential and prepayment rates in the expected direction. The other explanatory variables that were used are also found to be significant. The linear probability model that they use, however, can lead to some well known estimation and forecasting problems.[3]

Arak and Goodman use a similar methodology in their work. Their prepayment data, however, comes from GNMA pools. They find that (a) positive and negative interest rate differentials do not affect prepayment rates

---

[3]See Pindyck and Rubinfeld [1981].

symmetrically, (b) prepayment rates first increase with age but later decline, and (c) macroeconomic variables such as the unemployment rate and the aggregate volume of home sales add little explanatory power to the regression.

Navratil analyzes one year (September 1982 through August 1983) of monthly GNMA prepayment rates generically grouped by contract interest rate. To avoid the problems associated with the linear probability model, he uses a form of the aggregate logit model for his estimation.* He also finds that prepayment rates are increasing in the difference between the contract rate and the current rate. Furthermore, the sensitivity to interest rate changes was found to be greater when the current rate was below the contract rate than when it was above, a result similar to that found by Arak and Goodman. His data, however, is inadequate for testing the effect of mortgage age ("seasoning") on prepayment probabilities since pools of varying ages were joined together by interest rate.

Green and Shoven analyzed the prepayment experience of individual mortgages. They emphasized that the appropriate measure of the incentive to refinance depends on the amount of principal outstanding as well as the difference between current and contract rates. Specifically, they use the percentage "lock-in" to measure the incentive to refinance. This is defined as the difference between the book and fair

---

* For a more specific discussion of this type of empirical model, see Chapter VI.

market value of the mortgage divided by the initial principal amount (adjusted for inflation).[7] Using a proportional hazard framework, they find that this measure had a statistically significant influence on the probability of prepayment and that, even after controlling for this effect, the probability of prepayment was not constant over the life of a mortgage. Because they use the Cox method of partial likelihood[8] to estimate their model, they are unable to estimate a specific parametric form for this "pure aging effect."

Quigley relates the probability of moving to a number of individual household characteristics such as income, race, education, family size and so forth, using a non-proportional hazard model approach. He finds, for example, that an increase in family size has a significantly positive influence on the probability of moving. His most interesting finding from the perspective of the current thesis is that the "lock-in", as described above, had a significant influence on the probability of moving. Specifically, he finds that households with below market rate mortgages are less likely to move than those with above market rate mortgages, all else equal. This, of course, implies that households with below market rate mortgages are less likely to prepay.

---

[7]See Chapter IV for the discussion of a similar variable.

[8]See Kiefer [1988] for a simple presentation of the Cox method for estimating the proportional hazard model.

In their study of the valuation of GNMA's, Jacob, Lord, and Tilley argue that prepayments in a given period are likely to be a function of the historical path of interest rates as well as the current interest rate. They propose a model of the prepayment of a Ginnie Mae which indicates that the incentive to prepay can be measured as the cumulative difference between the contracted and market rate over the previous four months. They do not give details of estimation and, in particular, do not state whether they work with coupon aggregated data or estimate on the basis of specific pools. However, they state five regularities they claim characterize mortgage prepayments generally: (1) prepayments tend to increase over the life of a pool, (2) there is seasonality in prepayments, (3) there are geographic differences in prepayment rates, (4) prepayments respond inversely and with a lag to a drop in interest rates, and (5) if previous prepayments were high then current prepayments tend to be low. While these results are interesting, suggestive and consistent with the industry "folklore", they are not supported with explicit, reproducible empirical analysis.

In the working paper of Giliberto and Thibodeau, the individual's decision to prepay his mortgage is modeled explicitly and is made to depend upon such household characteristics as marital status, household income, etc. The model is estimated with individual data for the period 1981-1986 using the hazard function framework. They report as their most important empirical finding that increased interest rate volatility significantly reduces the probability of prepayment.

The most recent empirical work on mortgage prepayments appears in Schwartz and Torous. They estimate their model using monthly GNMA prepayment data. Like Green and Shoven, they use a proportional hazard model to asses the influence of interest rates on prepayments. Unlike Green and Shoven, however, they assume a specific parametric form for the baseline hazard. This allows them to estimate the "pure aging effect." They find evidence that prepayments first increase with age, reach a peak after about 6 years and decline thereafter. They also find that interest rates influence prepayments in the expected way and that this influence is more pronounced as market rates fall further below contracted rates.

CHAPTER III

THE GNMA PREPAYMENT DATA

The most important data in this empirical study of mortgage prepayment behavior are the historical monthly conversion-factor time series for individual GNMA I pools. The first section of this chapter describes the characteristics of these pools and the corresponding mortgage-backed pass-through securities ("Ginnie Maes"). The second section describes in detail the conversion-factor time series data themselves and how prepayments can be calculated from them. In the last section, the method of selecting the sample and some of the attributes of that sample will be discussed.

A. The GNMA Mortgage-Backed Pass-Through Security

The Government National Mortgage Association (GNMA) was created by the Housing and Urban Development Act of 1968 as a government corporation which is a part of the Department of Housing and Urban Development (HUD). The purpose was to have GNMA assist in the financing of housing by making real estate mortgages attractive to a wider range of potential investors. GNMA has performed this function primarily by acting as a guarantor of mortgage-backed securities.

The most important type of mortgage-backed security guaranteed by GNMA is its pass-through known as the Ginnie Mae I. The issuing process of the Ginnie Mae may be described briefly as follows. A financial institution creates a pool of first mortgages insured by either the Federal Housing Administration (FHA) or the Farmers Home Administration (FmHA) or guaranteed by the Veterans Administration (VA).[1] After proper application by the institution, GNMA submits its approval of the pool, provides its guarantee and issues Ginnie Mae certificates. The certificates are then sold to securities dealers who ultimately sell them to the investing public.

GNMA guarantees the timely payment of principal and interest to the holders of its mortgage-backed securities backed by the "full-faith and credit of the United States Treasury". A letter dated December 9, 1969 from the Assistant Attorney General to the Secretary of the Department of HUD assures GNMA that it can legally make this declaration and that the GNMA guarantee constitutes a "general obligation of the United States".[2] Another letter dated February 13, 1970 from the Secretary of the Treasury assures GNMA of its ability to borrow from the Treasury.[3] As a result we can conclude that, for all intents and purposes, Ginnie Maes are default risk-free.

---

[1]All of these types of mortgages are assumable and all carry no prepayment penalty.

[2]See GNMA Mortgage-Backed Securities Dealers Association [1978], p. 85.

[3]Ibid., p. 84.

GNMA has strict requirements concerning the characteristics of the mortgages underlying the Ginnie Mae I pools it guarantees. For example, the mortgages must be level payment, fixed-rate mortgages for single family homes. The mortgages must be "non-seasoned"; specifically, individual mortgages may be no older than 12 months when the GNMA pool is created. The pool size at origination must be $1 million or more with no fewer than 12 loans. Actual pools are generally larger. Most importantly, all mortgages in a pool must have the same original term to maturity and the same contracted rate of interest. This contracted rate of interest is always 50 basis points above the coupon interest rate on the Ginnie Mae pass-through; 44 basis points serve as a loan servicing fee and 6 basis points are for the GNMA guarantee.

B. Calculating Prepayments from the GNMA Factor Series

The pool conversion factors, reported monthly in The Bond Buyer, are computed by GNMA and represent the current aggregate principal balance of the mortgages in a pool divided by the face amount of the mortgages at the time the pool was originated. For example, a pool with an original face amount of $1 million with a current conversion factor of 0.673832 has outstanding principal of $673,832. In addition to the monthly factors, for each pool they also report its coupon interest rate and thus the contracted rate on the underlying mortgages as well as the issue date, original term to maturity, and original principal balance of the pool. We do not know the exact time to maturity for each of the individual mortgages, although we of course do know that the current term of at least one mortgage is the same as the current term of the pool itself.

The above information allows for the calculation of historical prepayments for individual pools under some simplifying assumptions:

(i) all mortgages have the current term of the pool,

(ii) there are no partial prepayments of principal (e.g., "double" monthly payments),

(iii) all mortgages in the pool have equal original principle balance.

Assumption (i) is quite weak given the discussion of the characteristics of GNMA I pools above. Assumption (ii) appears to be weak as well, especially over the period of this empirical study, which is from January 1979 to November 1986 inclusive. According to Hendershott, Hu and Villani [1983], households' incentive to accumulate wealth by reducing debt through the gradual prepayment of their mortgages has declined substantially since 1979 as alternative possible investments have become far more attractive. Assumption (iii) is the strongest. In the absence of specific information to the contrary, however, this is probably the most reasonable assumption to make. According to Senft [1985], all three of these assumptions are frequently made by the active participants in the GNMA pass-through market.

The total monthly prepayments for a pool of mortgages can be defined as principal payments in excess of the scheduled amortization payments of the remaining mortgages in the pool (i.e. those which have not already prepaid).[4] Let P_{jt} represent the prepayments for pool j when it is t months old, measured in dollars. P_{jt} can be represented by the following equation:

$$P_{jt} = B_{j,t-1} - B_{jt} - A_{jt} \tag{3.1}$$

where B_{jt} represents the actual principal balance of pool j when it is t months old and A_{jt} represents the total scheduled amortization of the remaining mortgages in pool j for month t.

The third term on the left above requires further comment. Define A^*_{jt} as the scheduled amortization on pool j for month t <u>assuming zero prepayments up to that time</u>. Then

$$A_{jt} = A^*_{jt}(B_{jt}/B^*_{jt}) \tag{3.2}$$

where B^*_{jt} is equal to the scheduled principal balance of pool j at month t assuming zero prepayments up to that time. Equation (3.2) states that the scheduled amortization of a pool, A_{jt}, is proportional to A^*_{jt}, where the factor of proportionality is the ratio of actual to scheduled principal outstanding.

The significance (3.2) is that A^*_{jt} and B^*_{jt} can be calculated directly from the monthly factor time series under the assumptions discussed above. To see this, define the following general "present value of an annuity interest

---

[4]No distinction is being made here between prepayments and proceeds from foreclosed loans.

factor":

$$(3.3) \qquad I_{rn} = \frac{1 - (1/(1 + r)^{n})}{r}$$

where r is an appropriate discount rate expressed as a monthly decimal and n is the number of time periods for the annuity in months. Then the following expression is easily arrived at using standard formulae for fixed-rate mortgages:

$$(3.4) \qquad A^{*}_{jt} = \frac{(I_{cr,m-(t-1)} - I_{cr,m-t})}{I_{cr,m}} B_{jo}$$

where cr is the contracted monthly rate of interest on the underlying mortgages in pool j and where m is the original term to maturity of the pool in months. B_{jo} is the original principal balance of the pool. In addition,

$$(3.5) \qquad B^{*}_{jt} = (I_{cr,m-t}/I_{cr,m})B_{jo} \ .$$

Equations (3.2) through (3.5) can be substituted into (3.1) to arrive at the desired measure of prepayments which can be calculated from the available GNMA I factor data.[5]

A pool's prepayment rate for month t is defined as the percentage of principal outstanding at the beginning of the month that prepays by the end of the month. It can be calculated as

$$(3.6) \qquad P^{r}_{jt} = P_{jt}/B_{j,t-1} \ .$$

3.3 <u>The Sample</u>

Historical factor series for over 50,000 GNMA I pools

---

[5]Violations of assumptions (i) through (iii) would complicate the calculation of scheduled amortization and thus prepayments.

were available on tape for the period beginning January 1979 and ending November 1986. These pools were newly originated over the same period. Because of the unmanageable size of this data set, a number of steps were taken to reduce the sample size. First the data was checked thoroughly for obvious errors and missing values. Only pools with complete factor series from their time of origination to November 1986, the closing date of the study, were included in the sample. The factor series were tested to ensure that they declined absolutely through time, which they must do given the fixed-rate fully-amortizing nature of the pools under consideration. Only pools with original terms of exactly 30 years were included.

The most significant step that was taken to reduce the sample size was to group the pools into cohorts based upon coupon rate and original issue month. Cohort prepayment rates can then be calculated as

$$P^r_{ct} = P_{ct}/B_{c,t-1} \tag{3.7}$$

where P_{ct} are total prepayments at t for cohort c, measured in dollars, and where B_{ct} is the principal balance of cohort c at time t. The P_{ct} and B_{ct} are easily obtained by aggregating the P_{jt} and B_{jt} by issue month and coupon rate. Under the assumption that any measurement error in the P_{jt} is random and independent across pools, the grouping will reduce the measurement error in the calculation of prepayments.[6]

One of the characteristics of this data and most time series data on prepayments is that it is right-censored. At the close of the observation period in November 1986, some of the mortgages in the cohorts will not have prepaid because the oldest mortgages, those originated in January 1979, will have been outstanding for only 8 years. It can not be known when or if these remaining mortgages will prepay. Censoring causes some problems in the context of the proportional hazard model; methods for dealing with it will be discussed more fully in Chapter V.[7] As a way to reduce the censoring problem and to reduce the sample size at the same time, however, pools issued on or after December 1984 have been eliminated. Therefore, the shortest factor series in the sample is 2 years long.

The proportional hazard model that will be described in Chapter V requires information on the specific number of individual mortgages prepaying in every month for each cohort. If the assumed common principal of mortgages in each pool were known at each point in time, this number could be calculated directly from the P_{jt}. Because this information is not available, however, an estimate of the common principle must be used.

Corresponding to each cohort c is the average amount of

---

[6]Further implications of this grouping will be discussed in the next chapter.

[7]Censoring does not pose a problem in the context of the aggregate logit model.

a new mortgage loan originated in the U.S. during the cohort's month of issue. Call this L_{co}. This information can be used to arrive at an estimate of the common principal of the cohort's underlying mortgages at each month t. Call this L_{ct}. Specifically,

$$(3.8) \qquad L_{ct} = (I_{cr,m-t}/I_{cr,m})L_{co}$$

where cr and m are specific to cohort c and the mortgages of which it is composed. L_{ct} represents the scheduled principal outstanding on an original loan of L_{co} after t months when the contract rate is cr. The number of mortgages in cohort c that prepay in month t can then be calculated as

$$(3.9) \qquad N_{ct} = P_{ct}/L_{c,t-1}.$$

The number of mortgages remaining in cohort c at time t, R_{ct}, can be calculated as

$$(3.10) \qquad R_{ct} = B_{ct}/L_{ct}.$$

The precision of the P_{jt} and the P_{ct} is independent of the assumed common value of the L_{ct}. Their accuracy depends upon the correctness of the assumptions made in section B. In fact, if the P_{jt} are approximately correct, the N_{ct} will correctly measure the number of mortgages prepaying in a cohort "on average" since overestimates and underestimates of the number of mortgages prepaying for each pool in a cohort should cancel upon aggregation. To better improve the accuracy of the N_{ct}, cohorts consisting of less than 50 pools were eliminated from the sample.

It can be noted at this time that the proportional hazard model estimated by Schwartz and Torous [1989] also required as input the number of mortgages prepaying at time t. They also only had data on aggregate prepayments in dollars. To implement their model they divided their measure of prepayments by a common principal of $100,000 across all times of issue. It is not clear if they adjust this amount for scheduled amortization, as in equation (3.8) above. In any case, they have an obviously inconsistent measure of the N_{ct}. They point out that the values of the maximum likelihood estimates of the parameters in their model will not depend upon the value of the assumed common principal but that standard errors will be affected by this measure. They correct for this problem by using a relatively complicated resampling plan. The resampling methodology is not used in this study because the estimates of the N_{ct} should be close in the large sample.

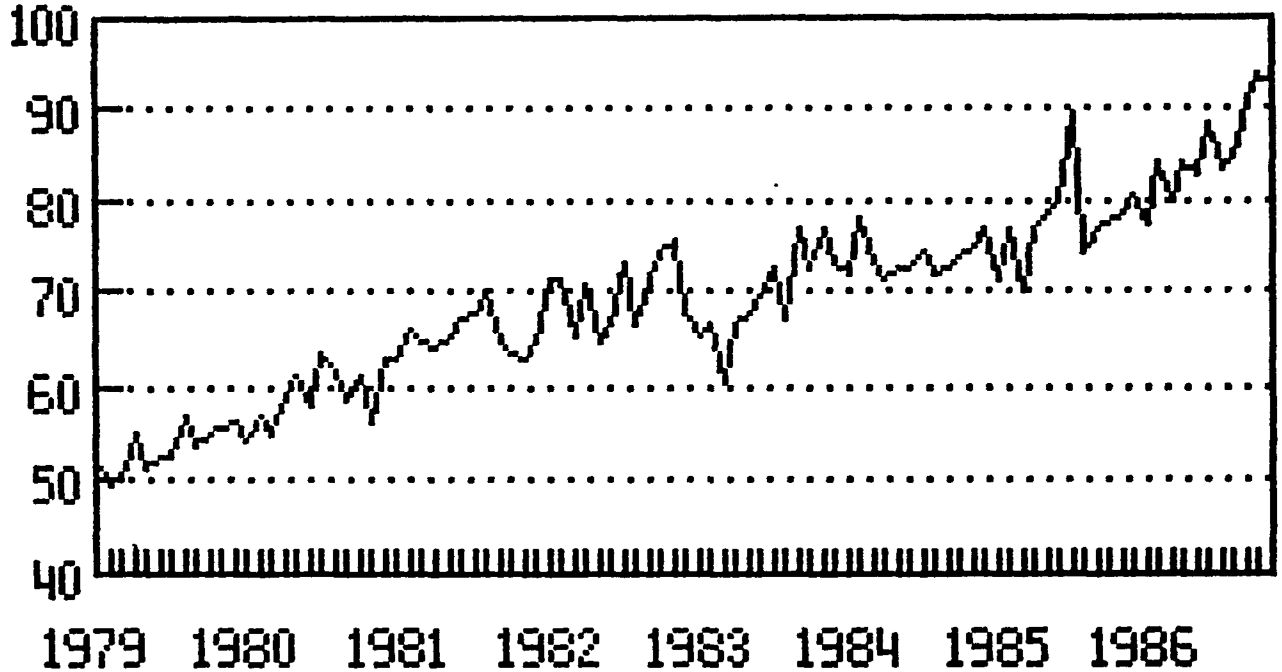
The average loan data used in this thesis is a weighted average based on a sample survey of mortgages originated by major institutional groups, compiled by the Federal Home Loan Bank Board in cooperation with the Federal Deposit Insurance Corporation. A plot of this data is given in Figure 3.1. The average mortgage loan in the U.S. rises from about $50 thousand in early 1979 to about $93 thousand in late 1986.

Some characteristics of the final GNMA sample are summarized in Table 3.1 where the listing is by coupon rate and issue month cohort. The 84 sampled cohorts are roughly

Figure 3.1

THE AVERAGE AMOUNT OF A MORTGAGE LOAN

thousands of dollars



Source: Federal Home Loan Bank Board

the same as those used by Schwartz and Torous, although their cohorts are listed by coupon-rate and issue year. For each cohort, the original number of mortgages is calculated as B_{co}/L_{co} where B_{co} is the original principal balance of cohort c. This amount is equal to the sum of all B_{jo} in c. The original number of mortgages ranges widely from a low of 1,811 for the September 1984 13's to a high of 32,579 for the April 1983 11.5's. The total number of mortgages in the sample is 672,438.

Also listed for each cohort is the percentage of the original number of mortgages that have prepaid as of November 1986. Note that all cohorts show some prepayment experience. Some cohorts, typically the low coupon cohorts, show relatively little prepayment experience. For example, only about 17 percent of the 9's and 9.5's issued in 1979 have prepaid. In contrast, approximately 75 percent of the 15's issued in 1981/1982 have prepaid. An interest rate effect is clearly operative although other factors appear likely to be significant as well. In the overall sample, 37.11 percent of the mortgages have prepaid as of the end of the observation period. This, of course, implies that 62.89 percent of the mortgages are right-censored at November 1986.

Also shown in Table 3-1 are maximum and minimum monthly prepayment rates for each cohort. Monthly prepayment rates range from 0 to 7 percent and more. Note in particular the small prepayment rates for the 9's and 9.5's of 1979.

Table 3.1
CHARACTERISTICS OF THE SAMPLE

This table lists the coupon-rate/issue-month cohorts included in the sample. I provide the estimated original number of mortgages in each cohort. I also tabulate for each cohort the number of mortgages that have prepaid (as a percentage of the original total) and the minimum and maximum monthly prepayment rates, all as of November 1986.

Cohort			Estimated Original Number of Mortgages	As of November 1986		
Issue Year	Issue Month	Coupon Rate (%)		Total Prepaid (%)	Monthly Prepayment Rate (%) Minimum	Maximum
1979	3	9.0	7,230	18.58	0.04	1.02
	4	9.0	9,630	17.72	0.03	0.84
	5	9.0	10,498	16.85	0.04	0.91
	6	9.5	7,867	17.96	0.02	0.94
	7	9.5	14,249	18.17	0.05	0.95
	8	9.5	14,809	17.69	0.01	1.09
	9	9.5	16,771	17.08	0.01	0.93
	10	9.5	17,553	16.65	0.01	0.89
	11	9.5	4,250	16.28	0.01	0.98
	11	10.0	6,788	18.00	0.01	1.00
	12	11.0	6,236	29.71	0.00	2.27
1980	1	11.0	7,784	29.21	0.00	2.25
	2	11.0	7,996	29.28	0.01	2.32
	3	11.0	6,357	28.54	0.01	2.47
	4	12.5	2,927	54.36	0.00	5.91
	5	12.5	4,055	52.38	0.02	5.58
	6	12.5	4,314	51.65	0.05	5.04
	6	13.5	1,908	64.88	0.08	6.43
	7	11.0	5,574	27.20	0.02	2.18
	7	12.5	2,359	53.79	0.08	5.87
	8	11.0	13,613	28.79	0.02	2.29
	9	11.0	14,245	27.24	0.02	2.23
	10	11.5	7,651	36.96	0.01	3.55
	11	12.5	8,812	51.02	0.01	5.08
	12	12.5	7,947	50.20	0.02	4.84
1981	1	13.0	3,153	53.73	0.01	5.06
	2	13.0	5,730	52.86	0.00	5.22
	3	13.0	4,721	55.64	0.02	6.20
	4	13.0	5,321	56.34	0.01	5.98
	5	13.5	4,046	63.03	0.00	6.20
	6	14.0	3,144	68.92	0.00	5.88
	7	15.0	3,641	78.33	0.00	5.83
	8	15.0	2,967	72.50	0.01	5.05
	9	15.0	2,929	75.55	0.00	6.08
	12	16.0	2,485	78.95	0.01	7.48

Table 3.1 (continued)
CHARACTERISTICS OF THE SAMPLE

Cohort				As of November 1986		
Issue Year	Issue Month	Coupon Rate (%)	Estimated Original Number of Mortgages	Total Prepaid (%)	Monthly Prepayment Rate (%) Minimum	Maximum
1982	2	15.0	2,328	75.04	0.04	5.12
	5	15.0	3,846	74.18	0.02	5.39
	6	15.0	4,305	74.84	0.04	6.02
	7	15.0	2,820	72.87	0.03	5.80
	8	15.0	6,091	74.50	0.06	5.62
	9	15.0	4,535	72.66	0.13	6.02
	10	13.5	2,107	64.17	0.05	8.67
	11	13.0	2,869	59.60	0.01	7.91
	12	12.0	7,604	42.87	0.04	5.76
1983	1	11.5	7,910	34.61	0.02	4.58
	1	12.0	5,971	43.34	0.05	5.66
	2	11.5	18,517	35.22	0.03	4.93
	2	12.0	3,806	43.22	0.11	6.17
	3	11.5	23,403	34.94	0.04	4.93
	4	11.5	32,579	33.74	0.03	4.79
	5	11.5	28,659	32.24	0.04	4.47
	6	11.0	3,362	19.04	0.01	2.62
	6	11.5	27,024	31.69	0.04	4.83
	7	11.0	11,873	19.62	0.01	2.61
	7	11.5	16,138	31.50	0.04	4.67
	8	11.5	19,087	30.99	0.02	4.34
	8	12.0	5,213	41.42	0.00	7.76
	9	12.0	12,226	39.03	0.02	5.94
	9	13.0	7,034	52.96	0.04	7.21
	10	12.5	11,207	46.21	0.02	6.77
	10	13.0	5,217	49.97	0.07	6.38
	11	12.5	13,314	45.51	0.02	6.57
	12	12.0	2,952	37.26	0.00	6.20
	12	12.5	8,091	42.27	0.03	6.00
1984	1	12.0	11,724	36.16	0.02	5.49
	1	12.5	2,061	38.82	0.05	5.78
	2	12.0	11,792	35.40	0.02	5.57
	3	12.0	12,758	35.06	0.01	5.48
	4	12.0	12,048	35.63	0.02	5.42
	5	12.0	6,594	34.39	0.01	5.15
	5	12.5	8,019	46.19	0.02	7.46
	6	12.0	3,070	34.95	0.01	5.77
	6	12.5	9,092	43.96	0.01	7.43
	6	13.0	2,337	50.11	0.01	8.06
	7	12.5	3,982	46.38	0.01	8.19
	7	13.0	4,358	48.90	0.01	7.50

Table 3.1 (continued)
CHARACTERISTICS OF THE SAMPLE

Cohort			Estimated	As of November 1986		
		Coupon	Original	Total	Monthly Prepayment Rate (%)	
Issue Year	Issue Month	Rate (%)	Number of Mortgages	Prepaid (%)	Minimum	Maximum
1984	7	13.5	3,401	54.28	0.01	7.91
	8	13.0	2,448	47.51	0.07	7.25
	8	13.5	7,324	53.80	0.02	7.00
	9	13.0	1,811	46.66	0.01	7.55
	9	13.5	6,331	49.83	0.03	6.58
	10	13.0	4,609	46.91	0.06	7.02
	10	13.5	2,320	41.90	0.02	5.85
	11	13.0	6,711	45.95	0.03	6.86
		Totals:	672,438	37.11		

CHAPTER IV

THE DETERMINANTS OF MORTGAGE PREPAYMENT BEHAVIOR

Both the proportional hazard model and the aggregate logit model allow the conditional probability of prepayment to depend upon any number of explanatory variables. Some examples of these types of variables have already been suggested in Chapter II. This chapter will extend that discussion to the specific variables included in this study.

The variables that influence prepayment behavior can be viewed as falling into two broad categories: (i) those designed to measure the mortgagor's financial incentive to refinance his mortgage, and (ii) those which capture the non-financial influences on prepayment behavior and which are generally related to the probability of moving.[1] Each of these categories will be discussed in turn.

It should be indicated at the outset that the choices of explanatory variables for the empirical analysis are limited by the aggregate character of the data; the variables that are ultimately chosen must be based upon characteristics of all mortgages in a given cohort at a

---

[1]This classification ignores the determinants of defaults which, in the case of the guaranteed Ginnie Mae, often result in prepayments. Fabozzi [1985] argues that defaults have historically been a relatively small proportion of total prepayments for these securities.

given point in time. This of course precludes the incorporation of variables that are unique to individual mortgagors as well as geographic and demographic variables that might otherwise have been included.

A. The Financial Incentive to Refinance

The most significant event that results in a mortgage prepayment occurs when a mortgagor refinances his existing mortgage at a lower rate of interest. As was already discussed in Chapter II, if refinancing involved no transactions costs, it would be optimal to refinance whenever the current rate on mortgages dropped below the contract rate on an existing mortgage. In fact, refinancing typically involves direct economic costs in the form of prepayment penalties and costs of originating the new mortgage (eg., points). Furthermore, refinancing can involve non-trivial amounts of time for the mortgagor.[2] Consequently, it is plausible to expect that the likelihood of prepayment will be a strictly increasing function of the incentive to refinance before transactions costs.

The incentive to refinance a mortgage is clearly related to the amount by which the contract rate on an existing mortgage exceeds the current market rate for new mortgages. Let cr be the contract rate and let r_t be the (contemporaneous) market rate at age t. An elementary

---

[2] Despite the fact that FHA and VA mortgages carry no prepayment penalties, refinancing these loans will involve the other costs mentioned.

measure of the incentive to refinance at t would then be

$$(4.1) \qquad REF1_t = Max(cr - r_{t-1}, 0).$$

The construction of the above variable recognizes that mortgagors are likely to respond to interest rate movements with a lag. This is because of inertia between the perception of a refinancing opportunity and the actual act of refinancing. The construction also imposes a non-symmetrical effect for positive and negative interest rate differentials.[3]

If one were dealing with mortgages that were grouped by interest rate only, one would be restricted to REF1 or slight transformations thereof (see, for example, Navratil [1985]). This, however, can seriously misrepresent the incentive to refinance. For example, when interest rates fall below the contract rate on an existing mortgage, the decision of whether or not to refinance will in general be affected by the amount of time remaining until the maturity of the mortgage; the prospect of reducing the contract rate one percentage point would be a relatively strong inducement to refinance if there were 25 years remaining on the mortgage but would be a minor inducement if there were only 2 years remaining. Consequently, we should expect that, all else equal, the impact of a given difference between contract and current market rates will change over the life of a mortgage.

---

[3]See section B of this chapter for a further discussion of this asymmetry.

A measure of the incentive to refinance which explicitly accounts for the above considerations is related to the difference between the fair market value of an existing mortgage, excluding the value of the prepayment option, and the principal outstanding. The fair market value can be approximated by discounting the remaining payments on an existing mortgage by the current market rate of interest. It is straightforward to show that the difference is identical to the present discounted value of the interest payments saved through refinancing, again using the current market rate of interest as the rate of discount. If we express these interest savings as a percentage of current outstanding principal, we can represent them by the following equation:

$$(4.2) \qquad PVS_t = \frac{I_{r_t,m-t} - I_{cr,m-t}}{I_{cr,m}} \cdot \frac{B_o}{B_t}$$

where all symbols are as defined in Chapter III. Two aspects of this measure are important. First, when r_t is less than cr, PVS is clearly positive and decreasing as the remaining term to maturity (m-t) decreases. This is shown in Figure 4.1 for a 30 year mortgage with cr equal to 12 percent and r_t equal to 10 percent. Second, when r_t is greater than cr, PVS is negative and increasing as the remaining term declines. (See Figure 4.2.) By this measure, two mortgages with the same cr but with different terms to maturity will have different incentives to refinance at any given time.

Figure 4.1

PVS WHEN THE CONTRACT RATE IS 12 PERCENT
AND THE MARKET RATE IS 10 PERCENT

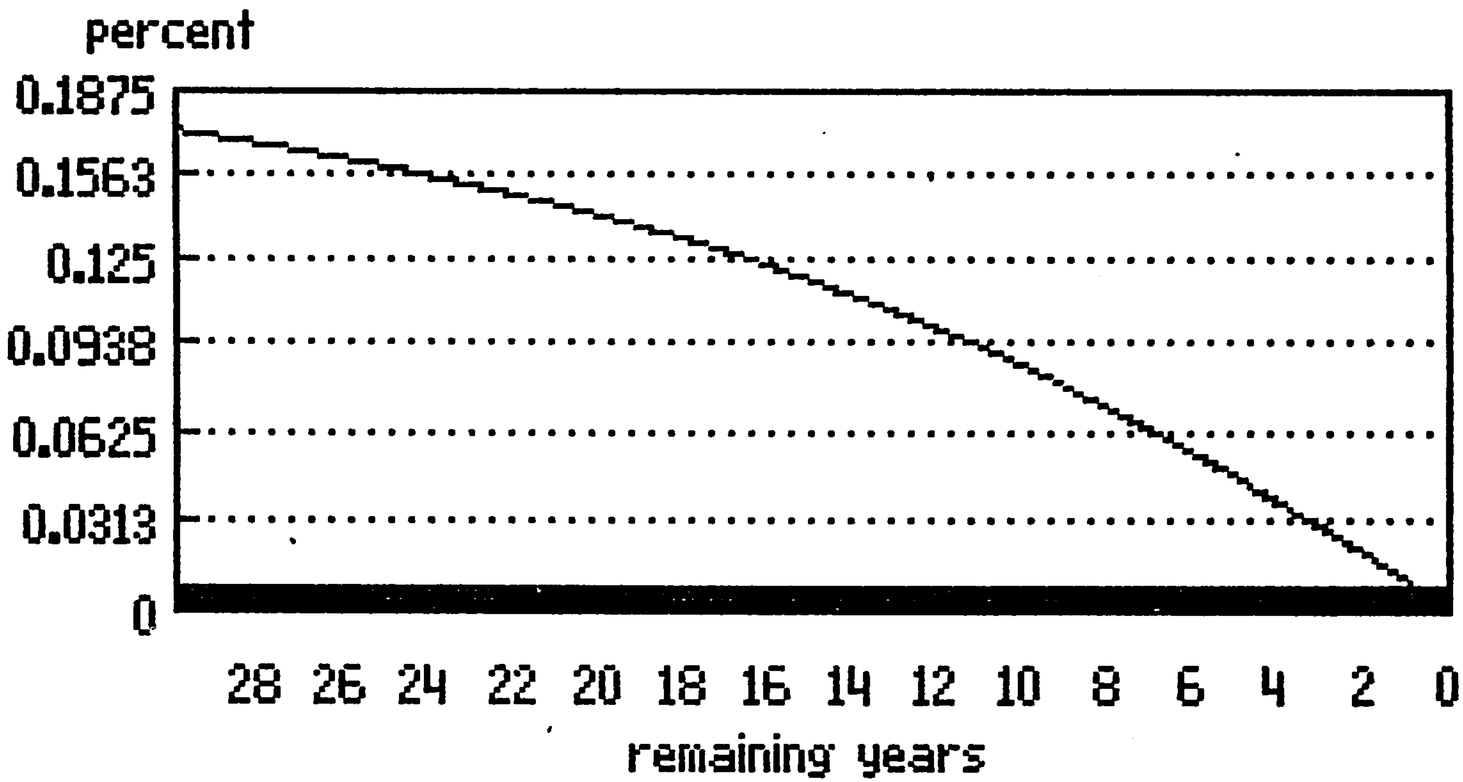
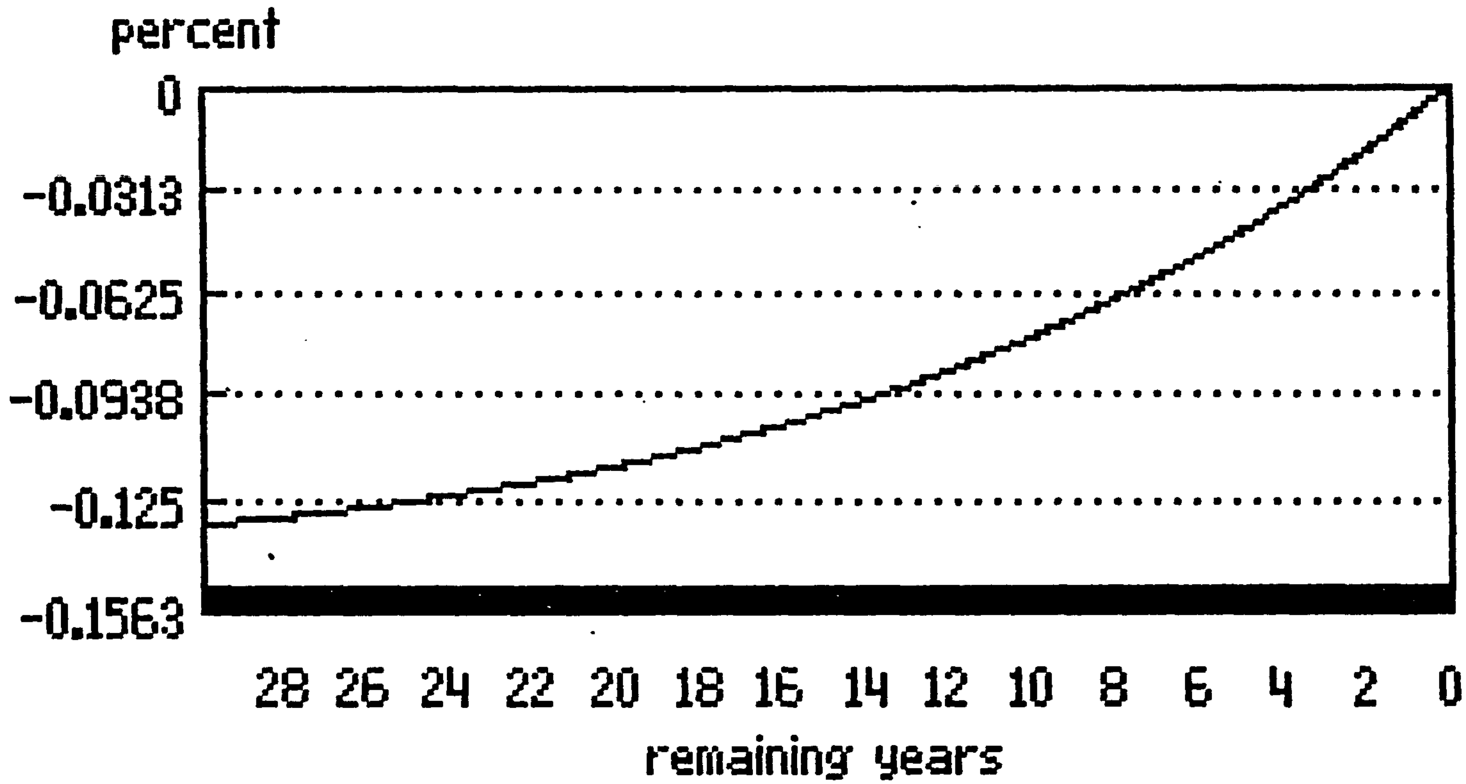


Figure 4.2

PVS WHEN THE CONTRACT RATE IS 12 PERCENT
AND THE MARKET RATE IS 14 PERCENT



The variable PVS described above is used to obtain a more refined measure of the incentive to refinance which is called REF2. REF2 can be defined by the following equation:

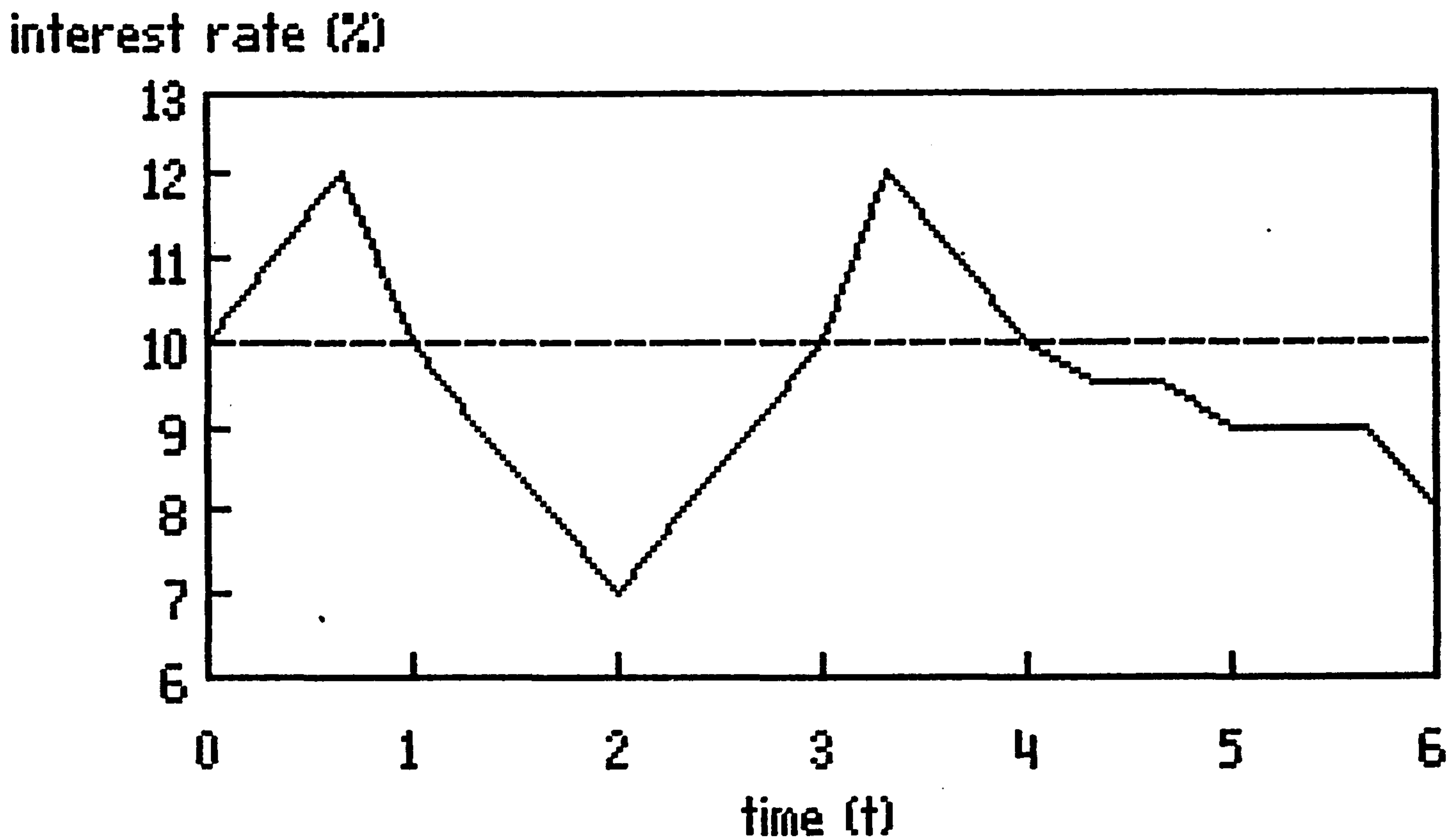
$$(4.3) \qquad REF2_t = \mathrm{Max}(\ PVS_{t-1} - FEES_{t-1},\ 0\ )$$

where FEES is a measure of the total transactions costs associated with refinancing expressed as a percentage of the current principal balance. REF2 can be interpreted as the "percentage net benefit to refinancing immediately" since it adjusts PVS for transactions costs. This variable, like REF1, allows for a lag in the mortgagor's response to changes in the incentive to refinance. It also imposes an asymmetry for positive and negative (net) values of PVS.

In the discussion above, it was noted that the financial incentive implied by a given interest rate differential changes as the time to maturity decreases. PVS, and thus REF2, account for this effect. In addition, the <u>responsiveness</u> to a given financial incentive to refinance may also change over time. This can understood clearly by reference to Figure 4.3 where there is graphed a hypothetical history of mortgage interest rates against time. Consider the situation at some time t_6, when the rate on new mortgages is 8%. Suppose one outstanding mortgage was originated at time t_0, and a second originated at t_3. Both carry a contract rate of 10%, and both imply an incentive to refinance before transactions costs. However, the situations of the mortgages are likely to be different. The reason is that the first mortgage has been in existence

Figure 4.3

HYPOTHETICAL MORTGAGE INTEREST RATES



during the period from t_1 to t_3 when interest rates fell below 10% reaching a minimum of 7% at t_2. In contrast, the second mortgage has never been exposed to a period when the interest rate lay below the contract rate by as much or for as long a time. It can be argued that the first mortgage should have responded to the strong prior incentive to refinance but because it is remaining at time t_6, it is signalling that the mortgagor faces relatively high transactions costs of refinancing. Knowing nothing else about the two mortgages, we would expect at t_6 that the probability of prepayment would be higher on the second mortgage than on the first. To secure the above described effect, a new variable is introduced. CUMREF is the cumulative prior incentive to refinance measured as the simple sum of all previous incentives. Geometrically, for the first mortgage it would be the area above the interest rate curve but below 10% from t_1 to t_3 and from t_4 to t_6. For the second mortgage it would include the second period only.[4] Formally,

$$(4.4) \qquad CUMREF_t = D \cdot \left(\sum_{j=0}^{t-1} REF_j \right)$$

where D is a dummy variable that takes on the value 1 if REF_t is positive and 0 otherwise. (REF can be either REF1 or REF2.) It is expected that the greater the value of CUMREF, given a positive value for REF, the smaller the

[4]For simplicity, this analysis was done using interest rate differentials (REF1). The same considerations apply to REF2.

probability of prepayment, all else equal.

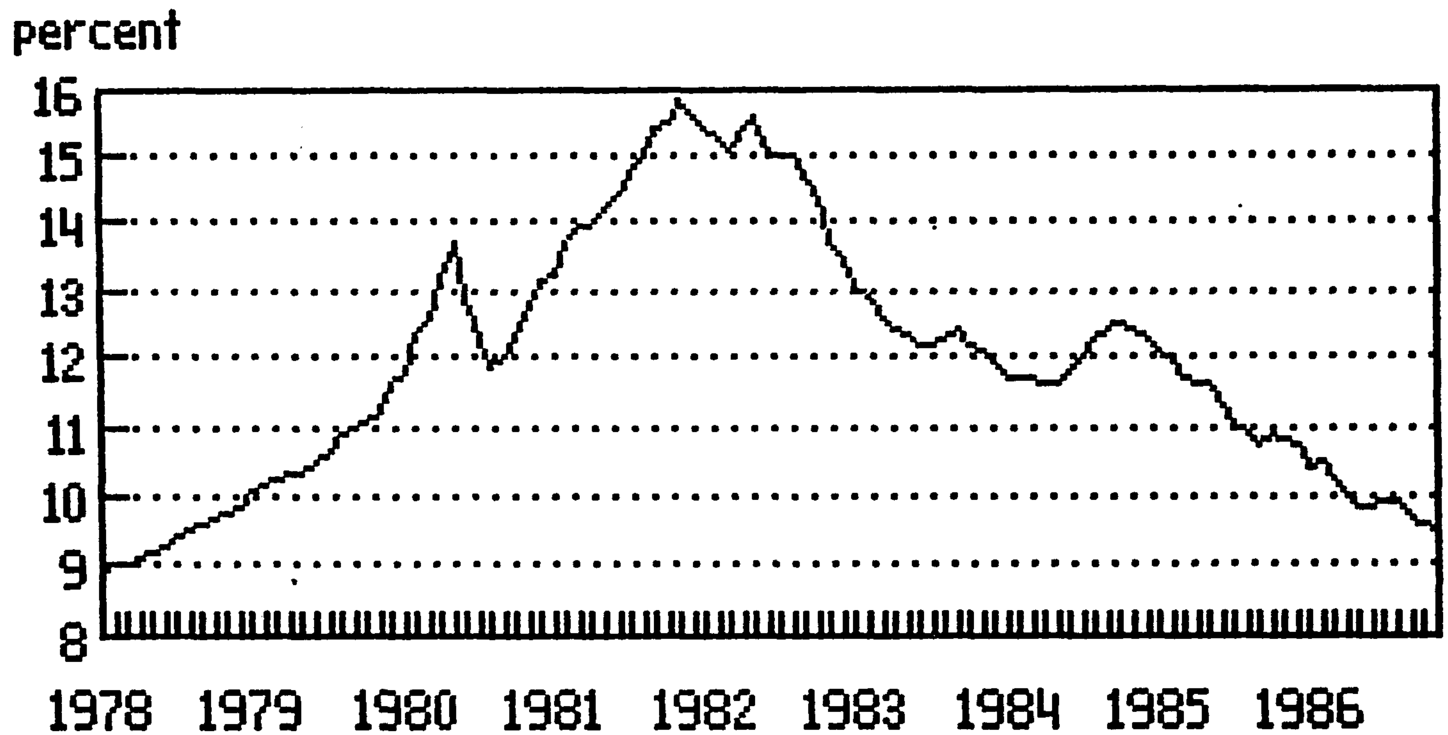
As discussed in Chapter II, the greater the volatility of interest rates, the more valuable is the mortgagor's prepayment option and, all else equal, the more likely he is to hold on to this option and wait for further declines in interest rates before refinancing. The variable VOLATILE is introduced to capture this effect. It is calculated as the standard deviation of the absolute changes in mortgage interest rates over the past twelve months. It is also multiplied by the dummy variable D described above. It is expected that the greater the value of VOLATILE, given a positive value for REF, the smaller the probability of prepayment, all else equal.

The mortgage interest rate data used to calculate the variables in this study is a weighted average based on sample surveys of mortgages originated by major institutional lender groups and compiled by the Federal Home Loan Bank Board in cooperation with the Federal Deposit Insurance Corporation. A plot of this data over the period 1978 through 1986 is shown in Figure 4.4. Note how rates reach a peak in late 1981 and early 1982. Figure 4.5 plots the corresponding measure of interest rate volatility, VOLATILE.

The source for the data on FEES is the same and includes all fees, commissions, discounts, and "points" paid in order to obtain a loan. A plot of this data is given in Figure 4.6.

Figure 4.4

MORTGAGE INTEREST RATES



Source: Federal Home Loan Bank Board

Figure 4.5

VOLATILITY OF MORTGAGE INTEREST RATES
(Standard Deviation)

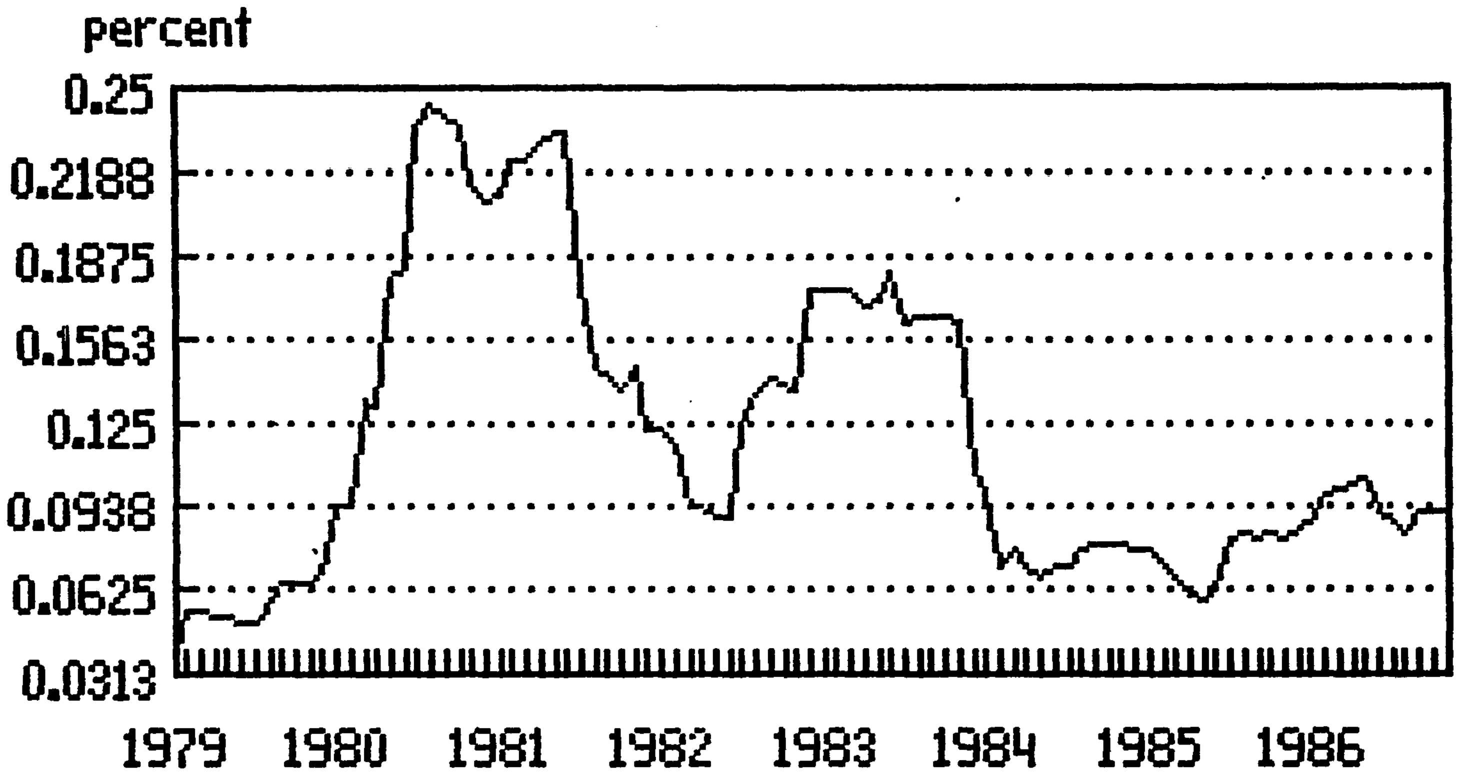
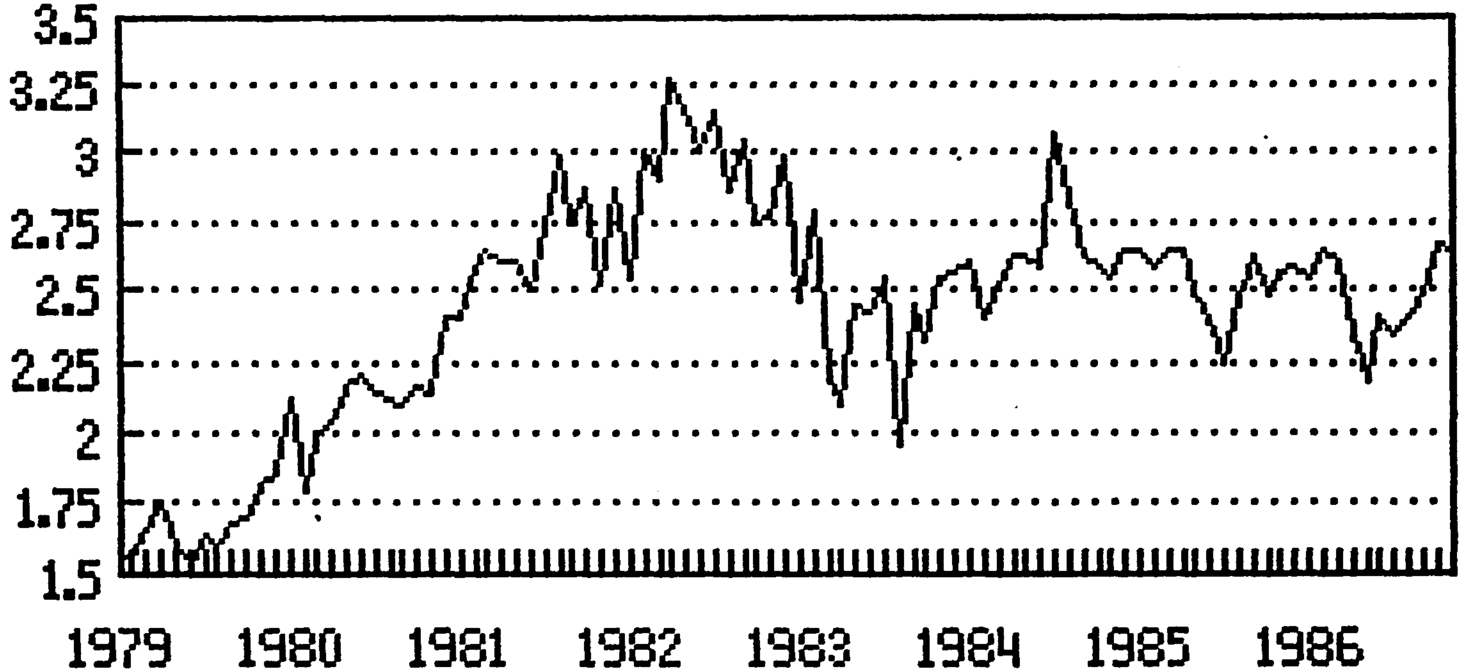


Figure 4.6

FEES AND CHARGES



Source: Federal Home Loan Bank Board

B. Other Determinants of Prepayment Behavior

The variables REF1 and REF2 measure the strength of the financial incentive to refinance a mortgage when the market interest rate lies below the contract rate. There is some evidence to show, however, that market interest rate movements influence prepayment behavior even when such rates are above the contract rate.[9] To see why such an effect might occur, suppose that the contract rate on an existing mortgage is 8 percent while the market rate on new mortgages is 13 percent. It can be argued that the mortgagor is not likely to move, and thus prepay, in such a situation because he would be giving up a valuable (to him) below-market rate mortgage. Even if he were to move, it is quite possible that the existing mortgage would be assumed by the buyer, again making prepayment unlikely. Of course the likelihood of prepayment would be even less if the market rate were even higher, say 15 percent. A variable can be introduced that captures this effect:

$$(4.5) \qquad ASSUM1_t = \mathrm{Max}(\ r_{t-1} - cr,\ 0\ ).$$

Clearly, the greater the value of ASSUM1 the less likely is prepayment. Note that ASSUM1 is analogous to REF1. There is no reason to believe, however, that the absolute effect of ASSUM1 on prepayment behavior is going to be equal to the absolute effect of REF1. In fact, it is expected that the absolute effect of REF1 will be greater than that of ASSUM1.

---

[9]See, for example, Navratil [1985], Green and Shoven [1986], and Quigley [1987].

Another variable that captures a similar effect can be defined as

$$(4.6) \qquad ASSUM2_t = Max(-PVS_{t-1}, 0).$$

ASSUM2 is proportional to the present value of interest savings associated with the assumption of a below-market rate mortgage and is analogous to REF2. This variable reflects the fact that low contract rate / long term mortgages are more likely to be assumed in a sale since a relatively large proportion of the value of the house is financed at the lower rate and because they require less cash outlay by the purchaser to effect the assumption. Clearly, the greater the value of ASSUM2, the less likely is prepayment, all else equal.

There is evidence that mortgage prepayment rates tend to be higher in the warm weather months of May, June, July, and August.* This reflects the fact that, all else equal, households are more likely to move, and thus prepay, during these months. To capture this seasonality in prepayment behavior, the following dummy variable can be created:

$$(4.7) \qquad SEASON_t = \begin{cases} 1 & \text{if t corresponds to May, June, July or August.} \\ 0 & \text{otherwise.} \end{cases}$$

The expected effect of SEASON on the probability of prepayment is positive.

Some empirical studies have included variables to

---

*See, for one example, Schwartz and Torous [1989].

capture macroeconomic influences on prepayment behavior.[7] The present study uses the single variable UNEMPLOY to control for such cyclical influences. $UNEMPLOY_t$ is measured as the difference between the civilian unemployment rate at time t-1 less 6.5 percent. There are no strong a priori beliefs concerning the expected direction of this macroeconomic effect, if any.

It is generally recognized that the prepayment behavior of recently issued mortgages tends to differ from that of long-standing mortgages. This is typically referred to as the "seasoning" of a mortgage. Seasoning is partially a financial phenomenon and has been discussed in the context of REF2, CUMREF, and ASSUM2. Seasoning may also arise because of non-financial considerations. Through the natural effect of the life-cycle, mortgagors typically change homes before the thirty-year span of a traditional fixed rate mortgage. All else equal, there may be a tendency for the prepayment probability to vary with the age of the mortgage. There is, however, no assurance that such a tendency moves continuously in the same direction. In the context of the proportional hazard model, this "pure aging effect" can be captured through appropriate specification of the baseline hazard function.[8] In the context of the aggregate logit model, the variable AGE and transformations thereof are introduced into the regression.

---

[7]See, for example, Arak and Goodman [1985].

[8]See the next chapter.

CHAPTER V

THE PROPORTIONAL HAZARD MODEL

This chapter fully develops a proportional hazard model of mortgage prepayment behavior that can be estimated empirically by the method of maximum likelihood. In the first section of the chapter, a very general prepayment function will be defined and related to a class of statistical models known as hazard functions. In addition, a general likelihood function will be constructed that is reflective of the right-censored character of the available GNMA cohort data.[1] The second section of the chapter will present the specific proportional hazard framework to be used in the empirical analysis. In this section, choices for the specification of the "baseline hazard" will be made and specific likelihood functions that reflect these choices will be constructed. The maximum likelihood estimation procedure will be described here as well.

A. Prepayment Functions and Hazard Models[2]

A mortgage prepayment function gives the probability of

---

[1]See Chapter III for a description of the cohort data.

[2]See Cox and Oakes [1984] and Kalbfleisch and Prentice [1980] for a complete and rigorous analysis of hazard models. See Kiefer [1988] for applications in economics.

an individual mortgagor prepaying his mortgage in the next instant, given that he has foregone prepaying the mortgage up to that time. This function also systematically relates the instantaneous conditional probability of prepayment to a set of exogenous variables that are called covariates.[3] To formalize these concepts, let T be a continuous random variable that represents the length of time before a mortgagor prepays. Equivalently, T is the age of the mortgage contract at the time of prepayment; t represents a specific value of T. If X_{it} represents the (1xK) row vector of covariates that influence the conditional probability of instantaneous prepayment for mortgage i at age t, then the mortgage prepayment function can be defined as[4]

$$(5.1) \quad \pi(t;X_{it},\Theta) = \lim_{s \to 0^+} \frac{\mathrm{Prob}(t \le T < t + s \mid T \ge t)}{s}$$

where Θ is a vector of parameters that is assumed constant across mortgagors and through time.

A hazard function represents the conditional probability intensity that a t-unit old item will fail. If the prepayment of a mortgage is defined as "failure", then equation (5.1) can be interpreted as a general hazard function.

Hazard functions have a number of general properties that will be useful in the construction of likelihood

---

[3]Potential covariates were discussed in Chapter IV.

[4]See Schwartz and Torous [1989] for an essentially identical representation of the mortgage prepayment function.

functions. First, for any specification of (5.1) in terms of a hazard function there is a mathematically equivalent specification in terms of an unconditional probability density function. The p.d.f. corresponding to equation (5.1) is

$$(5.2)\quad f(t;X_{it},\theta) = \lim_{s \to 0^+} \frac{\text{Prob}(t \leq T < t + s)}{s} .$$

No new parameters are identified in this transformation.

The survivor function is also related to the hazard function and is defined as

$$(5.3)\qquad S(t;X_{it},\theta) = \text{Prob}(T \geq t).$$

The survivor function is simply the complement of the more familiar cumulative distribution function $F(t;X_{it},\theta)$.[5] As a result of this,

$$(5.4)\qquad -dS(\bullet)/dt = f(\bullet) .$$

From the definition of conditional probability one can get the following:

$$(5.5)\qquad \pi(\bullet) = f(\bullet)/S(\bullet) .$$

Substituting equation (5.4) into (5.5) one obtains

$$(5.6)\qquad \pi(\bullet) = -d\ln S(\bullet)/dt$$

which shows the relationship between the hazard and survivor functions.

Another related function is the integrated hazard function $H(t;X_{it},\theta)$. It can be expressed as

$$(5.7)\qquad H(t;X_{it},\theta) = \int_0^t \pi(u;X_{iu},\theta)du .$$

The following useful relationship can be derived

---

[5]In the context of hazard models, the c.d.f. usually represents Prob(T < t) with strict inequality.

algebraically from equations (5.6) and (5.7):

$$(5.8) \qquad \ln S(\bullet) = -H(\bullet) \; .$$

In words, the natural log of the survivor function is equal to the opposite of the integrated hazard.

The mathematical relationships described above can be used to derive a general likelihood function for the hazard model that takes into account the specific characteristics of the sample data. Before proceeding, however, it is necessary to introduce some addition notations and definitions.

Let C_{ct} represent the number of mortgages in cohort c that are right-censored at age t. Let t′ represent mortgage age as of November 1986, the close of the sample observation period. Then, for this sample, C_{ct} must be defined as follows:

$$(5.9) \qquad \begin{aligned} C_{ct} &= 0 \quad \text{if } t < t' \\ &= R_{ct} \quad \text{if } t = t' \end{aligned}$$

where R_{ct} is the number of mortgages remaining in cohort c at time t, the measurement of which was described in Chapter III.

When an observation on a mortgage is right-censored at time t, the only information about its age at prepayment that is knowable is that T is greater than t. In this situation, the mortgage's contribution to the general likelihood function is S(t) where S(•) is the survivor function as defined in equation (5.3). If a mortgage is observed to prepay at time t, its contribution to the likelihood function is the density function as defined in equation (5.2). Therefore, under the assumption of the conditional independence of individual mortgages, the log-likelihood function for the sample can be expressed as follows:

$$(5.10) \qquad \ln\mathcal{L} = \sum_{c}\sum_{t} [N_{ct}\ln f(t) + C_{ct}\ln S(t)]$$

where the N_{ct} and C_{ct} essentially serve as weights.

It is often algebraically simpler to express the log-likelihood function in terms of the hazard function $\pi(\bullet)$ and integrated hazard function $H(\bullet)$. Using equations (5.5) and (5.8) and making the appropriate substitutions gives us the following form for the general log-likelihood function:

$$(5.11) \quad \ln\mathcal{L} = \sum_{c}\sum_{t} \{N_{ct}[\ln\pi(t) - H(t)] - C_{ct}H(t)\} .$$

B. <u>The Proportional Hazard Specification</u>

Hazard models can take many specific forms. One of the most popular forms, generally for reasons of its relative simplicity, is the proportional hazard model. Proportional hazard models were used to analyze prepayment behavior most notably by Green and Shoven [1986] and Schwartz and Torous [1989]. The general form of this model is the following:

$$(5.12) \qquad \pi(\bullet) = h_o(t)\exp(X_t b)$$

where h_o is the "baseline" hazard and where X_t is a row vector of covariates measured contemporaneously at t; b is a conformable column vector of unknown parameters. If the vector of covariates X is defined suitably, the baseline hazard will give the conditional probability of failure at each time t under so called "homogenous" conditions (when X = 0). In the context of the prepayment model, the baseline hazard can be given an interpretation as the "pure aging effect" or "pure seasoning effect" on prepayments.

The proportional hazard model has some other interesting properties. Note, for example, that the effect of the covariates is to act multiplicatively on the baseline hazard and that this effect is independent of time t.[6] The parameter vector b has a neat interpretation that is analogous to the interpretation of a regression coefficient:

$$(5.13) \qquad \delta \ln \pi(.)/\delta x_i = b_i \ .$$

In addition,

$$(5.14) \qquad \delta \ln \pi(\cdot)/\delta \ln x_i = b_i x_i$$

where this expression can be given an elasticity interpretation.

The parameter vector b can be estimated independently of the baseline hazard by the Cox method of partial likelihood, which can be used for any arbitrary specification of the baseline.[7] This is the methodology employed by Shoven and Green in their study. However, as pointed out by Schwartz and Torous, efficiency can be improved if we simultaneously estimate the baseline parametrically. In choosing a parametric form for their baseline, Schwartz and Torous point to evidence that the

---

[6]Though their <u>effect</u> is independent of time t, the covariates themselves may or may not be independent of time.

[7]See Kiefer [1988].

pure aging effect is non-monotonic. That is, all else equal, prepayments at first increase with age, reach a maximum and then decline.* In light of this evidence they choose a function derived from the log-logistic distribution as the baseline hazard. This function is given by

$$(5.15) \qquad h_o = (\alpha p t^{p-1})/(1+\alpha t^p) \ .$$

There are two important properties exhibited by this baseline hazard function. First, if p is less than or equal to one, then the baseline hazard declines monotonically with age. Second, if p is greater than one, then the hazard is non-monotonic and reaches a maximum at t^* where

$$(5.16) \qquad t^* = ((p-1)/\alpha)^{1/p} \ .$$

An alternative specification of the baseline hazard can be derived from the exponential distribution. In this case, the conditional probability of failure under homogenous conditions is a constant. Specifically,

$$(5.17) \qquad h_o = \alpha \ .$$

This form of the baseline implies that there is no "pure aging effect" on prepayments after controlling for relevant covariates. This specification will be empirically compared in Chapter VII to the log-logistic specification.

In order to construct specific likelihood functions for the proportional hazard model, expressions for the integrated hazard H(•) must be obtained. In cases where the covariates are time-independent, the integrated hazard can be gotten quite simply by substituting equation (5.15) or

---

*See Chapter II.

(5.17) into (5.12) and integrating over time as in equation (5.7). In cases where the covariates are time-dependent, however, the integration of the hazard becomes extremely more complex since exp(Xb) can no longer be treated as a constant when integrating. Schwartz and Torous operationalize their model by essentially assuming that the covariates are time-independent even though they are not strictly so. This allows them to approximate the integrated hazard by treating exp(Xb) as a constant when integrating. This thesis will by necessity employ the same approximation. The adequacy of the approximation will be explored in Chapter VII when the empirical results are presented.

Under the assumption of time-independence of covariates, in the log-logistic case the integrated hazard is

$$(5.18) \qquad H(t) = \ln(1+\alpha t^{p})\exp(Xb)$$

and in the exponential case it is

$$(5.19) \qquad H(t) = t\exp(Xb) \ .$$

Note that a normalization is used in the exponential case which will simplify estimation. Specifically, the covariate vector X includes as its first element the constant one. Therefore, α can be obtained as $\exp(b_1)$.

Substituting the specific forms of the hazard and integrated hazard into the general form of the log-likelihood given by equation (5.11) gives us the likelihood functions that will be estimated in this study. In the log-logistic case, the log-likelihood is

$$(5.20)\quad \ln\pounds = \sum_c \sum_t [\, N_{ct}\ln(((apt^{p-1})/(1+at^p))\exp(X_{ct}b)) -$$
$$N_{ct}\ln(1+at^p)\exp(X_{ct}b) -$$
$$C_{ct}\ln(1+at^p)\exp(X_{ct}b)\,] \,.$$

In the exponential case, the log-likelihood is

$$(5.21)\quad \ln\pounds = \sum_c \sum_t [\, N_{ct}\ln(\exp(X_{ct}b)) -$$
$$N_{ct}t\exp(X_{ct}b) -$$
$$C_{ct}t\exp(X_{ct}b)\,] \,.$$

Maximum likelihood estimates of the parameters of equations (5.20) and (5.21) were obtained by numerical methods using the programming language GAUSS. The algorithm used by GAUSS is that of Davidon, Fletcher and Powell. This is a quasi-Newton method in that the second derivatives are approximated at each iteration rather than calculated explicitly. Analytical first derivatives with respect to the parameters were supplied by the author. Because no Hessian is calculated explicitly by this method, inference and hypothesis testing must be conducted using likelihood ratio tests.

CHAPTER VI

THE AGGREGATE LOGIT MODEL

The proportional hazard model described in Chapter V is one technique for modeling mortgage prepayment probabilities. An alternative model is based upon the cumulative logistic probability function and can be called the aggregate logit model.[1]

Let π_{it} represent the probability that mortgage i prepays in month t and let X_{it} represent a (1xK) row vector of exogenous variables that influence the probability of prepayment for mortgage i at time t, including a constant. Then the logit model is

$$\pi_{it} = 1/(1 + \exp(-X_{it}b)) \tag{6.1}$$

where b is a conformable vector of parameters which is assumed constant across mortgages and through time. Defining $Z_{it} = X_{it}b$, equation (6.1) can be rewritten as

$$\pi_{it} = 1/(1 + \exp(-Z_{it})). \tag{6.2}$$

A graph of equation (6.2) is given in Figure 6.1. Note that the slope of this function is greatest near the middle of the distribution, reaching a maximum at $Z = 0$ ($\pi = 0.5$).

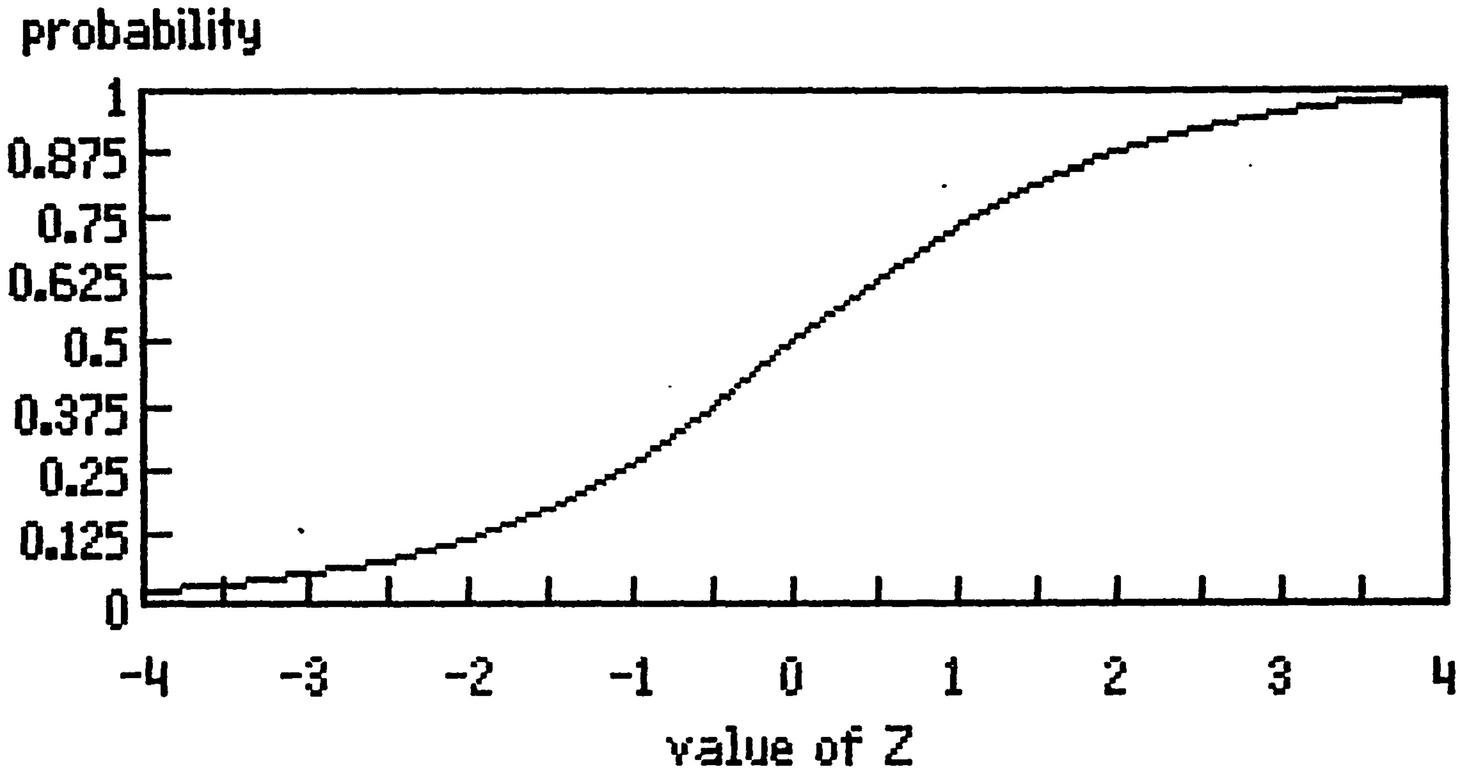
Algebraic transformation of equation (6.2) yields the following useful equation:

---

[1]The aggregate logit model was used in Navratil [1985].

Figure 6.1

THE CUMULATIVE LOGISTIC PROBABILITY FUNCTION



$$(6.3) \qquad \ln(\pi_{it}/(1-\pi_{it})) = Z_{it} = X_{it}b.$$

Equation (6.3) states that the natural log of the odds of prepayment for mortgage i at time t is a linear function of the explanatory variables X_{it}. The significance of this equation lies in the fact that, under two simplifying assumptions, unbiased and consistent estimates of the π_{it} can be obtained using the GNMA prepayment data. As a result, elementary linear techniques can be used to estimate the parameter vector b. The first assumption is that all the mortgages in a given interest rate / issue month cohort c have the same conditional probability of prepayment, π_{ct}. The second assumption is that the prepayments of mortgages in a given cohort are conditionally independent. The unbiased and consistent estimates of the π_{ct} are then given by the cohort prepayment rates, the P^r_{ct}, as defined in Chapter III.

The empirically estimable aggregate logit model is given by the following equation:

$$(6.4) \qquad \ln(P^r_{ct}/(1-P^r_{ct}) = X_{ct}b + \epsilon_{ct}$$

where ϵ_{ct} is the measurement error associated with the left-hand side variable. It is analogous to the random error term in an ordinary linear regression equation. Ordinary least squares (OLS) can be applied to equation (6.4) in order to obtain an estimate of the parameter vector b. However, the ϵ_{ct} do not satisfy all of the standard Gaussian assumptions. It can be shown that $E(\epsilon_{ct}) = 0$. It is possible to assume that the ϵ_{ct} are conditionally uncorrelated across cohorts and through time. However, the ϵ_{ct} are inherently heteroscedastic.[2] Specifically

$$(6.5) \qquad Var(\epsilon_{ct}) = 1/(R_{c,t-1}\pi_{ct}(1-\pi_{ct}))$$

where $R_{c,t-1}$ is the number of mortgages in cohort c which are outstanding at the end of month t-1.[3]

Given the nature of the heteroscedasticity as described above, weighted least squares can be used to estimate equation (6.4). However, because the π_{ct} are themselves unknown, they must be replaced by the P^r_{ct} in equation (6.5). The specific weights to be used are then the inverses of the estimated standard deviations of the ϵ_{ct}.[4] This feasible application of weighted least squares results in parameter estimates that are biased in small samples but are consistent and asymptotically normally distributed as the sample in each cohort (the $R_{c,t-1}$) gets large.[5]

---

[2]See Johnston [1984] for a thorough discussion of the properties of the ϵ_{ct} in this context.

[3]See Chapter III for a description of how the $R_{c,t-1}$ are calculated.

[4]Navratil [1985] apparently uses the wrong weights in his correction for heteroscedasticity. See p. 111.

[5]See Pindyck and Rubinfeld [1981].

CHAPTER VII

EMPIRICAL RESULTS

This chapter will present the results of the empirical analysis of mortgage prepayment probabilities. The first section will describe the results obtained from the estimation of the proportional hazard model. The second section will describe the results obtained from the estimation of the aggregate logit model. The third section will compare the two sets of results. General conclusions will be offered in the next chapter.

A. The Proportional Hazard Model

The proportional hazard model is estimated for each of the two sets of covariates listed below.[1]

COVARIATES

SET I	SET II
REF1	REF2
CUMREF1	CUMREF2
ASSUM1	ASSUM2
VOLATILE	VOLATILE
SEASON	SEASON
UNEMPLOY	UNEMPLOY
FEES	

The first set of covariates emphasizes the interest rate

[1]See Chapter IV for the descriptions of the individual covariates. REF1 and ASSUM1 have been rescaled as annual decimals.

differential REF1 as the appropriate measure of the incentive to refinance. An explicit measure of the transactions costs associated with refinancing, FEES, is also included.[2] The second set of covariates emphasizes the net benefit REF2 as the appropriate measure of the incentive to refinance. An explicit variable for the effect of refinancing transactions costs is excluded from this set since such an effect is already included in the construction of REF2. The model is estimated for each of the two sets of covariates using both the log-logistic baseline hazard and the exponential baseline hazard. As a result, four broad specifications of the proportional hazard model are considered.

All hypotheses tests concerning the maximum likelihood (ML) estimates of the parameters of the proportional hazard model are conducted using the likelihood ratio (LR) statistic.[3] This statistic can be calculated as follows: Suppose that θ is the (1xk) parameter vector of interest and that this vector can be partitioned into two sets (θ_1,θ_2) consisting of j and $k-j$ parameters respectively. Suppose the hypothesis to be tested is H_o: $\theta_1 = \theta_1^*$. Compute the ML estimate of θ without any restrictions, θ_u, and then compute the ML estimate of θ under the restrictions $\theta_1 = \theta_1^*$ and call this θ_r. Then, under the null hypothesis, the statistic $LR = 2[\ln\pounds(\theta_u) - \ln\pounds(\theta_r)]$ is asymptotically distributed

[2]$FEES = FEES_{t-1} \cdot D$. See Chapter IV.

[3]See Maddala [1977], pp. 179-180.

chi-square with j degrees of freedom, where £ is the appropriate likelihood function.

Table 7.1 presents results of models estimated with Set I covariates in conjunction with the log-logistic baseline hazard. Equation 1.A represents the unrestricted form of this class of the model. Given the ML estimates for α and p, the baseline hazard reaches a maximum at t = 154.90 months implying that, under strictly homogeneous conditions, the conditional probability of prepayment reaches a maximum at 12.91 years. The hypothesis H_o: p = 1 can clearly be rejected given the high value of the LR (chi-sqr) statistic, 48.3482. This means that we can reject the hypothesis that the baseline hazard is monotonically declining. The covariates REF1, CUMREF1, and SEASON are all highly significant in the expected direction whereas the covariates ASSUM1, VOLATILE, and FEES are all highly insignificant. UNEMPLOY has a significantly positive effect on the conditional probability of prepayment. The covariates taken as a set are highly significant as can be seen by the high chi-square value of 70.784.

The goodness of fit of the model is assessed by computing prepayment errors. These errors are calculated as actual prepayment rates less model predicted prepayment rates for each cohort and for each month.[4] A number of statistics summarizing the prepayment error distribution are listed in Table 7.1. For equation 1.A, the mean prepayment

---

[4]The total number of observations is 4526.

Table 7.1

Equation:			1.A	1.B	1.C
Log-likelihood:			-279.615	-281.870	-315.007
Baseline:		α	4.18E-6	3.74E-6	1.8E-4
		p	2.54182	2.42975	1.9888
		t^*	154.90	198.43	75.32
H_o: p = 1		chi-sqr	48.3482	45.127	
		prob>chi-sqr	1.5E-10	1.5E-10	
Covariates:					
	REF1	parameter	133.3192	79.7466	
		chi-sqr	15.808	12.1146	
		prob>chi-sqr	7.0E-6	0.00050	
	CUMREF1		-3.6282	-2.1162	
			10.81	6.565	
			0.001	0.0104	
	ASSUM1		4.8520		
			0.022		
			0.882		
	VOLATILE		-7.0158		
			0.384		
			0.535		
	SEASON		1.2824	1.4292	
			15.73	20.74	
			7.3E-6	5.2E-7	
	UNEMPLOY		0.8688	0.8981	
			8.844	22.6636	
			0.003	1.93E-7	
	FEES		-0.2451		
			0.266		
			0.606		
H_o: b = 0		chi-sqr	70.784	66.2748	
		prob>chi-sqr	1.5E-10	1.4E-13	
Prepayment Errors:					
	mean		-0.0188	-0.0162	0.00326
	mean absolute		0.02666	0.02417	0.00989
	upper quartile		0.00126	0.00090	0.00418
	median		-0.0020	-0.0021	-0.0022
	lower quartile		-0.0158	-0.0163	-0.0057
	root mse		0.06581	0.06252	0.01595
	root mse/mean prep		5.72168	5.43585	1.38660

error is -0.0188 (-1.88 percent). Also reported are the mean absolute error, the upper quartile, the median, the lower quartile, and the root mean squared error. These statistics indicate that the model generally overestimates the actual conditional probability of prepayment. Note that the mean monthly prepayment rate for the sample is 0.01161 (1.161 percent). This implies that the ratio of the root mean squared error to the mean prepayment rate is 5.72168. The model does not appear to fit the data very well.

The prepayment errors for equation 1.A are plotted against calendar time in Plot 7.1 where January 1979 = 1. The plot indicates that the prepayment errors are large generally and extremely large in late 1983 where the model in fact predicts a monthly prepayment rate of over 100 percent. Such errors suggest that the model may be incorrectly specified. This point will be discussed further in Chapter VIII.

Schwartz and Torous [1989] report three statistics that summarize the error distribution of their estimated proportional hazard model. These statistics are listed below for comparison purposes.[5]

upper quartile:	0.00127
median:	-0.00183
lower quartile:	-0.00662

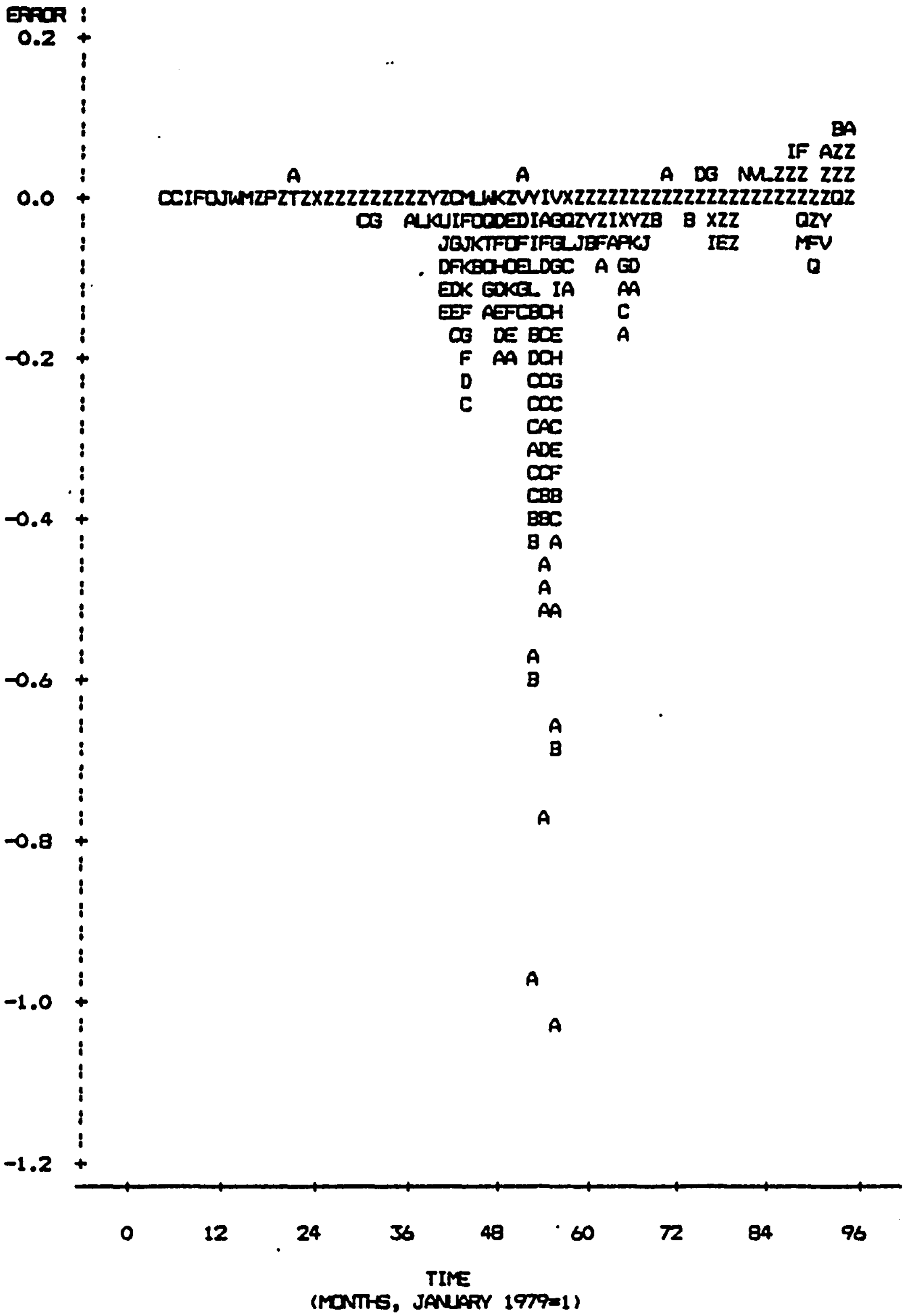
Their model generally overestimates the monthly conditional probability of prepayment as well. The upper quartile and

[5]Schwartz and Torous report annualized error statistics. The following statistics are monthly equivalents.

Plot 7.1

Equation 1.A: ERROR vs. TIME

Legend: A=1 observation, B=2 observations, ... , Z=26 or more observations.
Note: 1805 of 4526 observations are hidden.



median are nearly identical to the corresponding values for equation 1.A. However, the lower quartile of equation 1.A is more than twice as large as theirs is in absolute value. Little more can be said about the fit of their model since no additional information about the error distribution is provided.

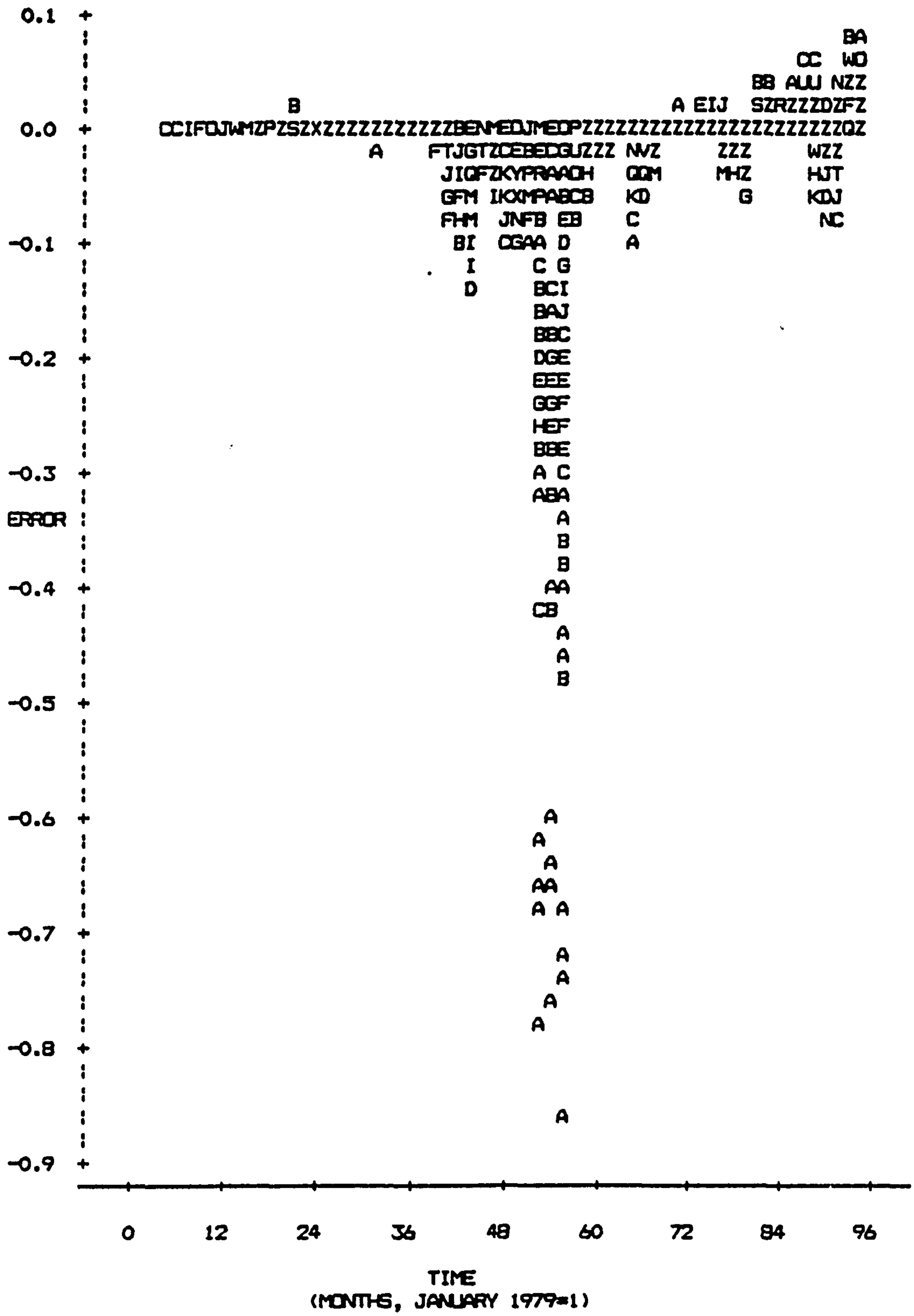
Equation 1.B in Table 7.1 is a restricted form of equation 1.A; the insignificant covariates ASSUM1, VOLATILE, and FEES have been dropped from the model. The baseline hazard reaches a maximum at t = 198.43 months. The hypothesis H_0: p = 1 is again clearly rejected. The remaining covariates are individually statistically significant and significant as a set. Equation 1.B appears to fit the data better than equation 1.A; the mean absolute prepayment error is reduced from 0.02666 to 0.02417 and the root mean squared error is reduced from 0.06581 to 0.06252. By these measures, the fit of the model has actually been improved by dropping covariates. The overall fit of the restricted model is, however, still quite poor. This can be seen by examining Plot 7.2 which again plots prepayment errors against calendar time. The general pattern is very similar to that found in Plot 7.1.

Equation 1.C in Table 7.1 provides an estimate of the log-logistic baseline hazard when all covariates are excluded from the model. Note how the mean absolute and root mean squared error are substantially reduced through such restrictions. This seems to provide additional

Plot 7.2

Equation 1.B: ERROR vs. TIME

Legend: A=1 observation, B=2 observations, ... , Z=26 or more observations.
Note: 1713 of 4526 observations are hidden.



evidence that the proportional hazard specification is incorrect. Again, this point will be explored further in the next chapter.

Table 7.2 presents results of models estimated with Set II covariates in conjunction with the log-logistic baseline hazard. Equation 2.A, the unrestricted form of this class of the model, reveals that the covariate ASSUM2 is statistically insignificant whereas the covariate VOLATILE is only marginally significant but of the correct sign. All other covariates are highly significant in the expected direction. Equation 2.B represents a restricted form of equation 2.A; ASSUM2 has been dropped from the model. Equation 2.C is identical to equation 1.C; it is repeated for comparison purposes.

The general fits of the models presented in Table 7.2 are of little improvement over the fits of the models presented in Table 7.1. This can be seen by examining Plot 7.3 which plots prepayment errors against calendar time for equation 2.B. Note the striking similarity to Plots 7.2 and 7.1. The Set II covariates appear no more powerful than the Set I covariates. This is of little surprise; since the period of the study is only 8 years long, the covariates REF1 and REF2 are going to be highly correlated in the sample. Specifically, the sample correlation coefficients for corresponding covariates from Set I and Set II are:

COVARIATES	r
REF1, REF2	0.99215
CUMREF1, CUMREF2	0.99680
ASSUM1, ASSUM2	0.99700

Table 7.2

Equation:			2.A	2.B	2.C
Log-likelihood:			-279.289	-279.321	-315.007
Baseline:		α	3.39E-6	3.28E-6	1.8E-4
		p	2.55838	2.55237	1.9888
		t^*	163.38	167.27	75.32
H_o: p = 1		chi-sqr	49.1192	49.155	
		prob>chi-sqr	1.5E-10	1.5E-10	
Covariates:					
	REF2	parameter	16.3586	16.3375	
		chi-sqr	16.815	16.7652	
		prob>chi-sqr	4.1E-6	4.2E-6	
	CUMREF2		-0.5421	-0.5422	
			11.6398	11.6172	
			0.00065	0.00065	
	ASSUM2		-1.2573		
			0.0642		
			0.79998		
	VOLATILE		-10.2702	-9.4563	
			2.7476	3.4456	
			0.0974	0.0634	
	SEASON		1.2406	1.2478	
			15.814	16.0954	
			7.0E-6	6.0E-6	
	UNEMPLOY		1.0333	1.0025	
			20.005	25.9832	
			7.7E-7	3.44E-8	
H_o: b = 0		chi-sqr	71.438	71.3736	
		prob>chi-sqr	2.1E-13	1.5E-10	
Prepayment Errors:					
	mean		-0.0187	-0.0191	0.00326
	mean absolute		0.02661	0.02703	0.00989
	upper quartile		0.00110	0.00119	0.00418
	median		-0.0017	-0.0018	-0.0022
	lower quartile		-0.0158	-0.0158	-0.0057
	root mse		0.06510	0.06582	0.01595
	root mse/mean prep		5.65967	5.72263	1.38660

Plot 7.3

Equation 2.8: ERROR vs. TIME

Legend: A=1 observation, B=2 observations, ... , Z=26 or more observations.
Note: 1605 of 4526 observations are hidden.

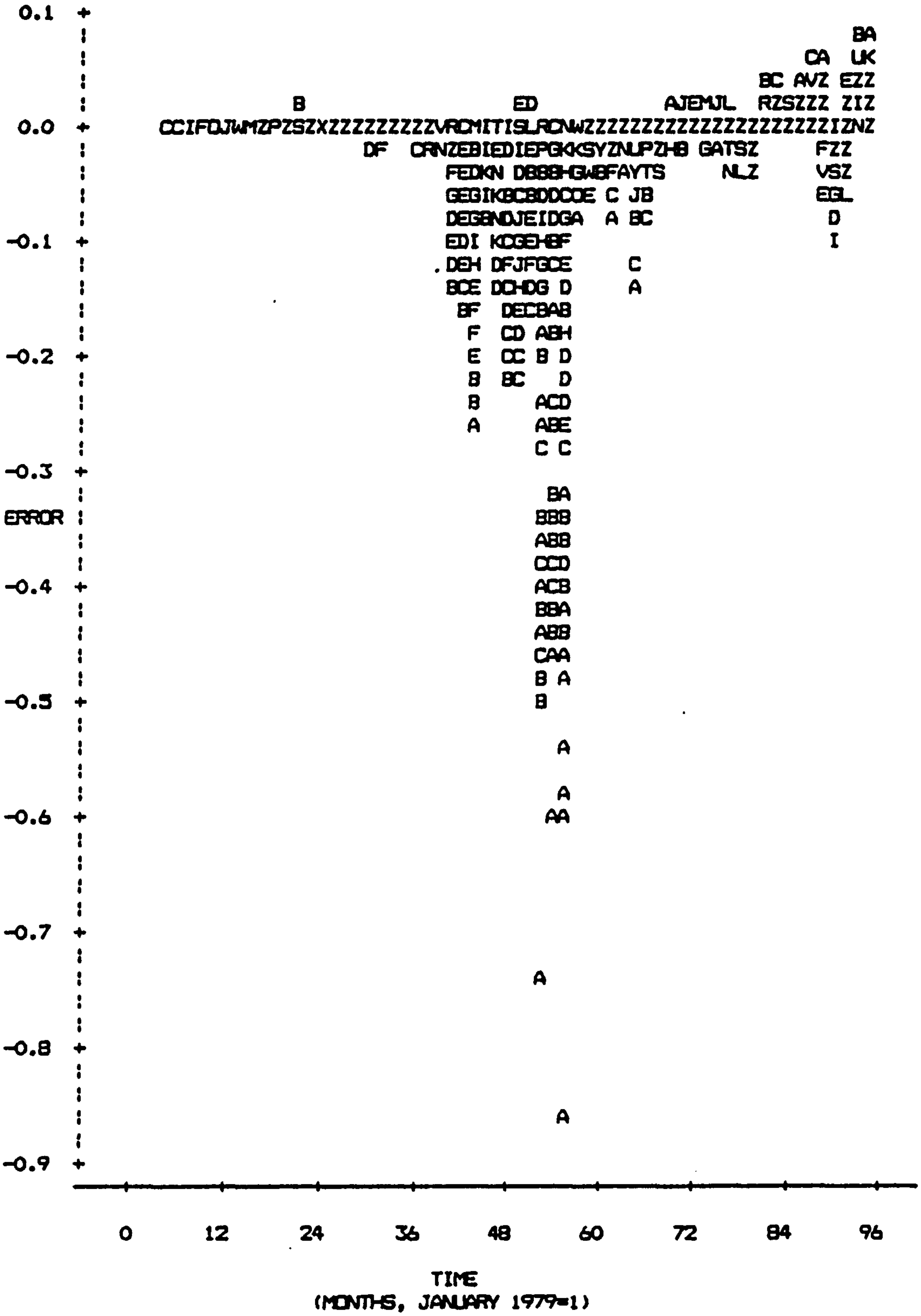


Table 7.3 presents results of models estimated with Set I covariates in conjunction with the exponential baseline hazard. Equation 3.A represents the unrestricted form of this class of the model. According to this equation, the (constant) conditional probability of prepayment under strictly homogeneous conditions is 0.347 percent per month. The covariates REF1, SEASON, and UNEMPLOY are all significant in the expected direction. The covariates CUMREF1, ASSUM1, VOLATILE, and FEES are all insignificant. Equation 3.B is a restricted form of equation 3.A where the insignificant covariates have been dropped from the model. In this equation the covariate REF1 is statistically insignificant. Equation 3.C is a restricted form of Equation 3.B where REF1 has been dropped from the model leaving only SEASON and UNEMPLOY as the significant covariates.

The fit of these models appears better than the corresponding models employing the log-logistic baseline hazard. For example, the root mean squared error is 0.02775 for equation 3.A whereas it is 0.06581 for equation 1.A. A plot of prepayment errors against calendar time is given in Plot 7.4 for equation 3.A. As can be seen, this equation still overestimates prepayments in late 1983. However, the magnitude of the prepayment errors for this time period has been reduced substantially over corresponding equation 1.A. Nevertheless, the ratio of the root mean squared error to the mean prepayment rate in the sample is still quite large

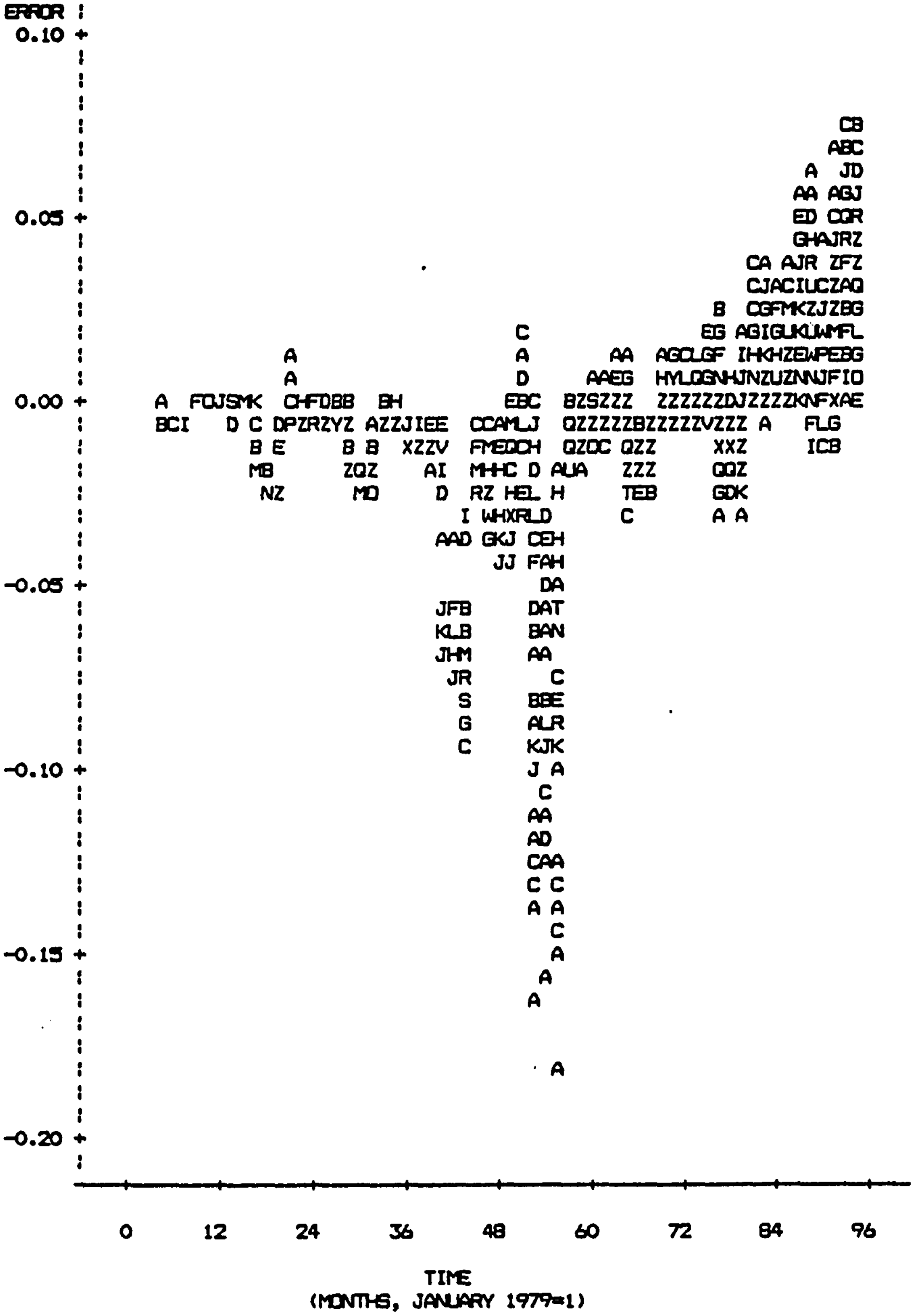
Table 7.3

Equation:		3.A	3.B	3.C
Log-likelihood:		-301.848	-303.185	-304.107
Baseline:	α	0.00347	0.00272	0.00370
Covariates:				
REF1	parameter	58.1483	13.3030	
	chi-sqr	3.8944	1.843	
	prob>chi-sqr	0.0484	0.1746	
CUMREF1		-1.5130		
		2.2508		
		0.1335		
ASSUM1		4.2691		
		0.022		
		0.881		
VOLATILE		-3.2846		
		0.1028		
		0.7485		
SEASON		1.2408	1.3524	1.3991
		15.615	19.637	21.015
		7.7E-6	9.3E-7	4.5E-7
UNEMPLOY		0.5496	0.5972	0.5445
		4.5468	12.457	11.021
		0.0330	0.0004	0.0009
FEES		-0.1730		
		0.1568		
		0.6921		
H_o: b = 0	chi-sqr	46.103	43.4272	41.5842
	prob>chi-sqr	8.5E-9	3.4E-10	9.3E-10
Prepayment Errors:				
	mean	-0.0064	-0.0067	-0.0080
	mean absolute	0.01770	0.01787	0.02007
	upper quartile	0.00283	0.00254	0.00304
	median	-0.0044	-0.0040	-0.0049
	lower quartile	-0.0144	-0.0147	-0.0191
	root mse	0.02775	0.02803	0.03048
	root mse/mean prep	2.41304	2.43668	2.64974

Plot 7.4

Equation 3.A: ERROR vs. TIME

Legend: A=1 observation, B=2 observations, ... , Z=26 or more observations.
Note: 861 of 4526 observations are hidden.



at 2.41304. This again suggests that the model does not fit the data very well.

The final set of results for the proportional hazard model is presented in Table 7.4. These equations are estimated with Set II covariates in conjunction with the exponential baseline hazard. Once again, the Set II covariates prove of little additional explanatory power when weighed against the Set I covariates. This can be seen by comparing the root mean squared error statistics in Table 7.4 to those in Table 7.3.

Equation 4.C provides an estimate of the exponential baseline hazard when all covariates are excluded from the model. This equation has the lowest mean squared error of any of the other reported equations employing the exponential baseline. However, it has a slightly higher mean squared error than equation 1.C, the estimate of the log-logistic baseline hazard when all covariates are excluded from the model. All of this provides further evidence that the proportional hazard specification is incorrect.

B. The Aggregate Logit Model

Table 7.5 presents the unweighted and weighted least squares estimates of the parameters of the aggregate logit model with explanatory variables corresponding to the Set I covariates described in Part A above. The variable LNAGE is equal to the natural log of AGE and is included in the model along with AGE in an attempt to capture a "pure-aging

Table 7.4

Equation:			4.A	4.B	4.C
Log-likelihood:			-301.744	-302.624	-324.899
Baseline:		α	0.00321	0.00253	0.00745
Covariates:					
	REF2	parameter	7.2287	4.0944	
		chi-sqr	4.295	2.534	
		prob>chi-sqr	0.038	0.1114	
	CUMREF2		-0.2328	-0.1284	
			2.6104	1.1652	
			0.1062	0.2804	
	ASSUM2		-0.3735		
			0.007		
			0.933		
	VOLATILE		-6.5838		
			1.314		
			0.252		
	SEASON		1.2304	1.3091	
			16.03	18.214	
			6.2E-6	2.0E-6	
	UNEMPLOY		0.6456	0.5925	
			10.367	12.421	
			0.0013	0.0004	
H_o: b = 0		chi-sqr	46.3108	44.5496	
		prob>chi-sqr	2.6E-8	4.9E-9	
Prepayment Errors:					
	mean		-0.0068	-0.0056	0.00405
	mean absolute		0.01821	0.01706	0.01052
	upper quartile		0.00302	0.00283	0.00627
	median		-0.0043	-0.0036	-0.0037
	lower quartile		-0.0141	-0.0131	-0.0060
	root mse		0.02872	0.02723	0.01681
	root mse/mean prep		2.49677	2.36725	1.46139

Table 7.5

Equation:	5.A	5.B
	unweighted	weighted
Parameters:		
CONSTANT	-9.34515	-9.14208
(t-value)	(-198.0)	(-134.9)
AGE	-0.01950	-0.01832
	(-21.21)	(-18.34)
LNAGE	1.257868	1.213363
	(54.231)	(41.269)
REF1	95.34320	107.9262
	(61.104)	(70.533)
CUMREF1	-1.54964	-2.03228
	(-27.72)	(-37.92)
ASSUM1	-24.6642	-22.2453
	(-30.15)	(-21.89)
VOLATILE	5.616258	6.238457
	(17.777)	(16.478)
SEASON	0.081996	0.073437
	(5.1260)	(4.7380)
UNEMPLOY	-0.00495	-0.07071
	(-0.598)	(-7.370)
FEES	-0.32605	-0.33136
	(-21.98)	(-19.41)
R-squared	0.8824	0.8774
Prepayment Errors:		
mean	0.001324	-0.00052
mean absolute	0.004634	0.004686
upper quartile	0.001060	0.000672
median	-0.00005	-0.00023
lower quartile	-0.00108	-0.00195
root mse	0.009592	0.009444
root mse / mean prep	0.826230	0.813486

effect" that would be comparable to the aging effect captured by the log-logistic baseline hazard. All of the explanatory variables are statistically significant on the basis of conventional t-tests. However, the variable VOLATILE is of the wrong sign. The reported parameter estimate implies that an increase in interest rate volatility would increase the probability of prepayment, all else equal. This is clearly contrary to prior expectations based on standard option pricing theory.* Table 7.6 presents results when this variable has been excluded from the model.

The fits of the models are again assessed by computing prepayment errors. Statistics that summarize the distributions of these errors are reported at the bottom of the tables. The fits appear good and are much better than the fits of the proportional hazard models reported in Part A above. Prepayment errors are plotted against calendar time for equation 5.B in Plot 7.5. Although this equation continues to overestimate prepayment rates in late 1983, the errors are much smaller in magnitude than those corresponding to the proportional hazard models. In addition, the means and medians of the error distributions are now quite close to zero implying that the model no longer generally overestimates the probability of prepayment.

---

*See Chapter IV.

Table 7.6

Equation:	6.A	6.B
	unweighted	weighted
Parameters:		
CONSTANT	-9.29977	-9.09540
(t-value)	(-190.7)	(-130.4)
AGE	-0.01691	-0.01487
	(-18.02)	(-14.78)
LNAGE	1.190256	1.127308
	(50.281)	(37.835)
REF1	102.7902	117.9301
	(66.101)	(81.535)
CUMREF1	-1.78219	-2.34006
	(-31.69)	(-45.24)
ASSUM1	-25.7217	-23.8837
	(-30.47)	(-22.94)
VOLATILE		
SEASON	0.071260	0.065878
	(4.3090)	(4.1290)
UNEMPLOY	0.064289	0.020797
	(8.5110)	(2.5810)
FEES	-0.14351	-0.13199
	(-12.96)	(-10.64)
R-squared	0.8741	0.8700
Prepayment Errors:		
mean	0.001463	-0.00052
mean absolute	0.004683	0.004809
upper quartile	0.001043	0.000615
median	-0.00007	-0.00024
lower quartile	-0.00100	-0.00190
root mse	0.009658	0.009717
root mse / mean prep	0.831989	0.837066

Plot 7.5

Equation 5.8: ERROR vs. TIME

Legend: A=1 observation, B=2 observations, ... , Z=26 or more observations.
Note: 1187 of 4526 observations are hidden.

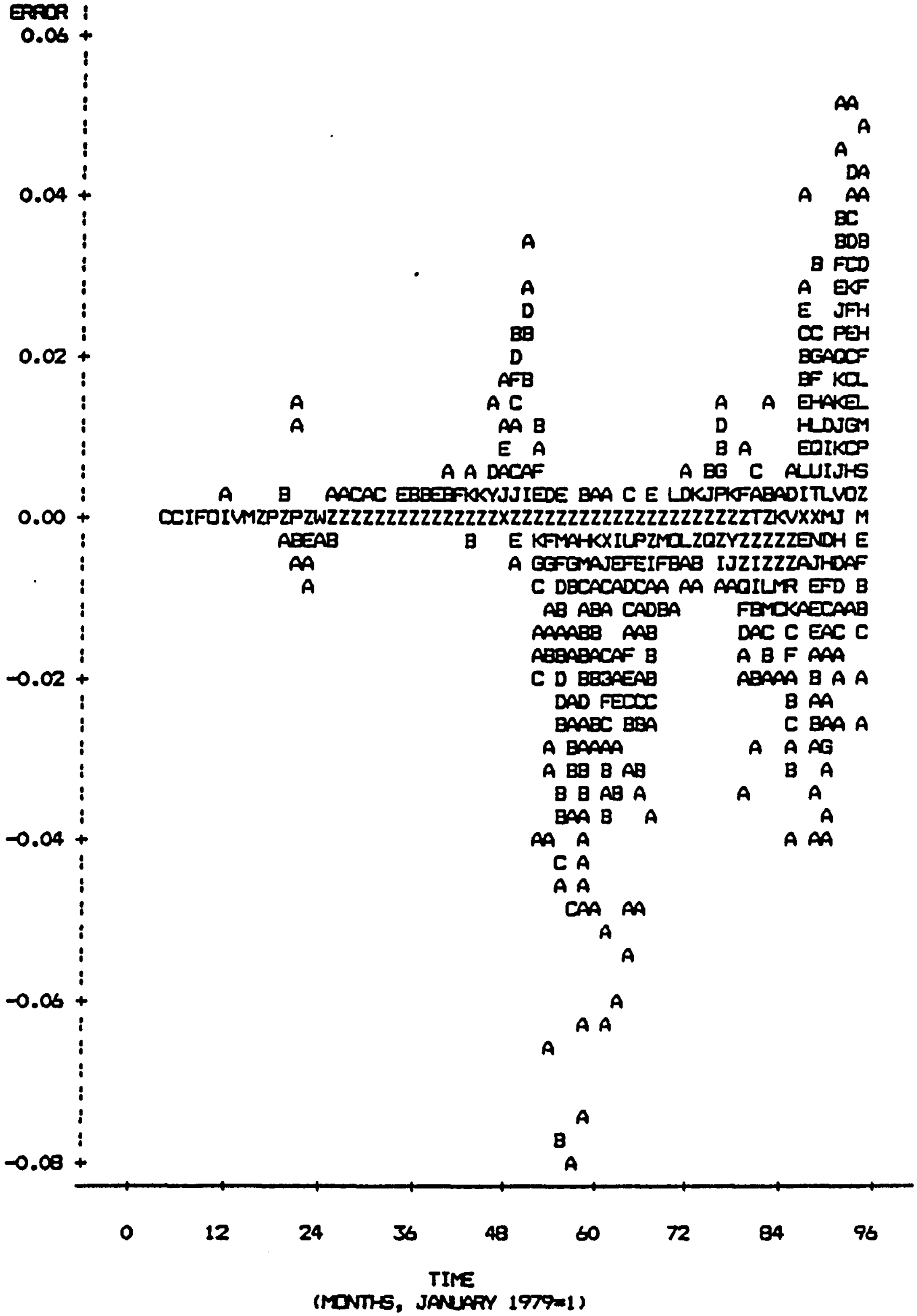


Table 7.7 presents results of models estimated with explanatory variables that correspond to the Set II covariates described in Part A. These variables improve the fit of the model slightly. This can be seen by comparing the root mean squared error of equation 7.B, 0.008300, to the root mean squared error of equation 5.B, 0.009444. All of the explanatory variables in Table 7.7 are statistically significant although once again the variable VOLATILE is estimated with the incorrect sign. Table 7.8 reports the results when this variable is restricted from the model.

C. Comparisons and Interpretations

It is already apparent that all estimated forms of the proportional hazard model fit the GNMA prepayment data quite poorly. In all cases, the ratio of the root mean squared error to the sample mean prepayment rate is greater than unity. Using this ratio as a goodness of fit criterion, the models employing the exponential baseline fit the data better than those employing the log-logistic baseline. However, their fit is still quite poor, many of the covariates are statistically insignificant and, in addition, past empirical evidence has consistently pointed to a non-constant "pure-aging effect".[7] The current results offer little solid basis for rejecting this past evidence. On the other hand, the aggregate logit model appears adequate in that the ratio of the root mean squared error to the sample

[7]See, for example, Schwartz and Torous [1989] and Arak and Goodman [1985].

Table 7.7

Equation:	7.A	7.B
	unweighted	weighted
Parameters:		
CONSTANT	-9.28442	-8.97471
(t-value)	(-205.2)	(-142.6)
AGE	-0.01459	-0.01129
	(-16.53)	(-12.33)
LNAGE	1.124720	1.014017
	(50.383)	(37.323)
REF2	12.46778	14.16050
	(60.664)	(76.023)
CUMREF2	-0.26556	-0.33672
	(-31.69)	(-45.24)
ASSUM2	-3.58385	-3.42085
	(-30.65)	(-24.54)
VOLATILE	1.411601	1.616464
	(6.2800)	(6.6850)
SEASON	0.062280	0.043211
	(3.9590)	(2.9450)
UNEMPLOY	0.077429	0.058724
	(10.346)	(7.3610)
R-squared	0.8864	0.8908
Prepayment Errors:		
mean	0.001280	-0.00039
mean absolute	0.004146	0.004183
upper quartile	0.001089	0.000691
median	-0.00003	-0.00022
lower quartile	-0.00094	-0.00162
root mse	0.008510	0.008300
root mse / mean prep	0.733065	0.714952

Table 7.8

Equation:	8.A	8.B
	unweighted	weighted
Parameters:		
CONSTANT	-9.24887	-8.93805
(t-value)	(-205.2)	(-141.9)
AGE	-0.01461	-0.01127
	(-16.47)	(-12.26)
LNAGE	1.125785	1.016847
	(50.217)	(37.250)
REF2	13.17923	14.90043
	(76.530)	(98.982)
CUMREF2	-0.28684	-0.35913
	(-37.26)	(-53.78)
ASSUM2	-3.90490	-3.86268
	(-36.97)	(-31.32)
VOLATILE		
SEASON	0.056849	0.033292
	(3.6040)	(2.2700)
UNEMPLOY	0.091245	0.079420
	(12.701)	(10.750)
R-squared	0.8854	0.8897
Prepayment Errors:		
mean	0.001240	-0.00041
mean absolute	0.004145	0.004222
upper quartile	0.001101	0.000709
median	-0.00004	-0.00021
lower quartile	-0.00088	-0.00149
root mse	0.008527	0.008491
root mse / mean prep	0.734548	0.731452

mean prepayment rate is at least less than unity in each case.

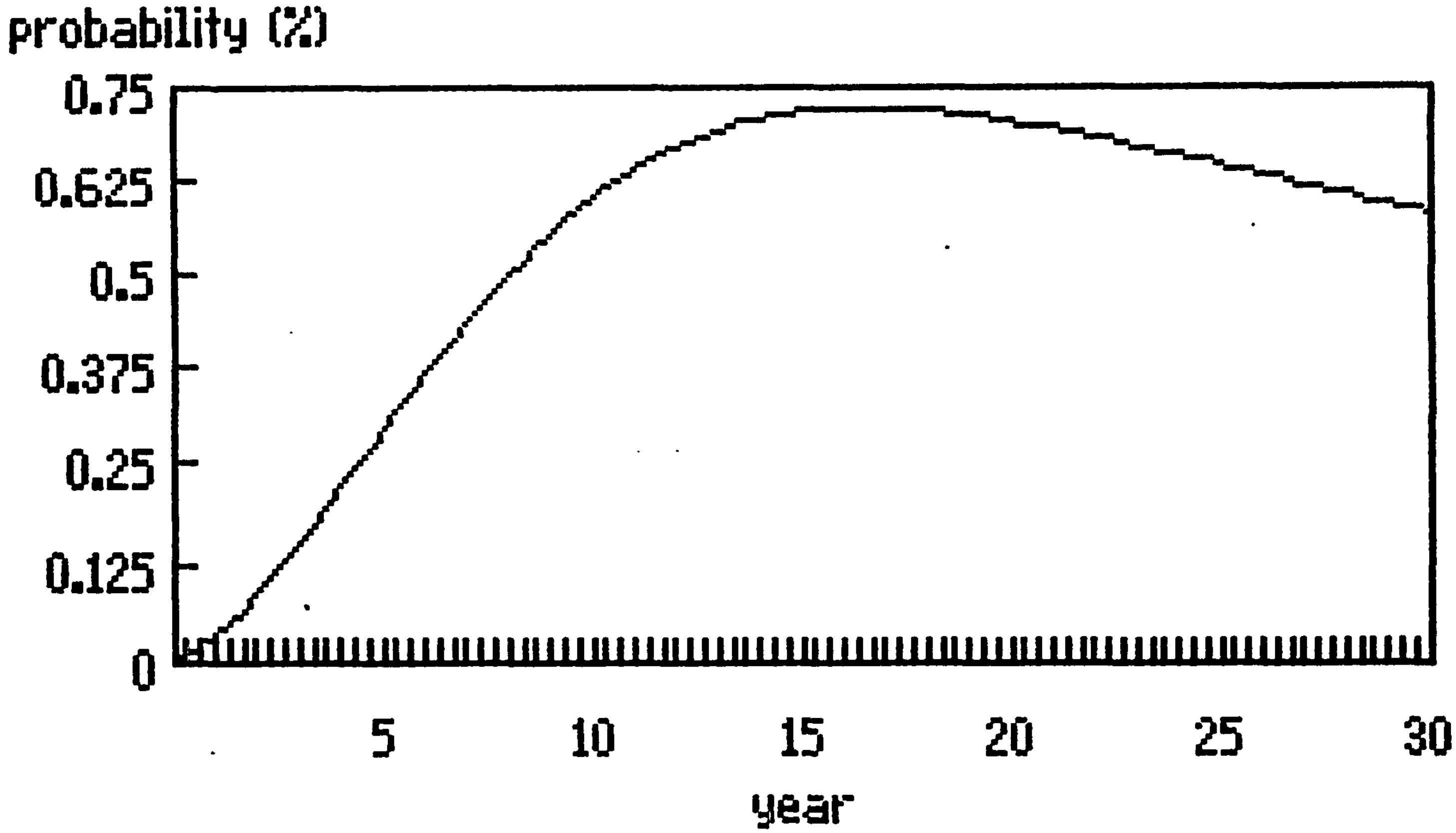
Further insights into the interpretation of the estimated models will be gained if we compute and compare model predicted mortgage prepayment probabilities for a few values of the exogenous factors. This analysis will be conducted using 3 of the estimated equations: 1.B, 6.B, and 8.B. Equation 1.B is the proportional hazard model with log-logistic baseline hazard in conjunction with restricted Set I covariates. It corresponds roughly with the model estimated by Schwartz and Torous [1989]. Equations 6.B and 8.B are weighted least squares estimates of the aggregate logit model with explanatory variables based on Set I and Set II covariates respectively. Both equations, however, exclude VOLATILE from the model.

The baseline hazard of equation 1.B gives an estimate of the conditional probability of a mortgage prepaying at each time t under strictly homogeneous conditions: market interest rates are presumed to remain fixed at the contract rate and unemployment is presumed to remain constant at 6.5 percent.* As such, the baseline hazard can be given an interpretation as the "pure-aging effect" on the probability of prepayment. The estimated baseline hazard for equation 1.B is graphed in Figure 7.1. The hazard increases rather sharply at first, reaches a maximum at 16.5 years where the conditional probability of prepayment is approximately 0.72

*This analysis ignores the effects of seasonality.

Figure 7.1

EQUATION 1.B: PURE-AGING EFFECT



percent, and declines slowly thereafter. As a point of comparison, the estimated baseline of Schwartz and Torous reaches a maximum at approximately 6 years.

The analogous aging effect for equation 6.B under identical homogeneous conditions is graphed in Figure 7.2. The curve showing the probability of prepayment at each time t rises extremely rapidly at first until it reaches a maximum at 6.3 years where the probability of prepayment is approximately 0.48 percent. This probability declines at a moderate pace thereafter. The aging effect for equation 8.B is quite similar; the maximum occurs at 7.5 years where the probability of prepayment is approximately 0.46 percent. These results seem more plausible and consistent with past evidence than do the corresponding results for equation 1.B.

Table 7.9 reports some calculated mortgage prepayment probabilities for equation 1.B. Here it is assumed that the unemployment rate is constant at 6.5 percent and that the market rate of interest is constant at the contracted rate until just before month X at which time the market rate falls below the contracted rate. The table shows what happens to the conditional probability of prepayment (measured as a percent) for various values of the decline in the market rate at various values for the month X. A one percentage point decline in the market rate raises the conditional probability of prepayment by a multiplicative

Figure 7.2

EQUATION 6.B: PURE-AGING EFFECT

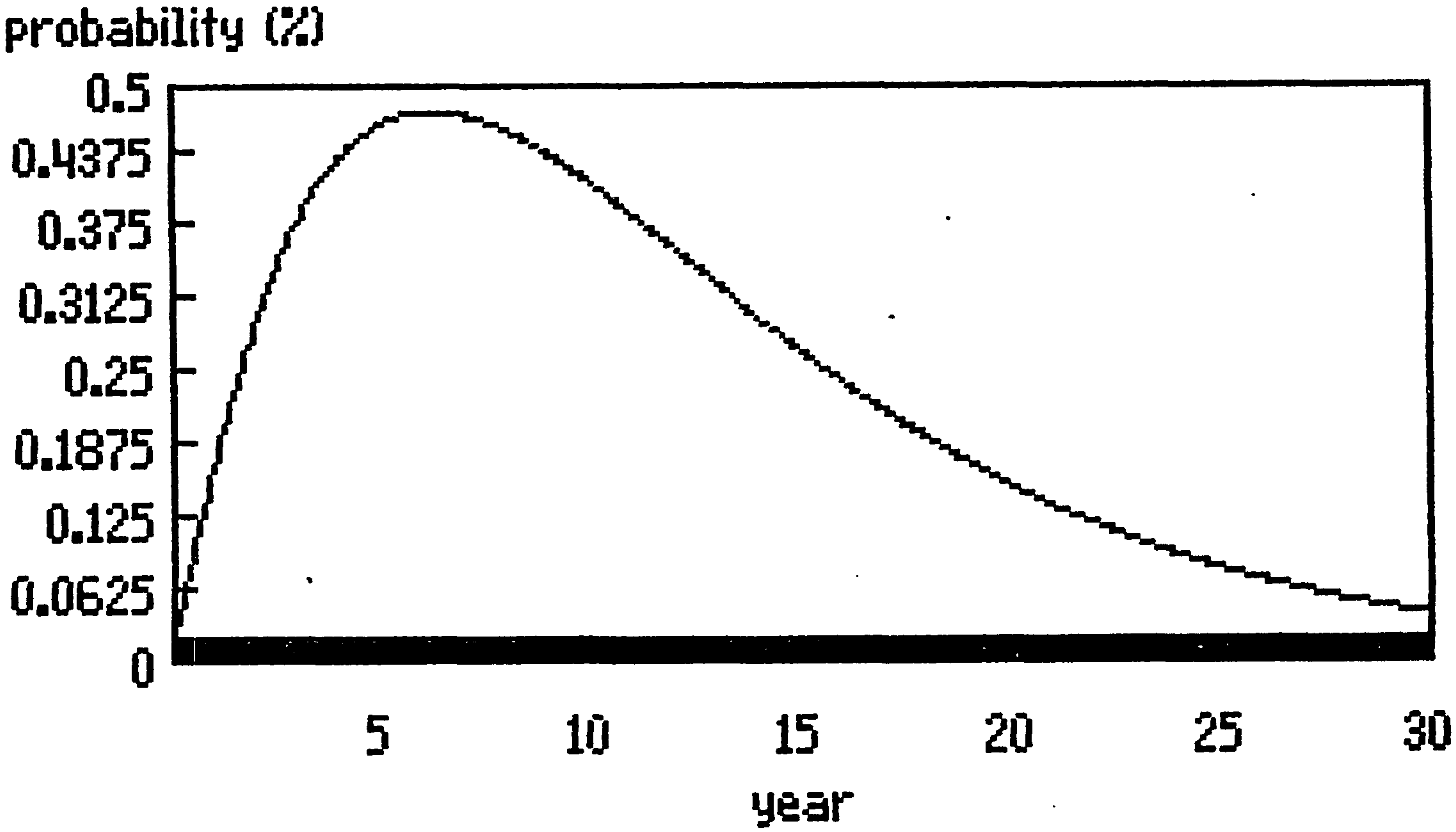


Table 7.9

Calculated Prepayment Probabilities (%): Equation 1.B

	Decline in Market Rate				
Month	0 %	1%	2%	3%	5%
12	0.03	0.07	0.16	0.35	1.71
60	0.29	0.65	1.45	3.21	15.84
120	0.60	1.33	2.96	6.57	32.36
180	0.72	1.59	3.53	7.83	38.59
240	0.70	1.56	3.46	7.69	37.89
300	0.64	1.43	3.18	7.05	34.76

factor of 2.22 independently of the month X.[9] The corresponding factors of proportionality for a 2, 3, and 5 percentage point decline are 4.93, 10.94, and 53.91 respectively. Thus, for example, a 5 percentage point decline just before month 120 would increase the probability of prepayment from 0.60 to 32.36 percent in that month. Of course, if the market rate were to remain below the contracted rate, the factor of proportionality would eventually decline over time as the negative coefficient on CUMREF1 would begin to play a role. In fact, the proportionality factor could eventually become less than unity.

Table 7.10 reports calculated mortgage prepayment probabilities for equation 6.B under a set of conditions similar to those used in drawing up Table 7.9.[10] A part of the difference between the reported probabilities in the two tables can be explained by the difference in the aging effects for equations 1.B and 6.B discussed above. However, equation 6.B appears to have an even stronger interest rate effect than equation 1.B. For example, a 5 percentage point decline in the market interest rate just before month 120 increases the probability of prepayment from 0.41 to 53.70

---

[9]This is a direct implication of the proportional hazard specification.

[10]Since equation 6.B includes the explanatory variable FEES, the assumption that FEES were equal to 2 percent was made.

Table 7.10

Calculated Prepayment Probabilities (%): Equation 6.B

	Decline in Market Rate				
Month	0 %	1%	2%	3%	5%
12	0.15	0.38	1.24	3.92	30.14
60	0.46	1.15	3.63	10.92	56.45
120	0.41	1.03	3.26	9.88	53.70
180	0.27	0.67	2.14	6.63	42.87
240	0.15	0.38	1.22	3.87	29.84
300	0.08	0.20	0.65	2.07	18.31

percent, a factor of 131.[11] The corresponding factor for equation 1.B was only 53.91. Similar relatively large interest rate effects hold for all of the other entries in the table as well.

Table 7.11 reports prepayment probabilities corresponding to equation 8.B. Values of the explanatory variable REF2 were computed for a mortgage with contract rate equal to 12 percent in order to perform the necessary calculations.[12] As was discussed above, equation 8.B fits the GNMA prepayment data only slightly better than equation 6.B and the pure-aging effects for the two equations are quite similar. However, the implied probabilities of the models are noticeably different as can be seen by comparing Table 7.10 to Table 7.11. Though the probabilities are close enough for 1 and 2 percentage point declines in the market rate, equation 8.B predicts substantially higher probabilities in the earlier months for a given 3 or more percentage point decline and substantially smaller probabilities in the later months as the time to maturity approaches. This is of course consistent with the discussion of REF2 in Chapter IV where it was pointed out that a given interest rate differential would have a greater impact on the financial incentive to refinance the longer is the time to maturity, all else equal.

---

[11]This factor is not independent of the month X as in the proportional hazard case.

[12]FEES were again assumed to be equal to 2 percent.

Table 7.11

Calculated Prepayment Probabilities (%): Equation 8.B

	Decline in Market Rate				
Month	0 %	1%	2%	3%	5%
12	0.14	0.35	1.32	5.96	75.79
60	0.43	0.96	3.29	12.47	82.56
120	0.44	0.88	2.61	8.43	63.20
180	0.34	0.58	1.42	3.72	27.11
240	0.23	0.32	0.61	1.22	5.43
300	0.15	0.15	0.22	0.32	0.68

CHAPTER VIII

CONCLUSIONS

The objective of this thesis was to estimate, interpret, and compare mortgage prepayment probabilities from two distinct types of empirical models: the proportional hazard model and the aggregate logit. Having carried out this objective, one of the most important conclusions that can be made is that the estimated proportional hazard models are completely inadequate as predictors of GNMA prepayment rates. On a number of criteria, these models simply miss the mark by a considerable amount. In fact, the examination of corresponding prepayment errors has suggested that the estimated models are incorrect specifications of the true stochastic process governing prepayments. The misspecification is possibly due to the "approximation" discussed in Chapter V and employed by Green and Shoven [1986] and Schwartz and Torous [1989]: can the time-varying character of the covariates be conveniently ignored as was done? Any further empirical research employing this model would have to explicitly resolve this dilemma. Another possible source of misspecification would be the appropriateness or inappropriateness of the assumption of proportionality. The specific choice for the parametric form of the baseline hazard is another potential source of error. It is doubtful, however, that any serious misspecification is due to the choices of the covariates themselves since these same covariates seem to "work" quite well in the context of the aggregate logit model.

The estimated aggregate logit models appear to predict GNMA prepayment rates quite well. Prepayment errors are relatively small and the R-squared statistics for the linear aggregate logit equations are greater than 85 percent in every case. Explanatory variables are all statistically significant in the expected direction with the exception of VOLATILE. The following conclusions can be made based on these empirical findings.

(i) The pure-aging effect is non-monotonic. That is, the probability of prepayment will at first increase with mortgage age, reach a maximum at approximately 7 years, and decline thereafter, ceteris paribus.

(ii) The probability of prepayment is directly related to the financial incentive to refinance. The best measure of such an incentive is given by the net present value of interest payments saved through refinancing (REF2) since this measure gives more plausible estimated prepayment probabilities than the simple interest rate differential (REF1).

(iii) The responsiveness to a given financial incentive to refinance is not constant over time but in fact declines.

This is seen by the significantly negative coefficient on the explanatory variable CUMREF.

(iv) Increases in the transactions costs associated with refinancing a mortgage will reduce the probability of prepayment, ceteris paribus. This is seen by the significantly negative coefficient on the explanatory variable FEES in Tables 7.5 and 7.6.

(v) Prepayments are seasonal. That is, prepayment rates are higher in the summer months of May, June, July, and August, ceteris paribus.

(vi) An increase in the market interest rate above the contracted rate significantly reduces the probability of prepayment as mortgagors postpone moving or have their outstanding mortgages assumed upon sale of the residence. This is indicated by the significantly negative coefficients on the explanatory variables ASSUM1 and ASSUM2. This effect, however, is not symmetrical to the effect when the market rate falls below the contracted rate. Since the absolute values of the coefficients on ASSUM1 and ASSUM2 are always less than the corresponding coefficients on REF1 and REF2, we can conclude that the interest rate effect is (not surprisingly) stronger when the market rate is below the contracted rate than when it is above by an equivalent absolute amount.

(vii) Increases in the unemployment rate generally increase the probability of prepayment, ceteris paribus. (An opposite effect occurs in Table 7.5 only.) This suggests that prepayment rates are counter-cyclical.

(viii) An increase in interest rate volatility does not have the effect on prepayments that is predicted by standard option pricing theory. According to the results presented in Tables 7.5 and 7.7, mortgagors prepay more quickly in a volatile interest rate environment rather than postponing refinancing as would be expected.

It appears that the estimated prepayment equations based on the aggregate logit model confirm most of our prior expectations about the determinants of prepayment behavior while adding new insights. The incorporation of such equations into a full model for the pricing of mortgage-backed pass-through securities should significantly enhance our understanding of value.

REFERENCES

Arak, Marcelle and Goodman, Laurie S., "Prepayment Risk in Ginnie Mae Pools," Secondary Mortgage Markets, Spring 1985, II, 20-25.

Black, Fischer and Scholes, Myron, "The Pricing of Options and Corporate Liabilities," Journal of Political Economy, May 1973, LXXXI, 637-54.

Brennan, Michael J. and Schwartz, Eduardo S., "Savings Bonds, Retractable Bonds and Callable Bonds," Journal of Financial Economics, August 1977, V, 67-88.

Cox, David R. and Oakes, David, Analysis of Survival Data. London and New York: Chapman & Hall, 1984.

Curley, Anthony J. and Guttentag, Jack M., "The Yield on Insured Residential Mortgages," Explorations in Economic Research, Summer 1974, I, 114-61.

Dunn, Kenneth B. and McConnell, John J., (1981a) "A Comparison of Alternative Models for Pricing GNMA Mortgage-Backed Securities," Journal of Finance, May 1981, XXXVI, 471-84.

Dunn, Kenneth B. and McConnell, John J., (1981b) "Valuation of GNMA Mortgage-Backed Securities," Journal of Finance, June 1981, XXXVI, 599-616.

Dunn, Kenneth B. and Singleton, Kenneth J., "An Empirical Analysis of the Pricing of Mortgage-Backed Securities," Journal of Finance, May 1983, XXXVIII, 613-23.

Fabozzi, Frank J., ed., The Handbook of Mortgage-Backed Securities. Chicago: Probus Publishing, 1985.

Follain, James R., Scott, Louis O. and Yang, Tai-Le, "Microfoundations of a Mortgage Prepayment Function." Paper presented at the annual meeting of the American Real Estate and Urban Economics Association, New York, December 1988.

Follain, James R. and Tzang, Dah-Nein, "Interest Rate Differential and Refinancing a Home Mortgage," The Appraisal Journal, April 1988, LVI, 243-251.

Giliberto, S. Michael and Thibodeau, Thomas G., "Modeling Residential Mortgage Prepayment." Paper presented at the annual meeting of the American Real Estate and Urban Economics Association, New York, December 1988.

GNMA Mortgage-Backed Securities Dealers Association, The Ginnie Mae Manual. Homewood, Illinois: Dow Jones-Irwin, 1978.

Green, Jerry and Shoven, John B., "The Effects of Interest Rates on Mortgage Prepayments," Journal of Money, Credit, and Banking, February 1986, XVIII, 41-59.

Hall, Arden R., "Valuing the Mortgage Borrower's Prepayment Option," Journal of the American Real Estate and Urban Economics Association, 1985, XIII, 229-47.

Hendershott, Patric H., Hu, Sheng and Villani, Kevin E., "The Economics of Mortgage Terminations: Implications for Mortgage Lenders and Mortgage Terms," Housing Finance Review, April 1983, II, 127-42.

Jacob, David P., Lord, Graham and Tilley, James A., "Price, Duration and Convexity of Mortgage-Backed Securities." In Mortgage-Backed Securities: New Strategies, Applications and Research, edited by Frank J. Fabozzi. Chicago: Probus Publishing, 1987.

Johnston, J., Econometric Methods, Third Edition. New York: McGraw-Hill, 1984.

Kalbfleisch, John D. and Prentice, Robert L., The Statistical Analysis of Failure Time Data. New York: John Wiley & Sons, 1980.

Kidder Peabody & Co., Mortgage-Backed Securities Report. September 19, 1985.

Kiefer, Nicholas M., "Economic Duration Data and Hazard Functions," Journal of Economic Literature, June 1988, XXVI, 646-79

Maddala, G. S., Econometrics. New York: McGraw-Hill, 1977.

Navratil, Frank J., "The Estimation of Mortgage Prepayment Rates," Journal of Financial Research, Summer 1985, VIII, 107-17.

Peters, Helen F., Pinkus, Scott M. and Askin, David J., "Figuring the Odds: A Model of Prepayments," Secondary Mortgage Markets, May 1984, I, 18-23.

Pindyck, Robert S. and Rubinfeld, Daniel L., Econometric Models and Economic Forecasts, Second Edition. New York: McGraw-Hill, 1981.

Quigley, John M., "Interest Rate Variations, Mortgage Prepayments and Household Mobility," Review of Economics and Statistics, November 1987, LXIX, 636-643.

Schwartz, Eduardo S. and Torous, Walter N., "Prepayment and the Valuation of Mortgage-Backed Securities," Journal of Finance, June 1989, XLIV, 375-92.

Senft, Dexter E., "Determining the Yield." In The Handbook of Mortgage Banking, edited by James M. Kinney and Richard T. Garrigan. Homewood, Illinois: Dow Jones-Irwin, 1985.

Siegel, Jeremy J., "The Mortgage Refinancing Decision," Housing Finance Review, January 1984, III, 91-97.