Pricing Mortgage-Backed Securities in a Multifactor Interest Rate Environment: A Multivariate Density Estimation Approach

Jacob Boudoukh[a], Matthew Richardson[a], Richard Stanton[b]

and Robert F. Whitelaw[a*]

December 18, 1996

*[a]Stern School of Business, NYU; [b]Haas School of Business, U.C. Berkeley. We would like to thank the Q Group, the Geewax-Terker Research Fund, the Berkeley Program in Finance, the U.C. Berkeley Committee on Research, and the U.C. Berkeley Center for Real Estate and Urban Economics for financial support. We are also grateful to Neal Pearson, Ravi Jagannathan, Bill Silber, Marti Subrahmanyam, an anonymous referee, and seminar participants at the Berkeley Program in Finance, the Chicago Quantitative Alliance, the NYU Salomon Center conference "Derivatives: The State of the Art", JP Morgan, Salomon Brothers, University of Chicago, U.C. Berkeley, Duke University, University of Texas at Austin, New York University, University of Michigan, University of Wisconsin, and the Utah Winter Finance Conference for helpful comments. We are grateful to Dong-Hyun Ahn, Amar Gande and Jeff Heisler for valuable research assistance.

Pricing Mortgage-Backed Securities in a Multifactor Interest Rate Environment: A Multivariate Density Estimation Approach

Abstract

This paper uses multivariate density estimation (MDE) procedures to investigate the pricing of mortgage-backed securities (MBS) in a multifactor interest rate environment. The MDE estimation suggests that weekly MBS prices from January 1987 to May 1994 can be well described as a function of the level and slope of the term structure. We analyze how this function varies across MBSs with different coupons and investigate the sensitivity of prices to the two factors. An important finding is that the interest rate level proxies for the moneyness of the option, the expected level of prepayments, and the average life of the cash flows, while the term structure slope controls for the average rate at which these cash flows should be discounted. Though the origination and prepayment behavior of mortgages differ substantially across coupons, there remains an unexplained common factor which explains 80–90% of the remaining variation of MBS prices. This factor does not seem to be related to the usual suspects, and therefore presents a puzzle to financial economists.

1 Introduction

The mortgage-backed security (MBS) market plays a special role in the U.S. economy. Originators of mortgages (S&Ls, savings and commercial banks) can spread risk across the economy by packaging these mortgages into investment pools through a variety of agencies, such as the Government National Mortgage Association (GNMA), Federal Home Loan Mortgage Corporation (FHLMC), and Federal National Mortgage Association (FNMA). Purchasers of MBSs are given the opportunity to invest in default-free, interest-rate contingent claims which offer different payoff structures from U.S. Treasury bonds. These elements combine to lower an individual's cost of obtaining a mortgage, creating welfare gains. Due to these gains, the MBS market is one of the fastest growing, as well as one of the largest financial markets in the United States. For example, in 1993, the face value of these securities outstanding was 1.5 trillion dollars, in comparison to approximately 100 million outstanding in 1980.

With increased holdings of MBSs, there have been well-documented cases of huge monetary losses incurred by financial institutions and investment groups. The risk management (or lack thereof) of S&Ls' mortgage portfolios is one example of financial institutions' vulnerability to price variations in the mortgage market. This vulnerability is partly due to the complexity of MBS pricing. On one level, pricing appears to be fairly simple. Fixed-rate mortgages offer fixed nominal payments; thus, fixed-rate MBS prices will be governed by pure discount bond prices. However, the mortgage holder has the option to prepay the existing mortgage and refinance the property with a new mortgage; hence, MBS investors are implicitly writing a call option on a corresponding fixed-rate bond. Furthermore, prepayment of mortgages (and thus the timing and magnitude of the MBS's cash flows) can also take place for reasons not related to the interest rate option. These complications lead to a nonlinear relation between MBS prices and interest rate and coupon-specific prepayment variables.

In this paper, we employ multivariate density estimation (MDE) procedures to estimate the functional relation between MBS prices and their fundamentals. There are several reasons why MDE is well suited to analyzing MBSs. First, although financial economists have good intuition for what the MBS pricing fundamentals are, the exact form is too complex (or assumption-specific) to be determined precisely from a parametric model. For example, while it is standard to assume at least two factors govern interest rate movements, the interaction between these factors and MBS prices is not well understood.[1] In contrast, MDE has the potential to capture previously unrecognized relations between MBSs and interest rate factors. Second, nonparametric procedures like MDE involve a tradeoff between consistent estimation of a general functional form and estimation error (given a finite sample). Here, due to the complex nonlinear payoffs of MBSs, the ability to estimate a flexible functional form is crucial. Moreover, derivative securities, like MBSs, are spanned by relatively few state variables and have little idiosyncratic error. Thus, sample size plays less of a role than would normally be true. Third, there have been few empirical investigations of the determinants of MBS prices — most investigations have focused on either hedging or prepayment issues.[2] Our paper is a departure from this previous work. It provides a nonparametric data analysis of the economic determinants of MBS prices (especially their relationship with interest rates), together with a detailed investigation of the possible sources of the pricing errors that remain after controlling for interest rate factors.

In order to study the finite sample properties of the MDE method, we first examine a simulated model of MBS prices. In this model, the economy is governed by a two-factor Cox, Ingersoll and Ross (CIR) (1985a,b) model. Mortgage prepayments are introduced using a modified version of the Schwartz and Torous (1989) model, which captures some of the salient features of prepayment behavior. The MDE approach is then applied to the simulated economy. For this particular model, the MDE approach approximates well the functional form of MBS prices for the majority of the sample's interest rate range. For extreme interest rates, however, the MDE method breaks down, and the pricing errors increase rapidly.

The MDE method is then applied to GNMA securities of various coupons over the period 1987-1994. The data are prices of weekly TBA (to be announced) GNMAs with coupons ranging from 7.5% to 10.5%.[3] The MDE methodology captures the well-known negative convexity of MBS prices. However, of particular importance, the relation between prices and the level of interest rates is also shown to be dependent on the slope of the term structure. We show that at least two factors are necessary to fully describe the effects of

[1] Current empirical evidence favors a multifactor approach to fixed-income pricing (e.g., Stambaugh (1988), Litterman and Scheinkman (1991) and Pearson and Sun (1994)), pointing to at least two factors. For example, nominal prices of fixed income securities may be governed by both real and nominal factors, indicating that models of interest rates (and prepayment) should contain at least two factors. Litterman and Scheinkman (1991), for example, approximate these factors by the level, slope and curvature of the yield curve. They find that the level and slope capture most of the movements of the term structure

[2] For empirical studies of MBS hedging see Breeden (1991), Harvey (1991), and Boudoukh, Richardson, Stanton, and Whitelaw (1995). Some of the better known papers on prepayment estimation include Richard and Roll (1989), Schwartz and Torous (1989), and Stanton (1995).

[3] A TBA contract is just a forward contract, trading over the counter. More details are provided in Section 3.

the prepayment option on prices. The analysis also reveals cross-sectional differences among GNMAs with different coupons, especially with regard to their sensitivities to movements in the two interest rate factors. For a given GNMA coupon, the interest rate level proxies for the moneyness of the option, the expected level of prepayments, and the average life of the cash flows, while the term structure slope controls for the average rate at which these cash flows should be discounted.

While the level and slope of the term structure explain most of the variation of MBS prices, there still remains significant common variation across coupons. This is surprising given that the most likely non-interest rate factors, i.e., origination and prepayment behavior of mortgages, differ substantially across coupons. Indeed, a thorough examination of additional interest rate and prepayment factors (e.g., *seasoning* and *burnout*) fails to provide an explanation. After taking the interest rate factors into account, we show that there remains an unexplained common factor which explains 80–90% of the remaining variation of MBS prices. The source of this variation, which does not seem to be related to the usual suspects, presents a puzzle to financial economists.

2 Mortgage-Backed Security Pricing: Preliminaries

Mortgage-backed securities represent claims on the cash flows from mortgages which have been pooled together and packaged as a financial asset. Investors in an MBS receive all payments (principal plus interest) made by mortgage holders in a particular pool (less some servicing fee). For many of these securities, the payments are guaranteed by government or private agencies so there is no question of default. In the case of a household default, the agency pays the remaining principal of that mortgage in the pool. Thus, the pricing of an MBS can be reduced to valuing the mortgage pool's cash flows at the appropriate discount rate. MBS pricing then is very much an issue of estimating the magnitude and timing of the pool's cash flows.[4]

However, pricing an MBS is not a straightforward discounted cash flow valuation. This is because the timing and nature of a pool's cash flows depends on the prepayment behavior of the holders of the individual mortgages within the pool. For example, independent of interest rates, mortgages might be prepaid by individuals with enhanced wealth or who relocate. These events will lead to early payments of principal to the MBS holders. In

[4]For a description of MBSs and their relevant characteristics, see *The Handbook of Mortgage-Backed Securities*, Fabozzi, editor, 1993.

addition, MBSs contain an embedded interest rate option. Mortgage holders have an option to prepay their existing mortgage and refinance their property at the lower interest rate. They are more likely to do so as interest rates, and hence refinancing rates, decline further below the rate of their current mortgage. This refinancing incentive tends to lower the value of the mortgage to the MBS investor because the mortgages' relatively high expected coupon payments are replaced by an immediate payoff of the principal. The equivalent investment alternative now available to the MBS investor is, of course, at the lower coupon rate. Therefore, the price of an MBS with an $X\%$ coupon is roughly equivalent to owning a default-free $X\%$ annuity bond and writing a call option on that bond (with an exercise price of par). This option component induces a concave relation between the price of MBSs and the price of default-free bonds (the so called "negative convexity").

2.1 MBS Pricing

Modeling and pricing MBSs involves two layers of complexity: (i) modeling the dynamic behavior of the term structure of interest rates, and (ii) modeling the prepayment behavior of mortgage holders. The standard way of solving this valuation problem has been to assume a stochastic process for term structure movements and to employ either an option pricing approach (e.g., Dunn and McConnell (1981a,1981b), Brennan and Schwartz (1985), Timmis (1985), Dunn and Spatt (1986), Johnston and Van Drunen (1988), and Stanton (1995)) or an empirical model (e.g., Schwartz and Torous (1989) and Waldman (1992)) for prepayment behavior. Given a set of m variables describing interest rates and prepayment behavior, and, using arbitrage-free valuation methods, mortgage-backed security prices can be written as

$$P_{mb,t} = V(\mathbf{x_t}, \theta)$$

where $P_{mb,t}$ is the price of the MBS at time t, and $V(\mathbf{x_t}, \theta)$ is a function of an m-vector of state variables $\mathbf{x_t}$ and an l-vector of parameters θ. The vector $\mathbf{x_t}$ includes interest rate variables (e.g., the level of interest rates) and possible prepayment specific variables (e.g., transaction costs of refinancing), while the vector θ is the set of parameters governing the economy. Both the rational and empirical approaches to prepayment modeling and MBS valuation depend crucially on the correct parameterization of prepayment behavior and on the correct model for interest rates. Previous empirical research focuses on estimation of these fundamental factors. Here, we take a different approach and study the direct relation between MBS prices and these fundamentals.

Note that $P_{mb,t}$ is a deterministic function of the state variables $\mathbf{x_t}$. For illustrative purposes, suppose that two state variables are sufficient both for interest rates and prepayments, and that the level of long-term interest rates, i_l, and the spread between long- and short-term rates, $i_l - i_s$ span this state space. Then two periods with the same interest rate environment, $(i_l^*, i_l^* - i_s^*)$, must also have the same MBS price. Even if the model is correctly specified, it is unrealistic to believe that the MBS price will be a deterministic function of the interest rate variables. In particular, at least for the MBS application studied in this paper, there are several sources for errors in MBS pricing. The first is that the MBS prices themselves are subject to measurement error. For example, bid prices vary slightly across dealers and may be asynchronous with respect to the interest rate quotes. Furthermore, the bid-ask spreads on the MBSs in this paper generally range from $\frac{1}{32}$nd to $\frac{4}{32}$nds, depending on the liquidity of the MBS. The second is that the MBS prices used in this paper refer to prices of unspecified mortgage pools in the marketplace (see Section 4.1). To the extent that the universe of pools changes from period to period, and its composition may not be in the agent's information set, this introduces an error into the pricing equation. Finally, there may be additional state variables which could lead to differential pricing. Therefore, we assume observed prices are given by

$$P_{mb,t} = V(\mathbf{x_t}, \theta) + \epsilon_t \tag{1}$$

where ϵ_t represent the aforementioned pricing errors.

In this paper, we estimate the functional form $V(\cdot)$ by estimating the joint density of MBS prices and various term-structure factors. Although it is necessary to choose the number of interest rate factors, the factors themselves can be unobservable. All that is required is that the interest rate variables used in the estimation are invertible to the true factors.[5] For example, if two factors can explain (i) the term structure (straight-bond component of MBSs), (ii) the mortgage rate (refinancing incentive), and (iii) prepayment characteristics (economic conditions), then, in the absence of estimation error, two interest rate variables should explain most of the important features of MBS pricing. In practice (i.e., in finite samples), these variables should be chosen to maximize the information content of the two

[5] In a related empirical investigation, Harvey (1991) also uses a density estimation procedure to estimate MBS prices. Harvey (1991), however, only considers one-factor models and focuses in particular on the ability of T-bond futures to hedge various GNMA prices during the 1986 period of volatile interest rates. Breeden (1992) provides a description of the MBS market during the 1980's, documenting negative convexity of MBS prices and the effects of prepayments and aggregate economic conditions. Of particular interest, he also uses a different nonparametric approach (based on the prevailing MBS market) to hedge GNMAs.

factors. We use the 10-year rate and the spread between the 10-year rate and the 3-month rate.

There are good reasons to choose the 10-year yield and the spread between this yield and the 3-month T-bill rate for capturing the salient features of MBSs. The MBSs analyzed in this paper have 30 years to maturity; however, due to potential prepayments and scheduled principal payments, their expected lives are much shorter. Thus, the 10-year yield should approximate the level of interest rates which is appropriate for discounting the MBS's cash flows. Further, the 10-year yield has a correlation of 0.98 with the mortgage rate (see Table 1B and Figure 3). Since the spread between the mortgage rate and the MBS's coupon determines the refinancing incentive, the 10-year yield should prove useful when valuing the option component.

The second variable, the slope of the term structure (in this case, the spread between the 10-year and 3-month rates) provides information on two factors: the market's expectations about the future path of interest rates, and the variation in the discount rate over short and long horizons. Steep term structure slopes imply lower discount rates for short-term cash flows and higher discount rates for long-term cash flows. Further, steep term structures may imply increases in future mortgage rates, which should increase the likelihood of mortgage refinancing.

Other factors, of course, may play an additional role in the valuation of MBSs. For example, Litterman and Scheinkman (1991) argue that the slope and level of the term structure are not sufficient to explain interest rate movements, and that the curvature of the term structure is also relevant (albeit less important). In addition, prepayments themselves may be related to factors other than those associated with interest rate movements. Market factors or structural factors may cause prepayment rates to change, affecting the pricing of MBSs. The importance of other factors is an empirical question that is addressed in some detail in Section 5.3.

2.2 MBS Pricing: An MDE Approach

We employ a kernel estimation procedure for estimating the relation between mortgage-backed prices and components of the term-structure of interest rates.[6] Kernel estimation

[6] For examples of MDE methods for approximating functional forms in the empirical asset pricing literature, see Pagan and Hong (1991), Harvey (1991) and Aït-Sahalia (1996). An alternative approach to estimating nonlinear functionals in the derivatives market is described by Hutchison, Lo and Poggio (1994). They employ methods associated with neural networks to estimate the nonlinear relation between option prices and the underlying stock price.

is a nonparametric method for estimating the joint density of a set of random variables. Specifically, given m-dimensional vectors $\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_T$ (e.g., MBS prices, the interest rate level, and the slope of the term structure) from an unknown density $f(\mathbf{z})$, then a kernel estimator of this density is

$$\hat{f}(\mathbf{z}) = \frac{1}{Th^m} \sum_{t=1}^{T} K\left(\frac{\mathbf{z} - \mathbf{z}_t}{h}\right), \tag{2}$$

where $K(\cdot)$ is a suitable kernel function and h is the window width or smoothing parameter.[7] This fixed window width estimator is often called the Parzen estimator. The density at any point is estimated as the average of densities centered at the actual data points. The further away a data point is from the estimation point, the less it contributes to the estimated density. Consequently, the estimated density is highest near high concentrations of data points and lowest when observations are sparse.

The econometrician has at his discretion the choice of $K(\cdot)$ and h. It is important to point out, however, that these choices are quite different from those faced by researchers employing parametric methods. Here, the researcher is not trying to choose functional forms or parameters that satisfy some goodness-of-fit criterion (such as minimizing squared errors in regression methods), but is instead characterizing the joint distribution from which the functional form will be determined.

One popular class of kernel functions is the symmetric beta density function, which includes the normal density, the Epanechnikov (1969) "optimal" kernel, and the commonly used biweight kernel as special cases. Results in the kernel estimation literature suggest that any reasonable kernel gives almost optimal results, though in finite samples there may be differences (see Epanechnikov (1969)). In this paper, we employ an independent multivariate normal kernel, though it should be pointed out that our results are relatively insensitive to the choice of kernel within the symmetric beta class.

The other parameter, the window width, is chosen based on the dispersion of the observations. For the independent multivariate normal kernel, Scott (1992) suggests the window width

$$\hat{h}_i = k_i \hat{\sigma}_i T^{\frac{-1}{m+4}},$$

where $\hat{\sigma}_i$ is the standard deviation estimate of each variable z_i, T is the number of observations, m is the dimension of the variables, and k_i is a scaling constant often chosen via

[7] If the smoothing factors vary across the variables, then h^m becomes $\prod_{i=1}^{m} h_i$, i.e., the product of the elements of the vector of window widths.

cross-validation. This window width (with $k_i = 1$) has the appealing property that, for certain joint distributions of the variables, it minimizes the asymptotic mean integrated squared error of the estimated density function. Unfortunately, our data are serially correlated and therefore the necessary distributional properties are not satisfied.

Consequently, we employ a cross-validation procedure to find the k_i which provides the right trade-off between bias and variance of the errors. Across all the data points, we find the k_i's which minimize the mean-squared error between the observed price and the estimated kernel price. This mean-squared error minimization is implemented using a Jackknife-based procedure. The implied MDE price at each data point is estimated using the entire sample, except for the actual data point and its nearest neighbors.[8] Once the k_i's are chosen, the actual estimation of the MBS prices and analysis of pricing errors involves the entire sample, albeit using window widths chosen from partial samples.

Consider the relation between three variables: MBS prices (P_{mb}), the level of long-term interest rates (i_l), and the slope of the term structure ($i_l - i_s$). Given the kernel estimate of the joint density of these variables, $\hat{f}(P_{mb}, i_l, i_l - i_s)$, the price at time t of an MBS at any interest rate level and term structure slope can be estimated by

$$\begin{aligned} \hat{P}_{mb,t}(i_{l,t}, i_{l,t} - i_{s,t}) &= V(i_{l,t}, i_{l,t} - i_{s,t}, \theta) \\ &= E[P_{mb,t} | i_{l,t}, i_{l,t} - i_{s,t}] \\ &= \int P_{mb,t} \frac{\hat{f}(P_{mb,t}, i_{l,t}, i_{l,t} - i_{s,t})}{\hat{f}(i_{l,t}, i_{l,t} - i_{s,t})} dP_{mb,t}, \end{aligned} \tag{3}$$

which is readily calculated from the kernel estimation of the joint density and the prices at each data point in the sample. Specifically,

$$\hat{P}_{mb}(i_l, i_l - i_s) = \frac{\sum_{t=1}^{T} P_{mb,t} K\left(\frac{i_l - i_{l,t}}{h_{i_l}}\right) K\left(\frac{[i_l - i_s] - [i_{l,t} - i_{s,t}]}{h_{i_l - i_s}}\right)}{\sum_{t=1}^{T} K\left(\frac{i_l - i_{l,t}}{h_{i_l}}\right) K\left(\frac{[i_l - i_s] - [i_{l,t} - i_{s,t}]}{h_{i_l - i_s}}\right)}, \tag{4}$$

where $K(z) = (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2} z' z}$. Thus, for any given long rate i_l^* and a given short rate i_s^*, there is a mapping to the MBS price $P_{mb}(i_l^*, i_l^* - i_s^*)$. These prices can then be used to evaluate how MBS prices move with fundamental interest rate factors.

Equation (4) takes a special form; the estimate of the MBS price can be interpreted as a

[8] Due to the serial dependence of the data, we performed the cross-validation omitting one year of data, i.e., six months in either direction of the particular data point in question.

weighted average of observed prices:

$$\hat{P}_{mb}(i_l^*, i_l^* - i_s^*) = \sum_{t=1}^{T} w_i(t) P_{mb,t} \ , \tag{5}$$

where

$$w_i(t) = \frac{K\left(\frac{i_l^* - i_{l,t}}{h_{i_l^*}}\right) K\left(\frac{[i_l^* - i_s^*] - [i_{l,t} - i_{s,t}]}{h_{i_l^* - i_s^*}}\right)}{\sum_{t=1}^{T} K\left(\frac{i_l^* - i_{l,t}}{h_{i_l^*}}\right) K\left(\frac{[i_l^* - i_s^*] - [i_{l,t} - i_{s,t}]}{h_{i_l^* - i_s^*}}\right)}.$$

Note that the weight $w_i(t)$ on observation i is a function of the "distance" of the state $(i_{l,t}, i_{l,t} - i_{s,t})$ from the current state $(i_l^*, i_l^* - i_s^*)$ (measured via the kernel function). The attractive idea behind MDE is that these weights are not estimated in an ad hoc manner, but instead depend on the true underlying distribution (albeit estimated) of the relevant variables. Thus, if the current state of the world, as measured by the state vector $(i_l^*, i_l^* - i_s^*)$, is not close to a particular point in the sample, then this sample price is given little weight in estimating the current price. Note, however, that MDE can give weight (possibly inconsequential) to all observations, so that the price of the MBS with $(i_l^*, i_l^* - i_s^*)$ also takes into account MBS prices at surrounding interest rates. This is an advantage of MDE, not a drawback. This is because MDE will help average out the different ϵ errors in equation (1) from period to period. Although our application utilizes only two factors, MDE will average out effects of other factors if they are independent of the two interest rate factors. Whether this is sufficient depends on the importance of these factors, and this issue is examined in Section 5.3.

The most serious problem with MDE is that it is data intensive. Much data is required in order to estimate the appropriate weights which capture the joint density function of the variables. The quantity of data which is needed increases quickly in the number of conditioning variables used in estimation. How well MDE does at estimating the relation between MBS prices and the interest-rate factors is then an open question, since the noise generated from the estimation error can be substantial.

In the next section, we examine the extent to which multivariate density estimation can uncover the relation between MBS prices and interest rates. As a first pass, we judge the MDE's ability to capture the salient features of MBS pricing in a simulated economy. While this economy is simple in structure, it provides a useful benchmark by which to judge the effectiveness of the MDE procedure. In particular, we wish to answer the following question: under which circumstances (i.e., number of observations, range of data) does the MDE perform well, that is, is the estimation error sufficiently small?

3 MBS Pricing in a Simulated Economy

In the appendix, we describe a simple multifactor MBS pricing model, based on the Schwartz and Torous (1989) model, which exhibits many of the commonly noted features of mortgage prepayment. In particular, the simulated prices have three important characteristics. First, the likelihood of prepayment increases as current interest rates fall relative to the coupon rate on the mortgage. In addition, the higher the fraction of a mortgage pool which has already prepaid, the lower the prepayment speed of the remaining loans in the pool (a simple form of burnout). Second, interest rates are described by a two-factor CIR model, using parameters estimated by Pearson and Sun (1994). To coincide with the range of interest rates, we look at the pricing of 7%, 10% and 13% MBSs. Third, since the pricing functional only has two sources of uncertainty, the MBS prices are deterministic functions of the two CIR factors. To coincide with a more realistic setting, a mean zero, uniformly distributed error over the range $[-\$0.50, \$0.50]$ is added to each MBS price. This can be viewed, for example, as a combination of the ϵ errors given in equation (1).

For this simulated model, Figure 1 graphs the prices of 7%, 10%, and 13% MBSs (with 30 years to maturity) against the 10-year yield for one particular simulation selected at random. The prices reflect both the negative convexity of MBSs and a second (albeit small) interest rate factor. At high interest rates, MBS prices behave much like those of a straight bond. They become concave in the interest rate level only at low interest rates when the refinancing incentive takes hold. This effect is apparent when comparing the 7%, 10%, and 13% MBSs. The higher the coupon, the higher the interest rate at which the refinancing option takes effect. The thickness of the pricing line, most evident at low interest rates, implies that there are multiple MBS prices for a given interest rate level, which can be explained by variation in the simulated model's second factor and the uniform pricing errors.

Since the pricing functional only has two main sources of uncertainty, a two-factor MDE should explain MBS prices well if there is no estimation error. Since the sample sizes are finite, however, estimation error is clearly present. In order to document the estimation error, we simulated 1000 independent economies, each with 500 observations on interest rates and MBS prices. For each economy, we estimated the MDE-implied pricing functionals for MBS prices using data on the first 50 and 150 observations, as well as using the full sample of 500 observations. In order to make the pricing functionals comparable across different sample sizes, the MBS prices were calculated using MDE for each sample size over the range of interest rates in the first 50 observations and using the bandwidth for the MDE computed with the standard deviation from the first 50 observations.[9] Absolute pricing errors were then calculated by taking the absolute value of the difference between the MDE's MBS price and the model's true MBS price (i.e., without pricing error) over this interest rate range for a cross-section of term structure spreads. Of particular interest, note that the MBS price is estimated from the MDE's pricing functional for sets of interest rates and spreads that may or may not have occurred in the sample. The only requirement is that these sets lie within the relevant interest rate and spread ranges.

Figure 2 documents the average absolute pricing error of the MDE procedure across the simulated economies and across several interest rate spreads. Note that the x-axis measures the level of the interest rate relative to the maximum and minimum observed in the first 50 observations. The points "0" and "100" correspond to the minimum and maximum, respectively. For a large range of interest rates in the sample, the MDE procedure's estimate of the MBS price coincides very closely with the true model price. For example, between the 15th and 85th percentile of interest rate ranges, the absolute pricing errors are around 20 cents, 40 cents and 60 cents for the 500-, 150- and 50-observation sample, respectively. Given that the observed prices are subject to uniformly distributed pricing errors, the MDE clearly performs well here. Since the par values are \$100, this represents approximately .2% absolute pricing error for the 500-observation sample. As we look to the higher interest rate levels within the sample, however, the estimation error increases.

Several observations on this estimation error increase are in order. First, the increase is worse for the smaller sample sizes. For example, the 50-observation sample has absolute pricing errors as high as \$1.50 (almost 1.5%), whereas the 500-observation sample has pricing errors of less than 70 cents (0.6%). Second, for the 50-observation sample, the pricing error increases dramatically outside of the interest rate range (i.e., below 0% and above 100%). This shows that MDE does not work well outside the data range; that is, the MDE interpolates the functional relation quite well, but does not extrapolate at the tails of the data.[10] Third, note that the 150- and 500-observation samples perform better than the 50-observation sample in the tails of the data. Recall that the interest rate range is chosen based on the 50-observation sample; thus, the other sample sizes may contain some observations in the $[-5\%, 0\%]$ and $[100\%, 105\%]$ ranges of the data, and the MDE will therefore not have

[9] While the range of interest rates over the simulations can be quite large, the range over the first 50 observations is considerably tighter.

[10] This problem is less apparent in the left tail of the interest rate distribution for the 13% GNMA. At very low interest rates, the option component is in-the-money and the MBS is much less sensitive to interest rate movements. Thus, extrapolation is not an issue.

to extrapolate MBS prices.

The overall conclusion from this particular simulated model is that the MDE works well, especially within the range of interest rates and spreads observed in the data. At the tails, however, the performance of the estimated pricing functional worsens. The ensuing errors seem to be monotonically related to the number of observations used in estimation. Nevertheless, even in the extremes of the data, the pricing errors are still on average less than 1%.

4 Data Description

4.1 Data Sources

Mortgage-backed security prices were obtained from Bloomberg Financial Markets covering the period January 1987 to May 1994. Specifically, we collected weekly data on 30-year fixed-rate Government National Mortgage Association (GNMA) MBSs, with coupons ranging from 7.5% to 10.5%.[11] The prices represent dealer-quoted bid prices on $X\%$ coupon-bearing GNMAs traded for delivery on a *to be announced* (TBA) basis.

The TBA market is most commonly employed by mortgage originators who have a given set of mortgages that have not yet been pooled. However, trades can also involve existing pools on an unspecified basis. Rules for the delivery and settlement of TBAs are set by the Public Securities Association (PSA) (see, for example, Bartlett (1989)). For example, an investor might purchase \$1 million worth of 8% GNMAs for forward delivery next month. The dealer is then required to deliver 8% GNMA pools within 2.5% of the contracted amount (i.e., between \$975,000 and \$1,025,000), with specific pool information to be provided on a TBA basis (just prior to settlement). This means that, at the time of the agreed-upon-transaction, the characteristics of the mortgage pool to be delivered (e.g., the age of the pool and its prepayment history) are at the discretion of the dealer. Nevertheless, for the majority of the TBA's, the delivered pools represent newly issued pools.

To the extent that some of the prices of these GNMA pools in the TBA market may be aged (and thus subject to prepayment history), we also collected monthly information on the general mortgage pool of all available 30-year $X\%$ GNMAs over the 1987-1994 period. In

[11] Careful filters were applied to the data to remove data reporting errors using prices reported in the Wall Street Journal. Furthermore, data are either not available or sparse for some of the GNMA coupons during the period. For example, in the 1980's, 6% coupon bonds represent mortgages originated in the 1970's, and not the more recent issues which are the focus of this paper. Thus, data on these MBSs were not used.

particular, we obtained information regarding the pools' conditional prepayment rate (CPR, the principal prepaid in the month divided by the principal outstanding at the beginning of the month), the dollar value of the original amount of all the mortgage pools, and the dollar value of the remaining principal of these pools. Prepayment data is of interest because it refers to all the available GNMA mortgage pools, and thus provides information on the universe of pools available for delivery in the TBA market.

With respect to the interest rate series, weekly data for the 1987-1994 period were collected on the average rate for 30-year mortgages (collected from Bloomberg Financial Markets),[12] and the yields on the 3-month Treasury bill, 1-year Treasury note, 5-year Treasury note and 10-year Treasury note (provided by the Board of Governors of the Federal Reserve).

4.2 Data Characteristics

Before describing the pricing results and error analysis for MBSs using the MDE approach, we briefly describe the environment for interest rates and mortgage rates during the sample period, 1987-1994.

Characteristics of Mortgages (1987-1994)

Since the mortgage rate represents the available rate at which homeowners can refinance, it plays an especially important role with respect to this incentive. For example, consider a 9% GNMA security. Note that this 9% GNMA is backed by 9.5% 30-year mortgages since there is a 0.5% servicing fee associated with GNMA pools. Figure 3 graphs the mortgage rate for 1987 through 1994. From 1987 to 1991, the mortgage rate varied from 9% to 11%. In contrast, from 1991 to 1994, the mortgage rate generally declined from 9.5% to 7%.

For pricing GNMA TBAs, it is most relevant to understand the characteristics of the universe of pools at a particular point in time. That is, the fact that a number of pools have prepaid considerably may be irrelevant if newly originated pools have entered into the MBS market. To get a better idea of the time series behavior of the GNMA TBAs during this period, Figure 4 graphs an artificially constructed index of all the originations of 7.5% to 10.5% GNMA pools from January 1983 to May 1994.[13]

[12] Bloomberg's source for this rate is "Freddie Mac's Primary Mortgage Market Survey", which reports the average rate on 80% of newly originated 30-year, first mortgages on a weekly basis.

[13] The dollar amount outstanding for each coupon is normalized to 100 in January 1987. Actual dollar amounts outstanding in that month were $10,172, $27,096, $10,277, $63,392, $28,503, $15,694, and $5,749 (in millions) for the 7.5% – 10.5% coupons, respectively.

There is a wide range of origination behavior across the coupons. As mortgage rates moved within a 9% to 11% band between 1987 to 1991, Figure 4 shows that GNMA 9s, 9.5s, 10s and 10.5s were all newly originated during this period. Consistent with the decline in mortgage rates in the post 1991 period, GNMA 7.5s, 8s and 8.5s originated while the GNMA 9s–10.5s became seasoned issues. Thus, in terms of the seasoning of the pools most likely to be delivered in the TBA market, there are clearly cross-sectional differences between the coupons.

Figure 4 shows that there are several reasons for choosing the TBA market during the post 1986 time period to investigate MBS pricing using the MDE methodology. First, during 1985 and 1986, interest rates dramatically declined, leading to mortgage originations for a wide variety of coupon rates. Thus, the GNMA TBAs in 1987-1994 correspond to mortgage pools with little prepayment history (i.e., no *burnout*) and long maturities. In contrast, prior to this period, the 7.5% to 10.5% GNMAs were backed by mortgages originated in the 1970's and thus represented a different security (in both maturity and prepayment levels). Second, MDE pricing requires joint stationarity between MBS prices and the interest rate variables. This poses a potential problem in estimating the statistical properties of any fixed maturity security, since the maturity changes over time. Recall that the TBA market refers to unspecified mortgage pools available in the marketplace. Thus, to the extent that there are originations of mortgages in the GNMA coupon range, the maturity of the GNMA TBA is less apt to change from week to week. Figure 4 shows that this is the case for the higher coupon GNMAs pre 1991, and for the low coupon GNMAs post 1991. Of course, when no originations occur in the coupon range (e.g., the GNMA 10s in the latter part of the sample), then the maturity of the available pool will decline. In this case, the researcher may need to add variables to capture the maturity effect and possibly any prepayment effects. In our initial analysis, we choose to limit the dimensionality of the multivariate system, and instead focus on the relation between MBS prices and the two interest rate factors. Section 5.3 provides an investigation of additional factors.

Characteristics of the Prepayment Option (1987-1994)

Table 1 provides ranges, standard deviations and cross-correlations of GNMA prices (Table 1A), and mortgage and interest rates (Table 1B) during the 1987-1994 period. Absent prepayments, MBSs are fixed-rate annuities, and the dollar volatility of an annuity increases with the coupon. In contrast, from Table 1, we find that the lower coupon GNMAs are more volatile than the higher coupon GNMAs. This suggests that the lower volatility of the higher coupon GNMAs is due to the embedded call option of MBSs. The important element of the option component for MBS valuation is the refinancing incentive. For most of the sample (especially 1990 on), the existing mortgage rate lies below 10.5% and the prepayment option is at- or in-the-money.[14] Historically, given the costs associated with refinancing, a spread of approximately 150 basis points between the old mortgage rate and the existing rate is required to induce rapid prepayments.[15] The lack of seasoning aside, this would suggest that the higher coupon GNMAs began to prepay in the early 90's.

As evidence of this, Figure 5 graphs the annualized conditional prepayment rate (CPR) for the 7.5% to 10.5% GNMAs throughout the 1987 to 1994 period. Figure 5 shows that, after 1991, the higher coupon GNMAs had monthly CPRs as high as 60% on an annualized basis. This fact points to several key features of MBS pricing. First, the moneyness (i.e., the value of the prepayment option) of the universe of higher coupon GNMA pools clearly increased as interest rates fell; that is, individuals began to rationally prepay their mortgages. Second, since a substantial number of these pools prepaid during this period, the possibility of *burnout* exists for the higher coupon GNMAs. Since *burnout* reflects that fact that some individuals only prepay at very low mortgage rates (due to high transactions costs) or indeed never prepay (due to inertia on their part), the pricing of higher coupon GNMAs may be quite complex in the latter part of the sample. This may not only lead to cross-sectional differences across coupons, but also point to additional factors necessary for GNMA pricing (see Section 5.3). Finally, note, however, that Figure 5 refers to the entire universe of GNMA pools and not necessarily the pools most likely to be delivered in the TBA market. Thus, the magnitude of prepayments may be a misleading measure of the prepayment history of the TBA pools. For example, in early 1987, the GNMA 10.5s had substantial prepayments. However, clearly, these prepaid pools are not the relevant ones for GNMA TBA pricing in our sample. This can be seen from Figure 4 which shows that there are substantial originations of GNMA 10.5s after the 1987 period.

Characteristics of Interest Rates (1987-1994)

[14] Figure 3 also graphs one of the interest rate factors, the 10-year yield. The correlation between the 10-year rate and the mortgage rate over our sample period is .980. However, there is a difference in the level between the two series (i.e., on average 1.56%), representing the cost of origination, the option value, and the bank profits, among other factors.

[15] See Bartlett (1989) and Breeden (1991) for some historical evidence of the relation between prepayment rates and the mortgage spread. Note that in the 1990's the 150 basis point spread has been somewhat lower — in some cases, 75 to 100 basis points. Some have argued that this is due to the proliferation of new types of mortgage loans (and ensuing marketing efforts by the mortgage companies) (Bartlett (1989)), though it may also be related to aggregate economic factors, such as the implications of a steep term structure.

As mentioned above, Figure 3 graphs the 10-year yield against the mortgage rate. During the 1987 to 1994 period, there are multiple observations of particular interest rates. Since these multiple observations occur at different points of the sample, this will help MDE isolate the potential impact of additional interest rate factors, as well as reduce maturity effects not captured by the MDE pricing (see *Characteristics of Mortgages* above). Similarly, while the spread between the 10-year yield and the 3-month rate is for the most part positive, there is still variation of the spread during the period of an order of magnitude similar to the underlying 10-year rate (see Table 1B). Moreover, the correlation between these variables is only -0.45, indicating that they potentially capture independent information, which may be useful for pricing GNMAs.

5 Empirical Results

5.1 One-Factor Pricing

As a first pass at the MBS data, we describe the functional relation between GNMA prices and the level of interest rates (the 10-year yield). As an illustration, Figure 6 graphs the estimated 9% GNMA price with the actual data points. The smoothing factor, which is chosen by cross-validation, is 0.35 (i.e., $k_i = 0.35$).

Several observations are in order. First, the figure illustrates the well-known negative convexity of MBSs. Specifically, the MBS price is convex in interest-rate levels at high interest rates (when it behaves more like a straight bond), yet concave at low interest rates (as the prepayment option becomes in-the-money). Second, the estimated functional relation is not smooth across the entire range of sample interest rates. Specifically, between 10-year yields of 7.1% to 7.8%, there is a *bump* in the estimated relation. While this feature is most probably economically spurious, it reflects the fact that the observed prices in this region are high relative to the prices at nearby interest rates. Increasing the degree of smoothing eliminates this bump at the cost of increasing the pricing errors. The source of this variation, which could be missing factors, MDE estimation error or structural changes in the mortgage market, is investigated further below. Third, there is a wide range of prices at the same level of interest rates. For example, at a 10-year yield of 8%, prices of GNMA 9s vary from 98% to 102% of par. Is this due to the impact of additional factors, measurement error in GNMA prices, MDE estimation error, or some other phenomenon?

Table 2 provides some preliminary answers to this question. Specifically, Table 2A reports summary statistics on the pricing errors (defined as the difference between the MDE estimated price and the observed MBS price) for the 7.5% to 10.5% GNMAs. As seen from a comparison between Table 1A and 2A, most of the volatility of the GNMA price can be explained by a 1-factor kernel using the interest rate level. For example, the volatility of the 9% GNMA is \$5.26, but its residual volatility is only \$0.83. However, while 1-factor pricing does well, it clearly is not sufficient as the pricing errors are highly autocorrelated (from 0.861 to 0.927) for all the GNMA coupons. Though this autocorrelation could be due to measurement error induced by the MDE estimation, it does raise the possibility that there is a missing factor. In addition, the residuals are highly correlated across the 7 different coupon bonds (not shown in the Table). Thus, the pricing errors contain substantial common information.

This correlation across different GNMAs implies that an explanation based on idiosyncratic information (such as measurement error in prices) will not be sufficient. Combined with the fact that the magnitude of the bid-ask spreads in these markets lies somewhere between $\frac{1}{32}$nd and $\frac{4}{32}$nds, clearly measurement error in observed prices cannot explain either the magnitude of the pricing errors with 1-factor pricing (e.g., \$2 – \$3 in some cases) or the substantial remaining volatility of the errors (e.g., \$0.70 to \$0.84 across the coupons).

Table 2B looks at the impact of additional interest rate factors. We run a regression of the pricing errors on the level and squared level to check whether any linear or nonlinear effects remain. For the most part, the answer is no. The level has very little explanatory power for the pricing errors, with R^2s ranging from 1.1% to 2.8%. Moreover, tests of the joint significance of the coefficients cannot reject the null hypothesis of no explanatory power at standard significance levels. Motivated by our discussion in Sections 2 and 3, we also run a regression of the pricing errors for each GNMA on the slope of the term structure (the spread between the 10-year yield and the 3-month yield) and its squared value. The results strongly support the existence of a second factor, with R^2s increasing with the coupon from a low of 2.0% to 40.1%. Furthermore, this second factor comes in nonlinearly as both the linear and nonlinear term are large and significant.

Most interesting is the fact that the slope of the term structure has its biggest impact on higher coupon GNMAs. This suggests an important relation between the prepayment option and the term structure slope. Due to the relatively lower value of the prepayment option, low coupon GNMAs behave much like straight bonds. Thus, the 10-year yield may provide enough information to price these MBSs. In contrast, the call option component of higher coupon GNMAs is substantial enough that the duration of the bond is highly variable.

Clearly, the slope of the term structure provides information about the variation in yields across these maturities; hence, its additional explanatory power for higher coupon GNMAs. The negative coefficient on the spread implies that the 1-factor MDE is underpricing when the spread is high. In other words, when spreads are high, and short rates are low for a fixed long rate, high coupon GNMAs are more valuable than would be suggested by a 1-factor model. The positive coefficients on the squared spread suggest that the relation is nonlinear, with a decreasing effect as the spread increases. Note that in addition to information about variation in discount rates across maturities, the spread may also be proxying for variation in expected prepayment rates that is not captured by the long rate.

5.2 Two-Factor Pricing

Motivated by the results in Table 2B, it seems important to consider a second interest rate factor for pricing MBSs. Therefore, we describe the functional relation between GNMA prices and two interest-rate factors, the level of interest rates (the 10-year yield) and the slope of the term structure (the spread between the 10-year yield and the 3-month yield). In particular, we estimate the pricing functional given in equation (5) for each of the GNMA coupons. For comparison purposes with Figure 6, Figure 7 graphs the 9% GNMA against the interest rate level and the slope. The smoothing factor for the long rate is fixed at the level used in the 1-factor pricing (i.e., 0.35), and the cross-validation procedure generates a smoothing factor of 1.00 for the spread.

The well-known negative convexity of MBSs is very apparent in Figure 7. However, this functional form does not hold in the northwest region of the figure, that is, at low spreads and low interest rates. Recall from Section 3 that the MDE approach works well in the regions of the available data, but extrapolates poorly at the tails of the data and beyond. Figure 8 graphs a scatter plot of the interest rate level against the slope. As evident from the figure, there are periods in which large slopes (3%-4%) are matched with both low interest rates (in 1993-1994) and high interest rates (in 1988). However, few observations are available at low spreads joint with low interest rates. Thus, the researcher needs to be cautious when interpreting MBS prices in this range.

Within the sample period, the largest range of 10-year yields occurs around a spread of 2.70%. Therefore, we take a slice of the pricing functional for the 8%, 9% and 10% GNMAs, conditional on this level of the spread. Figure 9 graphs the relation between GNMA prices for each of these coupons against the 10-year yield. Several observations are in order. First, the negative convexity of each MBS is still apparent even in the presence of the second factor. Though the *bump* in the functional form is still visible, it has been substantially reduced. Thus, multiple factors do play a key role in MBS valuation. Second, the price differences between the various GNMA securities narrow as interest rates fall. This just represents the fact that higher coupon GNMAs are expected to prepay at faster rates. As GNMAs prepay at par, their prices fall because they are premium bonds, thus reducing the differential between the various coupons. Third, the GNMA prices change as a function of interest rates at different rates depending on the coupon level, i.e., on the magnitude of the refinancing incentive. Thus, the effective duration of GNMAs varies as the moneyness of the prepayment option changes.

The results of Section 5.1, and Figures 8 and 9, suggest the possible presence of a second factor for pricing MBSs. To understand the impact of the term structure slope, Figure 10 graphs the various GNMA prices against interest rate levels, conditional on two different spreads (2.70% and 0.30%).[16] Recall that the slope of the term structure is defined using the yield on a full-coupon note, not a ten-year zero-coupon rate. As a result, positive spreads imply upward sloping full-coupon yield curves and even more steeply sloping zero-coupon yield curves. In contrast, when the spread is close to zero, both the full-coupon and zero-coupon yield curves tend to be flat. Thus, holding the 10-year full-coupon yield constant, short-term (long-term) zero-coupon rates are lower (higher) for high spreads than when the term structure spread is low.

In terms of MBS pricing, note that at high interest rate levels, the option to prepay is out-of-the-money. Consequently, many of the cash flows are expected to occur as scheduled, and GNMAs have long expected lives. The appropriate discount rates for these cash flows are therefore longer-term zero-coupon rates. Consider first the effects on the price of an 8% GNMA. Since this security has its cash flows concentrated at long maturities, its price should be lower for higher spreads, just as we observe in Figure 10. On the other hand, the option component of the 10% GNMA is much closer to being at-the-money, even for the highest interest rates shown in the figure. Hence, at these interest rates, 10% GNMA prices do not follow the same ordering as 8% GNMAs vis-a-vis the level of the spread.

As interest rates fall, prepayments become more likely, and the expected life of the MBS falls for GNMAs of all coupons. As this life declines, the levels of the shorter-term zero-coupon rates become more important for pricing. In this case, high spreads imply lower

[16]The spreads and interest rate ranges are chosen to coincide with the appropriate ranges of available data, to insure that the MDE approach works well.

discount rates at the relevant maturities, for a fixed 10-year full-coupon yield. Consequently, when the GNMAs are priced as shorter-term securities due to high expected prepayments, high spreads imply higher prices for all coupons. This implication is illustrated in Figure 10. While prices always increase for declining long rates, the increase is much larger when spreads are high. For the 8% GNMA, this effect causes the prices to cross at a long rate of approximately 8.3%, while for the 10% GNMA it causes the pricing functionals to diverge further as rates decrease. The effect in Figure 10 is primarily driven by changes in expected cash flow life. The 10-year yield proxies for the moneyness of the option, the expected level of prepayments, and the average life of the cash flows. The addition of the second factor, the term structure slope, also controls for the average rate at which these cash flows should be discounted.

In order to understand the impact of 2-factor pricing more clearly, Table 3 provides an analysis along the lines of Table 2 for 1-factor pricing. Specifically, Table 3A reports some summary statistics on the pricing errors for the 7.5% to 10.5% GNMAs. The addition of a second interest rate factor reduces the pricing error volatility across all the GNMA coupons, i.e., from \$0.70 to \$0.65 for the 7.5s, \$0.83 to \$0.61 for the 9s, and \$0.84 to \$0.52 for the 10.5s. Most interesting, the large reduction in pricing error volatility occurs with the higher coupon GNMAs, which confirms the close relation between the slope of the term structure and the prepayment option. Table 3B looks at whether there is any remaining level or slope effect on the 2-factor MBS prices. We run nonlinear regressions of the pricing errors on the level and the slope separately. Neither the level nor the slope have any remaining economic explanatory power for the pricing errors, with R^2s ranging from 3.6% to 4.5% for the former and R^2s under 1.0% for the latter. The tests of joint significance of the coefficients exhibit marginal significance for the level, suggesting that reducing the smoothing parameter will generate a small improvement in the magnitude of the pricing errors.

While these results suggest that a 2-factor kernel is sufficient to explain level and slope effects, the results are not entirely satisfactory for several reasons.[17] First, substantial autocorrelation of the pricing errors persist. Whether this autocorrelation is due to estimation error induced by MDE or additional factors is an open question. Second, the volatility of the pricing errors (i.e., around \$0.50) is higher than that implied by measurement error in observed prices. Third, and most important, the pricing errors across the GNMA coupons are

[17]While it may appear that the level and slope are, by construction, necessarily explained within the sample, this is not the case. Unlike nonlinear least squares, MDE does not fit the functional form in such a way that the regression errors are constrained to be uncorrelated with the predictive variables.

highly cross-correlated. Table 4 provides estimates of these correlations across the GNMA 7.5s to 10.5s for the 2-factor pricing case. The correlations for nearby coupons are as high as 0.982 (i.e., between the 8.5s and 9s), and generally are above 0.650 (the only exception being the GNMA 10.5s).

A different way of looking at the problem is to perform a principal components analysis of the pricing errors across the 7.5%–10.5% GNMAs. Although this measure is linear, it provides a first pass at documenting the size of the common components across the coupons. We find that the first component explains 85% of the variation. Moreover, to the extent that the 10.5% GNMA is less correlated with the other coupons, the 85% estimate is actually higher across the remaining GNMA bonds. This principal components analysis suggests that some common information across the coupons is still missing from the MDE pricing of the GNMAs. Most puzzling is the fact that the most obvious type of information ignored in the pricing, such as maturity effects and non-interest rate specific prepayment effects, seems to be an unlikely candidate. Figures 4 and 5 show that origination behavior and prepayment histories differ substantially across the GNMA coupons, which (if important for pricing) would drive a wedge between the cross-correlations of the pricing errors. Nevertheless, it seems worthwhile exploring the potential impact of other factors on MBS pricing.

6 The Cross-Correlation Puzzle

The results in Sections 4 and 5 present a puzzle to financial economists. On the one hand, there are substantial differences across the GNMA 7.5s to 10.5s in their point of origination, their prepayment behavior and their price behavior. On the other hand, even though we have accounted for these effects in our empirical analysis, three important characteristics of the pricing errors remain: (i) dollar volatility of the order of $0.52 to $0.65, (ii) persistence of these errors with autocorrelations as high as 0.85, and (iii) large cross-correlations of these errors across the coupon bonds. There are four possible explanations for this puzzle, namely (1) an additional factor, (2) spurious regression, (3) a structural break, or (4) estimation error generated by MDE.

6.1 Other Factors

If there is a factor missing from the model, and if that factor exhibits autocorrelation, we should expect to see both cross-correlation and serial correlation in our pricing errors. We therefore investigate the relationship between the pricing errors and three likely candidates for such an omitted factor: (i) an additional interest rate factor, (ii) a prepayment factor based on the *seasoning* of the pools, and (iii) a prepayment factor based on the *burnout* of the pools.

Interest Rate Factor

Litterman and Scheinkman (1991) suggest the possible presence of a third interest rate factor based on the curvature of the yield curve. Here, we define curvature as half the sum of the 5-year and 3-month rates minus the 1-year rate. Table 5 reports nonlinear regressions of various factors on the pricing errors of the GNMA 7.5s to 10.5s from the 2-factor kernel estimation. There is some evidence that MBSs contain a third interest rate factor. Curvature reduces the volatility of the pricing errors across all the coupons, and explains roughly 11%-12% of the remaining variation of the pricing errors across the GNMAs 7.5s to 10.5s. The coefficients are jointly significant at the 1% level in every case. Because curvature's effect is similar across all the bonds, it suggests its impact may have more to do with the straight bond component of MBSs than the embedded call option of refinancing.

The results in Table 5 imply that curvature has a similar effect on all the GNMA bonds; therefore, perhaps curvature can explain the common behavior of the pricing errors. Although the residuals from the nonlinear regression of the GNMA pricing errors on curvature are correlated across coupons, it may be that the second order Taylor series parameterization is not well enough specified. Table 6 provides empirical results for a 3-factor kernel using the level, the slope and the curvature of the yield curve. On a positive note, the range of volatilities of the pricing errors drops to between \$0.44 and \$0.47. Furthermore, the autocorrelations of these errors, while still high, drop about 15–20%. The interest rate and prepayment factors also lose much of their explanatory power for the pricing errors, and this is true across coupons. On a negative note, however, the cross-correlation of these errors hardly changes at all. Therefore, even though a third interest rate factor may exist, it is not the main determinant of the common source of variation across coupons.

Prepayment Factors

The prepayment behavior of mortgage holders does not vary one-to-one with interest rates because individuals prepay for reasons not associated with the value of the option to refinance the mortgage. For example, mortgages are prepaid for "non-economic" reasons such as relocation, change in the household's family status, or default. In this context, two commonly given explanations for a pool's prepayment behavior are:

1. The pool's *seasoning*.

   Prepayment rates on mortgages will initially tend to increase with the age of the mortgage, since there are frictions to household changes. For example, brand new mortgages are unlikely to have been taken out if the holders thought they were to relocate or default.

2. The pool's *burnout*.

   Aged (and substantially prepaid) pools in a positive coupon spread environment, there is a tendency for low future prepayments. The intuition is that if a mortgage holder were going to prepay, then he/she would have already done so. This *burnout* effect could reflect nonoptimal behavior on the part of some mortgage holders, or frictions they face in trying to refinance their property (e.g., the value of the house may have fallen by such an amount that refinancing is no longer possible, yet there are sufficient costs to defaulting).

We use the following measures for the prepayment factors:

1. For each GNMA coupon, we measure *seasoning* by the time passed since its last set of originations. (See Figure 4 for a graph of each coupon's origination history).

2. For each GNMA coupon, we measure *burnout* in two ways:

   (a) The cumulative product of the CPRs since the last origination. This variable captures the prepayment history of the universe of pools for each coupon; however, as mentioned previously, the universe of pools includes aged pools which may not be relevant for TBA pricing.

   (b) The number of periods in which the mortgage rate lies below the coupon rate since the last origination, adjusted for the magnitude of the difference. Specifically, following Schwartz and Torous (1989), we use $\sum_{t=-\tau}^{0} \max(C_{GNMA} - i_{mt}, 0)$, where τ marks the period of last origination and i_{mt} represents the mortgage rate at time t.

Although the nature of the TBA market mitigates these effects, as a pools' *seasoning* or *burnout* may not reflect the pools most likely to be delivered, it seems worthwhile exploring their impact on MBS TBA pricing.

As seen from Table 5, the three measures have limited explanatory power for the pricing errors from the 2-factor kernel. The exceptions are the low coupon GNMAs which have as much as 28% of their pricing error explained by the pool's seasoning and by the pool's cumulative CPR. One explanation for this result is that the prepayment option has little value for the lower coupon GNMAs. Thus, prepayments of these lower coupon GNMAs can be almost entirely explained by "non-economic" factors. To the extent that, between 1987 to 1991, there were almost no originations of lower coupon bonds, the TBA market contained *seasoned* pools for lower coupon GNMAs. Thus, some of the variation of the MBS prices for these GNMAs must be due to uncertainty about non-economic prepayments. These prepayments are, however, so minor relative to the magnitude of the prepayment option for higher coupon GNMAs that they have very little impact on higher coupon bonds.

6.2 Spurious Regression

The high autocorrelation of all the pricing errors suggests a partial explanation of the high cross-correlation of these errors. Even if there is no relation between these errors across coupons, the errors might be subject to the well known spurious regression bias. That is, in small samples, unrelated autocorrelated series give the appearance of being related. One way to test this hypothesis is to examine the first differences of the series. That is, if the source of the common variation is spurious, we should expect that the cross-correlations of the change in the pricing errors to be much smaller than the cross-correlations of the errors. Table 7 reports these cross-correlations for the 3-factor kernel. A comparison of Table 6 and Table 7 reveals only minor differences between the cross-correlation patterns of the pricing errors and the change in these errors.[18] Apparently, spurious regression is not the explanation for the cross-correlation puzzle.

6.3 Structural Break

The above explanations assume that there were no structural shifts in the mortgage market during the 1987–1994 period; thus, sample statistics, e.g., correlations across coupons, are well specified. During the early 1990's, a number of innovations occurred in the mortgage market, such as lower points, increased marketing efforts, and shorter term mortgages. While these innovations may be due to the economic environment, as described by, for example, the

[18]These results carry through for the 1-factor and 2-factor kernels as well.

term structure, some researchers believe these represent a structural change in the market. To the extent this structural change impacts prepayment behavior, it may be possible that the high correlation between pricing errors is reflecting the omission of this structural shift. As an example, Figure 11 provides a scatter plot of the 9% GNMA prices against interest rates, separating the observations out into two distinct subperiods. A cursory look at the figure suggests a possible shift upward in the prices of GNMAs in the latter period. That is, for a given interest rate level, prices are much higher. Note that a shift up in prices is the opposite of what would be expected given innovations that make refinancing less costly. However, this figure helps explain why MDE generates a *bump* in the MBS price's functional form — the structural change shifts prices upward, leading to a jump in the estimated conditional mean of the MBS price. Of course, an alternative explanation might be that there are different term structure spreads within these periods, or different origination or prepayment behavior not related to structural changes in the market.

In order to resolve this question, Table 8A reports a subperiod analysis of these pricing errors. In the first subperiod, the mean of the pricing errors across the GNMAs tends to be positive ($0.26), while they are negative in the latter period (-$0.15). However, the average cross-correlation between the errors does not decline, and in fact is remarkably similar, and high, in the two periods. Of course, the errors are derived from the difference between an MDE implied price and the observed price. Since the MDE implied price uses all the observations in the sample, we also performed MDE separately for the two subperiods. These separate estimations rid the two samples of the bias (i.e., the mean) of the pricing errors, and lower the volatility of these errors by 20-25%. The cross-correlations between the errors, however, are not reduced in any meaningful way in either subperiod. Thus, although these results do suggest a possible structural break in the data (e.g., zero means and lower volatility), this break is clearly not the source of the common variation that still exists across the GNMA coupons.

6.4 Estimation Error

Two results emerge from the empirical analysis of Section 5. First, there is a common factor that explains much of the variation across the pricing errors. Second, the MBS prices across the coupons react similarly to the interest rate factors, albeit with some differences due to the value of the prepayment option. Perhaps, given these two facts, the MDE procedure itself leads to *excess* correlation across the GNMAs. This excess correlation may be due to similar measurement error induced by the MDE procedure. Since all the pricing errors are residuals from estimation using the same predictor variables (i.e., the level and the slope of the term structure), measurement error is a logical candidate. The most common measurement error with MDE occurs when the data is sparse at the evaluation point of the functional (i.e., the MBS price at a given level and slope).[19] Since the sparseness of the data is common across all coupons, one might expect some degree of correlation across the errors for this reason alone.

While this measurement error is difficult to test for explicitly, we broke the sample into two types of observations, namely ones which are close to other data and ones which are far away.[20] Table 8B reports summary statistics for the pricing errors generated from these two different samples. In terms of the relation between the pricing errors and other interest rate factors, there does not seem to be a systematic effect based on whether the data is dense or sparse in a particular region of these interest rate factors. More interesting, the average cross-correlations across the pricing errors are maintained at high levels when the data is broken down into the dense and sparse regions of the sample. Of course, while this analysis does not rule out all forms of MDE estimation error, it does suggest that the most likely candidate is not the source of the correlation across the errors.

7 Conclusion

In this paper, we provide an empirical analysis of mortgage-backed security pricing using multivariate density estimation techniques. Instead of postulating and estimating parametric models for both interest rate movements and prepayments, as in previous approaches to mortgage-backed security valuation, we estimate directly the functional relation between mortgage-backed security prices and the level of economic fundamentals. This approach can yield consistent estimates without the need to make the strong assumptions about the processes governing interest rates and prepayment required by previous approaches.

Several interesting observations can be made. First, using simulated data, we confirm that the MDE procedure works well except when trying to extrapolate beyond the range of the data. Second, using weekly prices for GNMA MBSs between 1987 and 1994, we find that these prices can be well described as a function of the level of interest rates and the

[19]See, for example, the discussion on page 6 and the simulation exercise of Section 3.

[20]Specifically, observations are ranked based on the distance to their 10th nearest neighbor. The observations with distances less than the median distance are considered to be in "dense" regions of the data, the observations in the other half of the sample are considered to be in "sparse" regions of the data.

slope of the term structure. A single interest rate factor, as used in most previous mortgage valuation models, is insufficient. The relation between prices and interest rates displays the usual stylized facts, such as negative convexity in certain regions, and a narrowing of price differentials as interest rates fall. Most interesting, the term structure slope plays an important role in valuing MBSs via its relation to the interest rate level and the refinancing incentive associated with a particular MBS. Third, even after taking account of the nonlinear relation between MBS prices and the interest rate factors, a common component across all the MBS coupons is still present. In fact, a principal components analysis suggests 85% of the remaining variation across the coupons can be explained by a common component. The size of this common component is surprising given that each GNMA coupon displays different origination and prepayment behavior. Fourth, although this common component can be related to a variety of economic variables (e.g., other interest rate and prepayment factors), the explained variation falls well short of the 85% magnitude that exists across the coupons. Moreover, other logical explanations (not based on additional factors) do not explain this magnitude either. We leave the solution of this cross-correlation puzzle to future research.

On a more general note, the MDE procedure will work well (in a relative sense) under the following three conditions. First, since density estimation is data intensive, the researcher either needs a large data sample or an estimation problem in which there is little disturbance error in the relation between the variables. Second, the problem should be described by a relative low dimensional system, since MDE's properties deteriorate quickly when variables are added to the estimation. Third, and especially relevant for comparison across methods, MDE will work relatively well for highly nonlinear frameworks. As it happens, these features also describe derivative pricing. Hence, while the results we obtain here for GNMAs are encouraging, it is likely that the MDE approach would fare well for more complex derivative securities. Though the TBA market is especially suited for MDE analysis due to its reduction of the maturity effect on bonds, it may be worthwhile investigating the pricing of interest only (IO) and principal only (PO) strips, and collateralized mortgage obligations (CMOs). Since the relation between the prices of these securities and interest rates is more highly nonlinear than that of a GNMA, a multifactor analysis might shed light on the interaction between various interest rate factors and the underlying prices. The advantage of the MDE approach is its ability to capture arbitrary nonlinear relations between variables, making it ideally suited to capturing the extreme convexity exhibited by many derivative mortgage-backed securities.

Appendix: Theoretical MBS Pricing Model

Interest Rates

Assume interest rates are described by the two-factor interest rate model estimated and tested by Pearson and Sun (1994). The two factors are the instantaneous riskless real interest rate, r, and the expected inflation rate, y. The real interest rate is given by $r_t = r'_t + \overline{r}$, where $\overline{r}$ is a constant, and

$$dr'_t = \kappa_1(\theta_1 - r'_t)\,dt + \sigma_1\sqrt{r'_t}\,dZ^1_t. \tag{6}$$

The expected inflation rate moves according to the equation

$$dy_t = \kappa_2(\theta_2 - y_t)\,dt + \sigma_2\sqrt{y_t}\,dZ^2_t, \tag{7}$$

where the two Brownian motions dZ^1_t and dZ^2_t are uncorrelated. The price level, p, moves according to the equation

$$dp_t = y_t p_t\,dt + \sigma_p p_t\sqrt{y_t}\,dZ^3_t,$$

where $E(dZ^2_t dZ^3_t) = \rho\,dt$. When $\overline{r} = 0$, this reduces to the standard 2 factor CIR model. The equilibrium risk premium for real bonds is $\lambda r'$. Pearson and Sun estimated the parameter values $\overline{r} = -10$, $\sigma_p = 0$, $\rho = 0$, $\kappa_1 = 7.4525$, $\sigma_1 = 0.0197$, $\theta_1 + \overline{r} = 0.0264$, $\lambda = -0.0048$, $\kappa_2 = 0.0797$, $\sigma_2 = 0.1170$, $\theta_2 = 0.093$.

Prepayment and Calculation of Cash Flows

To value mortgage-backed securities, we need a model which specifies the cash flows each period as a function of the history of interest rates.[21] We use a model based on that of Schwartz and Torous (1989). Prepayment is governed by a hazard function π_t,[22] defined by

$$\pi_t = 0.75\exp[\beta v(t)].$$

Here, $v(t)$ is a vector of explanatory variables, defined by

$$\begin{aligned} v_1(t) &= c - l_t, \\ v_2(t) &= (c - l_t)^3, \\ v_3(t) &= \ln(\text{proportion of pool not yet prepaid}), \end{aligned}$$

[21] Or a model of the expected cash flows each period, as long as the risk of deviation from this expected value is not priced.

[22] In other words, as Δt approaches zero, the probability of prepayment occurring in a time interval of length Δt approaches $\pi_t \Delta t$.

where c is the coupon rate on the mortgage, and l_t is the yield on a long-term government bond. We assume a one-year bond, and use parameters based on those estimated by Stanton (1992), $\beta_1 = 0.49$, $\beta_2 = -0.01$, $\beta_3 = 0.15$.[23]

Given π_t, the expected cash flow in month t, per dollar of initial principal, is given by

$$C_t = \mathrm{SF}_{t-1}\left(X + (1 - e^{\pi_t/12})\mathrm{BAL}_{t-1}\right),$$

where X is the scheduled monthly payment, given by

$$X = \frac{c/12}{1 - (1 + c/12)^{-360}},$$

BAL_t is the scheduled balance remaining on the loan at the end of month t,

$$\mathrm{BAL}_t = \frac{X}{c/12}\left[1 - (1 + c/12)^{-(360-t)}\right],$$

and SF_t is the probability that the mortgage has not prepaid prior to t, given by $\mathrm{SF}_0 = 1$, and

$$\mathrm{SF}_t = (1 - e^{\pi_t/12})\,\mathrm{SF}_{t-1}.$$

Valuation

Assets whose value depends only on current values of r and y can be valued by writing down and solving a partial differential equation with appropriate boundary conditions (see Cox, Ingersoll and Ross (1985a,b)). This approach cannot easily handle path dependence of the sort we have described, where an asset's cash flows depend on the entire history of interest rates, rather than just the current values. An alternative approach is based on the fact that, given the interest rate model described above, we can write V, the value of an asset which pays out nominal cash flows at a (possibly path dependent) rate C_t, in the form

$$V_t = E\left[\int_t^T e^{-\int_t^s (\widehat{r}_u + y_u)\,du} C_s\,ds\right], \tag{8}$$

where $\widehat{r_\tau} = \widehat{r'_\tau} + \overline{r}$, and $\widehat{r'}$ follows the "risk adjusted" process,

$$\begin{aligned} d\widehat{r'}_\tau &= \left[\kappa_1(\theta_1 - \widehat{r'}_\tau) - \lambda \widehat{r'}_\tau\right] dt + \sigma_1 \sqrt{\widehat{r'}_\tau}\, dZ^1_\tau \qquad \text{for all } \tau \geq t, \\ \widehat{r'}_t &= r'_t. \end{aligned} \tag{9}$$

[23]To prevent the cubic term from dominating for extreme interest rates, $(c - l_t)$ is replaced by either 4.05% or −4.05% if its magnitude exceeds 4.05%.

This says that the value of the asset equals the expected sum of discounted cash flows paid over the life of the asset, except that it substitutes the risk adjusted process $\widehat{r'}$ for the true process r'_t for r.

This representation leads directly to a valuation algorithm based on Monte Carlo simulation. For each (r_t, y_t) pair (simulated using the model described in equations (6) and (7)), 500 paths for $\widehat{r}$ and y were simulated using equations (9) and (7). Along each path, the cash flows C_t were calculated as above, then discounted back along the path followed by the instantaneous nominal riskless rate $\widehat{r} + y$. The average of the sum of these values taken over all simulated paths is an approximation to the value V. The more paths simulated, the closer this approximation.

TABLE 1: SUMMARY STATISTICS

Table 1A – GNMA Prices

	Coupon						
	7.5%	8.0%	8.5%	9.0%	9.5%	10.0%	10.5%
Mean	93.132	95.578	97.876	100.084	102.204	104.347	106.331
Max.	105.156	106.563	107.500	108.281	109.469	110.938	112.719
Min.	78.375	81.625	83.656	86.531	89.531	92.688	95.750
Vol.	6.559	6.287	5.831	5.260	4.722	4.294	3.978

	Correlations						
	7.5%	8.0%	8.5%	9.0%	9.5%	10.0%	10.5%
7.5%	1.000	0.998	0.993	0.986	0.981	0.983	0.977
8.0%	0.998	1.000	0.997	0.992	0.987	0.987	0.979
8.5%	0.993	0.997	1.000	0.998	0.995	0.993	0.982
9.0%	0.986	0.992	0.998	1.000	0.999	0.995	0.983
9.5%	0.981	0.987	0.995	0.999	1.000	0.997	0.985
10.0%	0.983	0.987	0.993	0.995	0.997	1.000	0.994
10.5%	0.977	0.979	0.982	0.983	0.985	0.994	1.000

Table 1B – Interest Rates

	Long Rate	Spread	Curvature	Mortgage Rate
Mean	7.779	2.119	-0.324	9.337
Max.	10.230	3.840	0.430	11.580
Min.	5.170	-0.190	-1.065	6.740
Vol.	1.123	1.101	0.398	1.206

	Correlations			
	Long Rate	Spread	Curvature	Mortgage Rate
Long Rate	1.000	-0.450	0.733	0.980
Spread	-0.450	1.000	-0.753	-0.518
Curvature	0.733	-0.753	1.000	0.782
Mortgage Rate	0.980	-0.518	0.782	1.000

Summary statistics for prices of TBA contracts on 7.5% to 10.5% GNMAs, the long rate (10-year), the spread (10-year minus 3-month), the curvature(the average of the 5-year and 3-month minus the 1-year), and the average mortgage rate. All data are weekly from January 1987 through May 1994. Interest rates are in percent per year.

TABLE 2: 1-FACTOR GNMA PRICING

Table 2A – Pricing Errors

	Coupon						
	7.5%	8.0%	8.5%	9.0%	9.5%	10.0%	10.5%
Mean	0.003	0.006	0.007	0.010	0.010	0.010	0.009
Mean Abs.	0.529	0.605	0.649	0.679	0.660	0.597	0.666
Vol.	0.703	0.747	0.800	0.832	0.824	0.767	0.841
Autocorr.	0.861	0.898	0.918	0.927	0.921	0.917	0.916

Table 2B – Pricing Error Regression Analysis

	Coupon						
	7.5%	8.0%	8.5%	9.0%	9.5%	10.0%	10.5%
Const.	3.226	3.613	3.887	3.857	3.721	3.249	2.755
	(3.190)	(3.413)	(4.133)	(4.419)	(4.402)	(4.028)	(4.320)
Long Rate	-0.963	-1.062	-1.130	-1.117	-1.074	-0.942	-0.805
	(0.887)	(0.941)	(1.135)	(1.216)	(1.210)	(1.119)	(1.216)
$(\text{Long Rate})^2$	0.069	0.075	0.080	0.078	0.075	0.066	0.057
	(0.059)	(0.063)	(0.075)	(0.081)	(0.080)	(0.075)	(0.082)
R^2	0.028	0.026	0.023	0.020	0.018	0.017	0.011
Joint Test	2.795	2.157	1.709	1.441	1.327	1.228	0.707
p-value	0.247	0.340	0.426	0.487	0.515	0.541	0.702
AC(e)	0.853	0.891	0.913	0.923	0.916	0.912	0.914
Const.	0.491	0.236	0.411	0.607	0.887	1.137	1.494
	(0.148)	(0.194)	(0.212)	(0.211)	(0.194)	(0.165)	(0.135)
Spread	-0.673	-0.373	-0.446	-0.605	-0.948	-1.234	-1.582
	(0.275)	(0.330)	(0.365)	(0.374)	(0.342)	(0.286)	(0.261)
$(\text{Spread})^2$	0.165	0.098	0.095	0.120	0.199	0.261	0.328
	(0.074)	(0.090)	(0.101)	(0.103)	(0.094)	(0.079)	(0.072)
R^2	0.072	0.020	0.033	0.068	0.145	0.276	0.401
Joint Test	6.480	1.281	2.608	5.916	14.102	34.899	79.195
p-value	0.039	0.527	0.271	0.052	0.001	0.000	0.000
AC(e)	0.848	0.895	0.914	0.920	0.904	0.877	0.847

Summary statistics and regression analysis for the pricing errors from a 1-factor (long rate) MDE GNMA pricing model. The regression analysis involves regressing the pricing errors on linear and squared explanatory variables. Heteroscedasticity and autocorrelation consistent standard errors are reported in parentheses below the corresponding regression coefficient. AC(e) is the autocorrelation of the residuals from the regression.

TABLE 3: 2-FACTOR GNMA PRICING

Table 3A – Pricing Errors

	Coupon						
	7.5%	8.0%	8.5%	9.0%	9.5%	10.0%	10.5%
Mean	0.018	0.020	0.022	0.023	0.025	0.023	0.018
Mean Abs.	0.503	0.489	0.499	0.494	0.483	0.412	0.396
Vol.	0.646	0.616	0.627	0.623	0.613	0.532	0.523
Autocorr.	0.832	0.843	0.859	0.869	0.864	0.840	0.819

Table 3B – Pricing Error Regression Analysis

	Coupon						
	7.5%	8.0%	8.5%	9.0%	9.5%	10.0%	10.5%
Const.	3.470	3.924	4.193	4.188	4.088	3.559	3.002
	(2.751)	(2.784)	(3.112)	(3.064)	(2.924)	(2.228)	(1.875)
Long Rate	-1.036	-1.150	-1.215	-1.207	-1.176	-1.030	-0.878
	(0.763)	(0.754)	(0.830)	(0.814)	(0.780)	(0.605)	(0.526)
(Long Rate)^2	0.075	0.082	0.085	0.085	0.082	0.072	0.062
	(0.051)	(0.049)	(0.054)	(0.053)	(0.051)	(0.040)	(0.036)
R^2	0.041	0.045	0.043	0.040	0.039	0.042	0.036
Joint Test	4.775	5.846	5.185	5.059	4.608	5.332	3.792
p-value	0.092	0.054	0.075	0.080	0.100	0.070	0.150
AC(e)	0.825	0.834	0.852	0.865	0.856	0.833	0.814
Const.	0.124	0.096	0.082	0.093	0.117	0.151	0.210
	(0.200)	(0.182)	(0.175)	(0.171)	(0.178)	(0.171)	(0.151)
Spread	-0.203	-0.158	-0.105	-0.105	-0.126	-0.183	-0.308
	(0.279)	(0.265)	(0.265)	(0.255)	(0.243)	(0.210)	(0.177)
(Spread)^2	0.057	0.045	0.028	0.027	0.031	0.046	0.081
	(0.072)	(0.068)	(0.069)	(0.065)	(0.061)	(0.052)	(0.043)
R^2	0.009	0.006	0.002	0.002	0.003	0.009	0.027
Joint Test	0.665	0.486	0.172	0.173	0.270	0.777	3.474
p-value	0.717	0.784	0.917	0.917	0.874	0.678	0.176
AC(e)	0.830	0.841	0.857	0.868	0.862	0.836	0.809

Summary statistics and regression analysis for the pricing errors from a 2-factor (long rate, spread) MDE GNMA pricing model. The regression analysis involves regressing the pricing errors on linear and squared explanatory variables. Heteroscedasticity and autocorrelation consistent standard errors are reported in parentheses below the corresponding regression coefficient. AC(e) is the autocorrelation of the residuals from the regression.

TABLE 4: CORRELATION OF 2-FACTOR PRICING ERRORS

	Coupon						
	7.5%	8.0%	8.5%	9.0%	9.5%	10.0%	10.5%
7.5%	1.000	0.942	0.889	0.841	0.785	0.658	0.401
8.0%	0.942	1.000	0.968	0.938	0.889	0.774	0.495
8.5%	0.889	0.968	1.000	0.982	0.947	0.829	0.524
9.0%	0.841	0.938	0.982	1.000	0.977	0.871	0.566
9.5%	0.785	0.889	0.947	0.977	1.000	0.922	0.634
10.0%	0.658	0.774	0.829	0.871	0.922	1.000	0.836
10.5%	0.401	0.495	0.524	0.566	0.634	0.836	1.000

Correlations across coupons of the pricing errors from a 2-factor (long rate, spread) MDE GNMA pricing model.

TABLE 5: DETERMINANTS OF 2-FACTOR PRICING ERRORS

	Coupon						
	7.5%	8.0%	8.5%	9.0%	9.5%	10.0%	10.5%
Curvature R^2	0.113	0.120	0.111	0.110	0.107	0.107	0.129
Joint Test	12.055	13.258	11.488	12.513	11.888	8.775	10.981
p-value	0.002	0.001	0.003	0.002	0.003	0.012	0.004
AC(e)	0.811	0.822	0.841	0.854	0.849	0.825	0.796
CPR R^2	0.261	0.286	0.108	0.128	0.143	0.033	0.045
Joint Test	18.673	28.704	42.118	55.699	46.451	5.739	4.349
p-value	0.000	0.000	0.000	0.000	0.000	0.057	0.114
AC(e)	0.766	0.775	0.842	0.851	0.841	0.835	0.810
Age R^2	0.262	0.283	0.084	0.077	0.128	0.038	0.041
Joint Test	19.050	31.037	5.496	9.388	30.604	4.325	3.965
p-value	0.000	0.000	0.064	0.009	0.000	0.115	0.138
AC(e)	0.765	0.775	0.845	0.859	0.844	0.834	0.811
Burnout R^2	0.006	0.009	0.114	0.147	0.152	0.033	0.054
Joint Test	NA	NA	35.151	62.180	59.718	7.326	7.969
p-value	NA	NA	0.000	0.000	0.000	0.026	0.019
AC(e)	0.831	0.842	0.841	0.848	0.841	0.835	0.808

Further regression analysis for the pricing errors from a 2-factor (long rate, spread) MDE GNMA pricing model. The regression analysis involves regressing the pricing errors on explanatory variables. AC(e) is the autocorrelation of the residuals from the regression.

TABLE 6: 3-FACTOR GNMA PRICING

Table 6A – Pricing Errors

	Coupon						
	7.5%	8.0%	8.5%	9.0%	9.5%	10.0%	10.5%
Mean	0.019	0.020	0.022	0.021	0.022	0.019	0.010
Mean Abs.	0.389	0.353	0.360	0.355	0.364	0.354	0.338
Vol.	0.478	0.440	0.451	0.443	0.453	0.443	0.443
Autocorr.	0.697	0.701	0.739	0.749	0.757	0.766	0.738

	Correlations						
	7.5%	8.0%	8.5%	9.0%	9.5%	10.0%	10.5%
7.5%	1.000	0.897	0.807	0.722	0.628	0.501	0.377
8.0%	0.897	1.000	0.945	0.892	0.807	0.680	0.512
8.5%	0.807	0.945	1.000	0.968	0.904	0.779	0.590
9.0%	0.722	0.892	0.968	1.000	0.960	0.854	0.675
9.5%	0.628	0.807	0.904	0.960	1.000	0.933	0.783
10.0%	0.501	0.680	0.779	0.854	0.933	1.000	0.897
10.5%	0.377	0.512	0.590	0.675	0.783	0.897	1.000

Table 6B – Pricing Error Regression Analysis

	Coupon						
	7.5%	8.0%	8.5%	9.0%	9.5%	10.0%	10.5%
Long Rate R^2	0.058	0.069	0.066	0.064	0.057	0.049	0.043
Spread R^2	0.021	0.019	0.010	0.010	0.010	0.011	0.019
Curvature R^2	0.057	0.059	0.048	0.045	0.041	0.041	0.046
CPR R^2	0.107	0.107	0.023	0.035	0.036	0.011	0.007
Age R^2	0.105	0.103	0.039	0.018	0.026	0.009	0.007
Burnout R^2	0.012	0.016	0.027	0.044	0.038	0.012	0.007

Summary statistics and regression analysis for the pricing errors from a 3-factor (long rate, spread, curvature) MDE GNMA pricing model. The regression analysis involves regressing the pricing errors on linear and squared explanatory variables.

TABLE 7: CORRELATION OF CHANGES IN 3-FACTOR PRICING ERRORS

	Coupon						
	7.5%	8.0%	8.5%	9.0%	9.5%	10.0%	10.5%
7.5%	1.000	0.840	0.791	0.751	0.665	0.535	0.424
8.0%	0.840	1.000	0.905	0.871	0.788	0.647	0.522
8.5%	0.791	0.905	1.000	0.946	0.868	0.729	0.589
9.0%	0.751	0.871	0.946	1.000	0.917	0.776	0.641
9.5%	0.665	0.788	0.868	0.917	1.000	0.877	0.745
10.0%	0.535	0.647	0.729	0.776	0.877	1.000	0.830
10.5%	0.424	0.522	0.589	0.641	0.745	0.830	1.000

Correlations across coupons of the change in the pricing errors from a 3-factor (long rate, spread, curvature) MDE GNMA pricing model.

TABLE 8: MORE ON 2-FACTOR PRICING

Table 7A – Subperiod Analysis

	Pricing Errors					R^2			
	Mean	Mean Abs	Vol.	AC	Corr.	Long Rate	Spread	Curvature	CPR
1/87-3/90	0.262	0.457	0.521	0.753	0.798	0.034	0.079	0.019	0.093
2/91-5/94	-0.147	0.486	0.611	0.881	0.717	0.208	0.056	0.137	0.296
					Re-Estimated				
1/87-3/90	0.009	0.327	0.415	0.665	0.770	0.025	0.075	0.050	0.041
2/91-5/94	0.012	0.378	0.477	0.789	0.654	0.005	0.004	0.237	0.089

Table 7B – Estimation Error Analysis

	Pricing Errors				R^2			
	Mean	Mean Abs	Vol.	Corr.	Long Rate	Spread	Curvature	CPR
Dense	-0.079	0.527	0.665	0.772	0.003	0.007	0.101	0.149
Sparse	0.120	0.410	0.502	0.811	0.056	0.096	0.112	0.021

Averages across coupons of summary statistics and regression R^2s for pricing errors from a 2-factor (long rate, spread) MDE GNMA model.

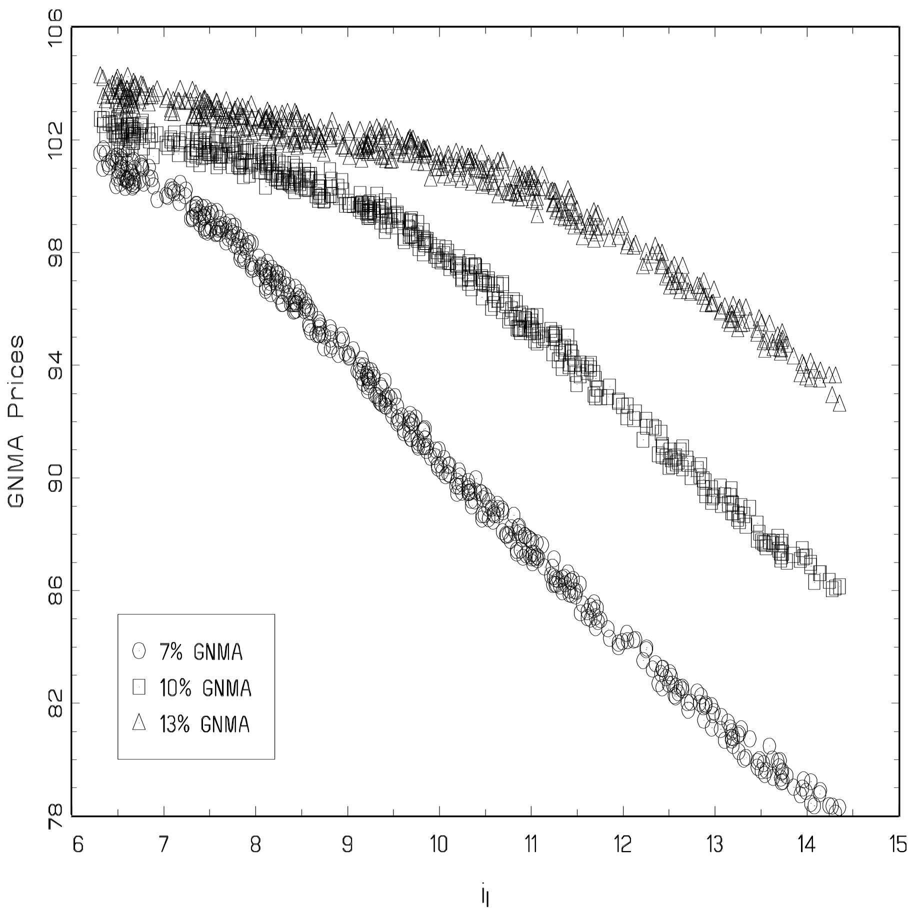

Figure 1: Scatter plot of simulated 7%, 10%, and 13% GNMA prices in a two factor economy. The model used is discussed in detail in the Appendix.

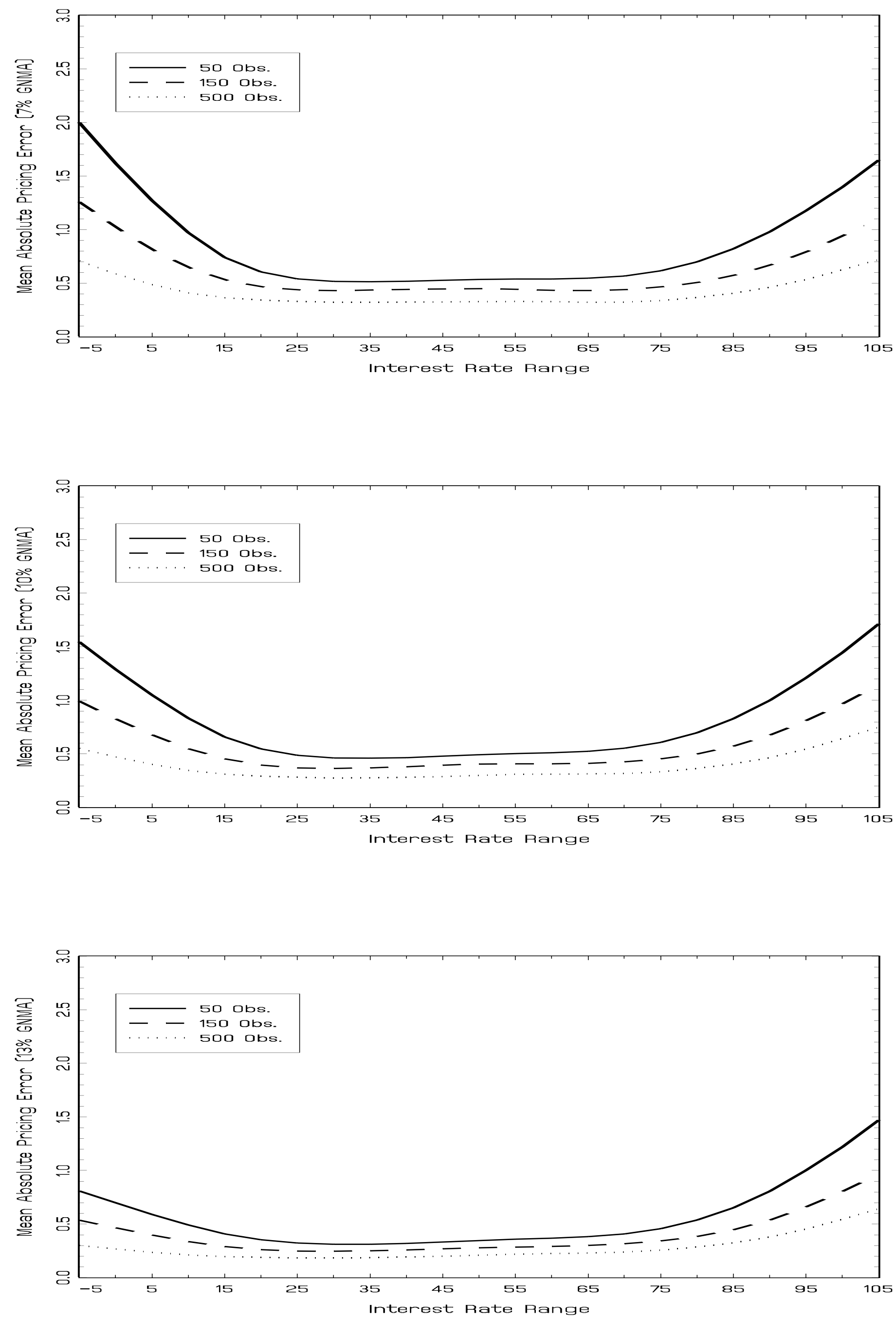

Figure 2: Average absolute pricing errors resulting from applying the MDE approach to the simulated price data for a 7%, 10%, and 13% GNMAs. Note that the x-axis measures the level of the interest rate relative to the maximum and minimum in the first 50 observations. The points “0” and “100” correspond to the minimum and maximum, respectively.

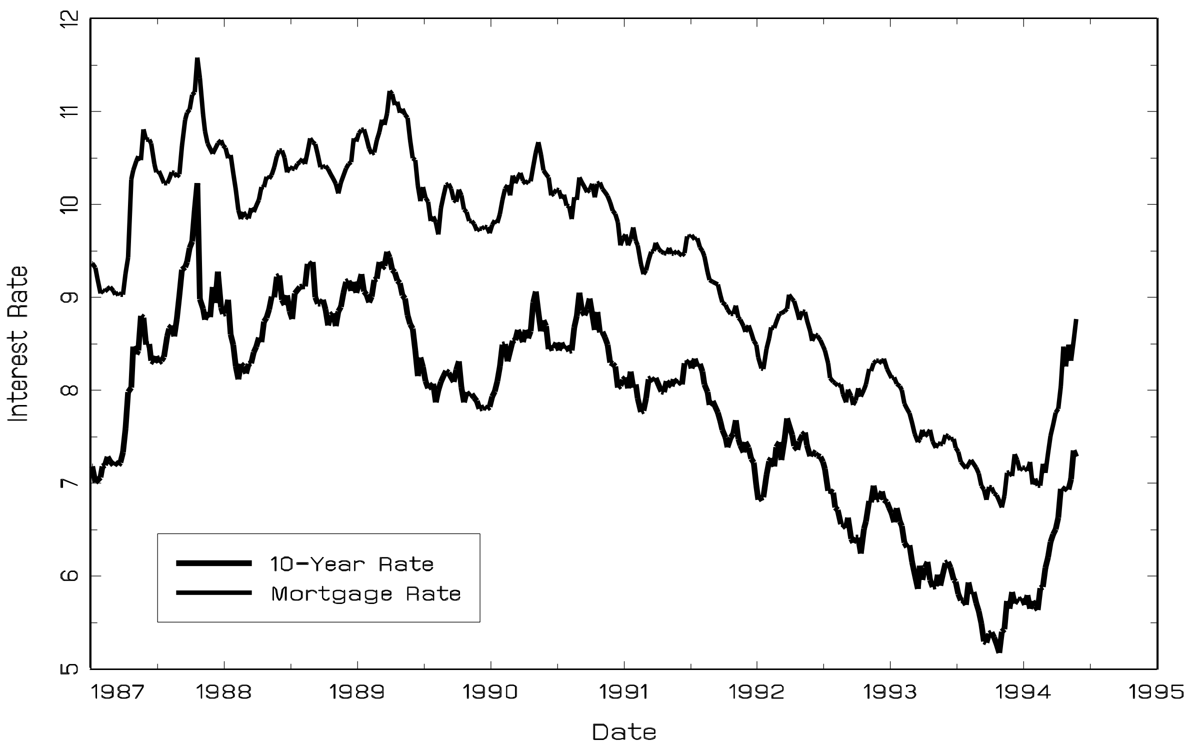

Figure 3: The yield on the "on-the-run" 10-year Treasury note and the average 30-year mortgage rate, from January 1987 to May 1994.

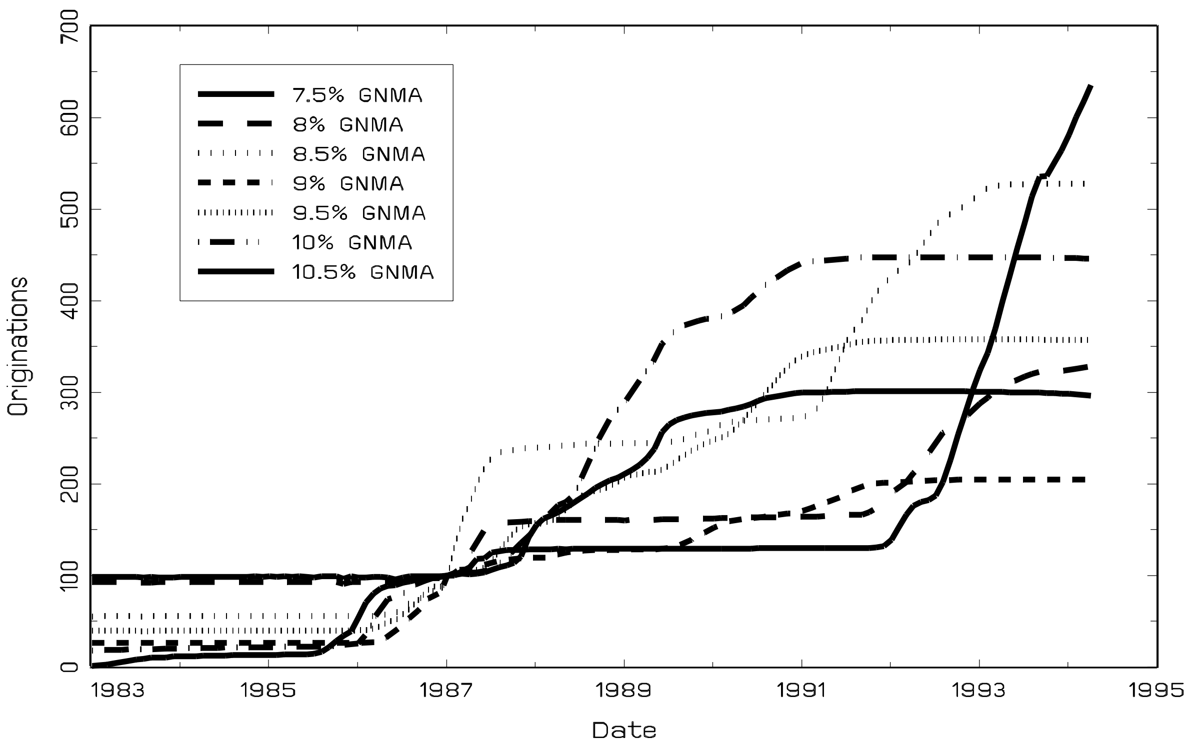

Figure 4: Originations of 7.5%–10.5% GNMAs from January 1983 to April 1994. The dollar amount outstanding is normalized to 100 in January 1987.

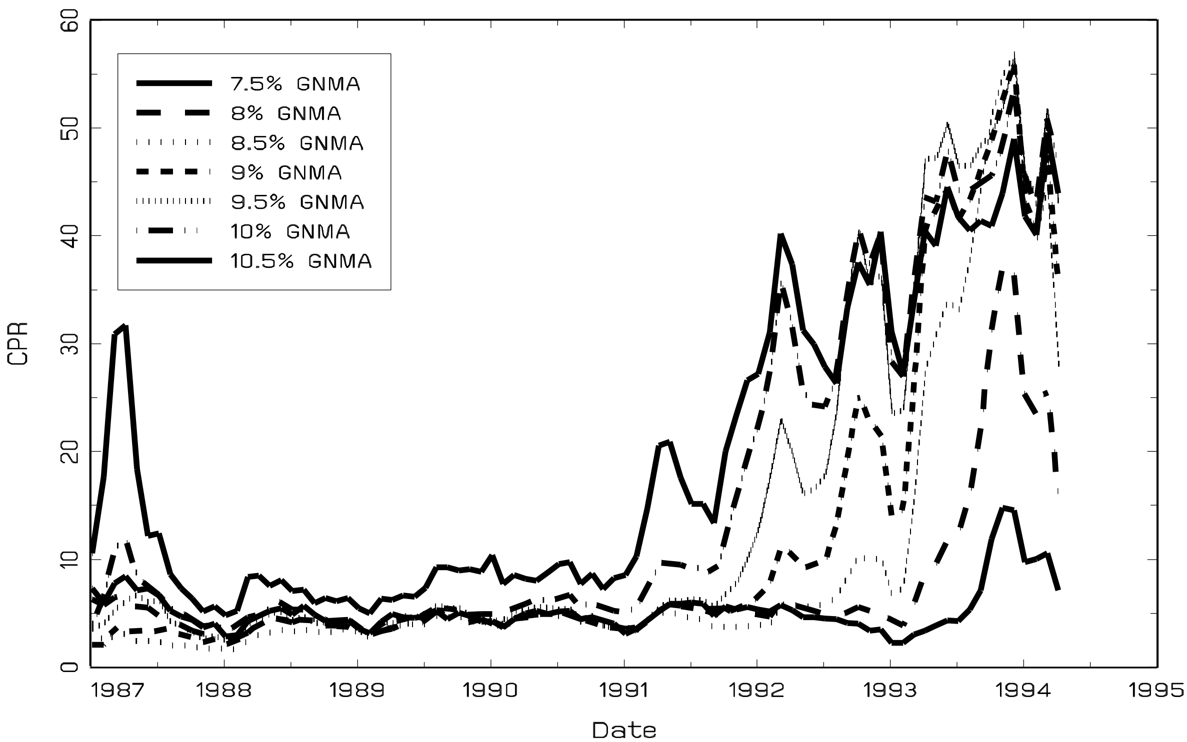

Figure 5: Annualized monthly conditional prepayment rates of 7.5%–10.5% GNMAs from January 1987 to April 1994, in percent.

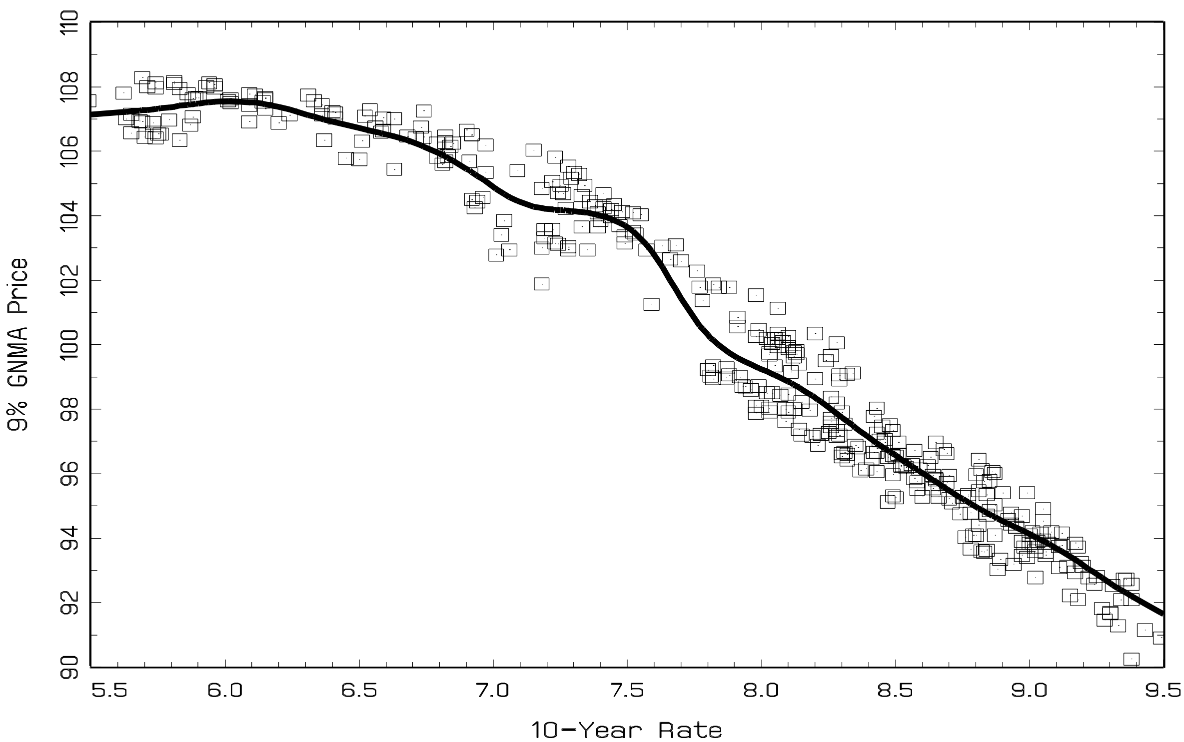

Figure 6: Observed weekly prices and estimated prices from a 1-factor (long rate) MDE model for a 9% GNMA for the period January 1987 to May 1994.

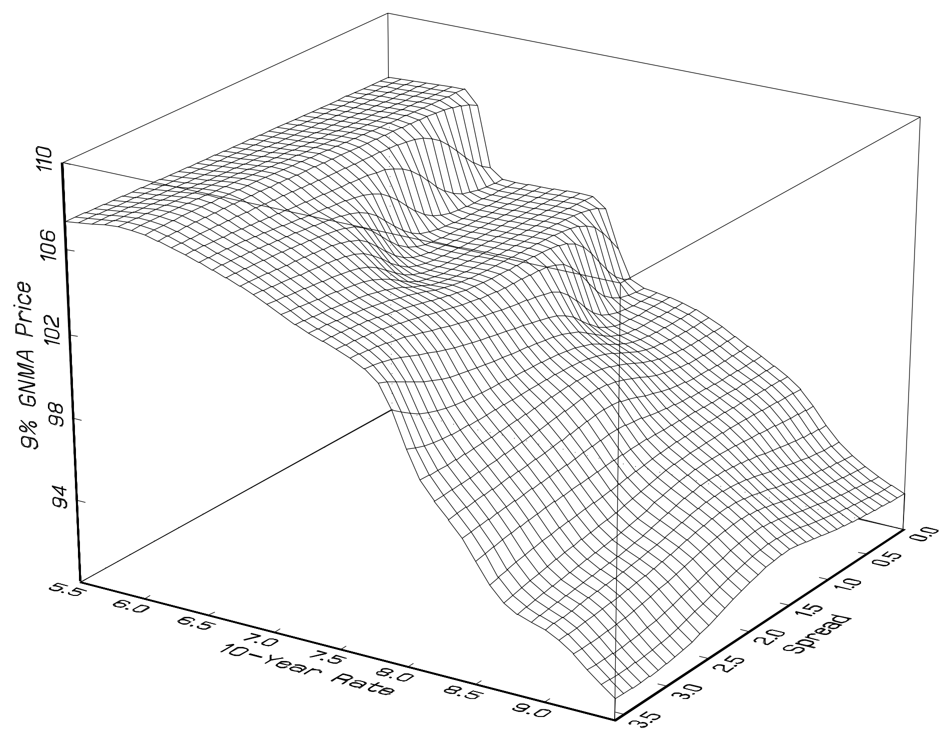

Figure 7: The price of a 9% GNMA as a function of the pricing factors: the long rate and the spread. The pricing functional is estimated using the MDE approach and weekly data from January 1987 to May 1994.

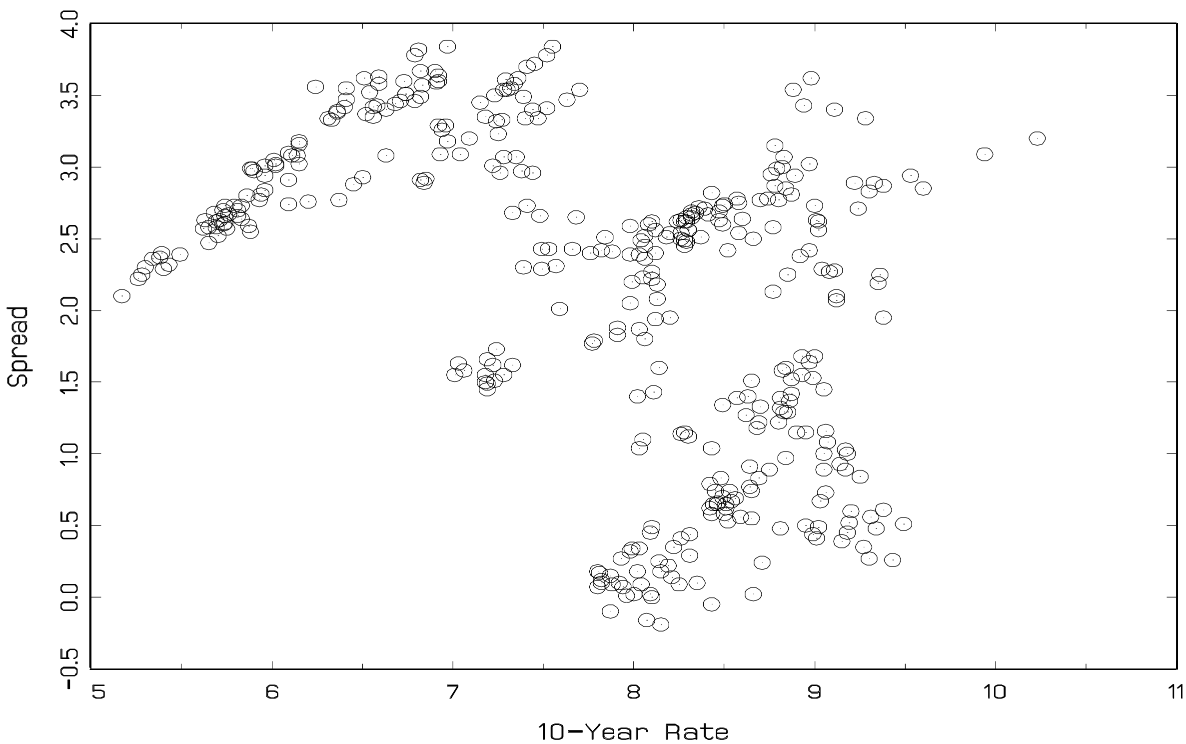

Figure 8: A scatter plot of the pairs of data available for the 10-year rate and the spread between the 10-year rate and the 3-month rate, from January 1987 to May 1994.

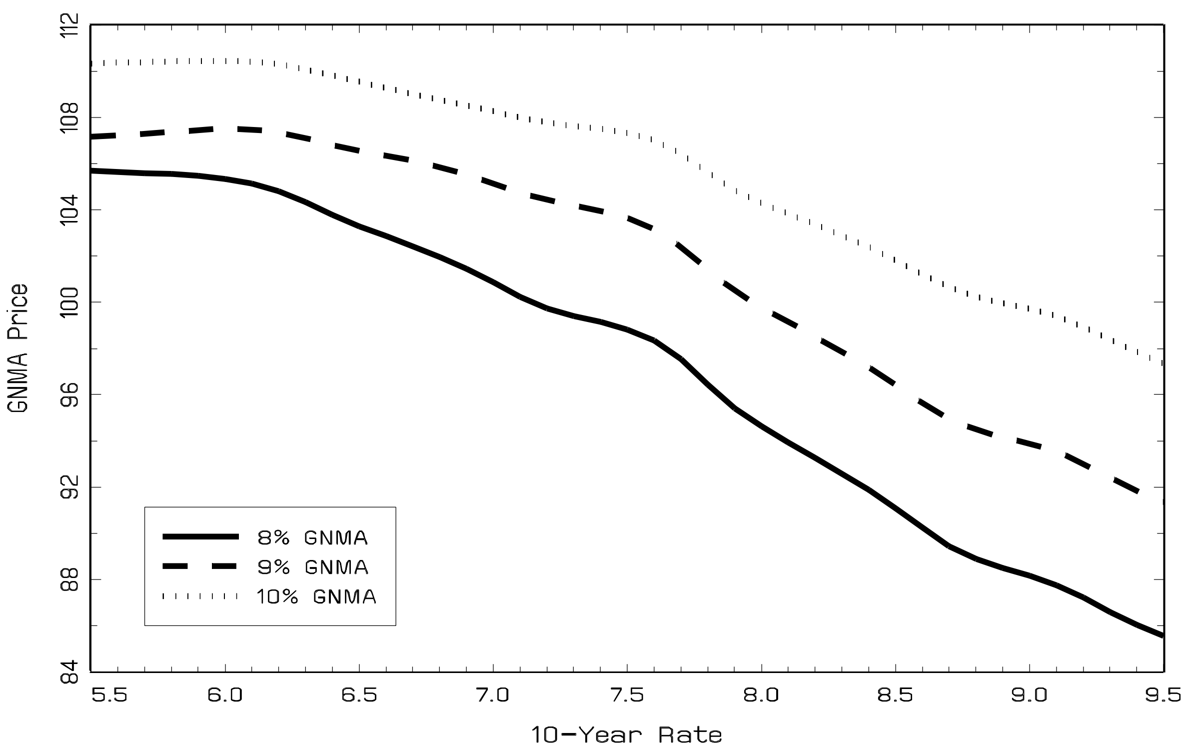

Figure 9: Prices of 8%, 9% and 10% GNMAs for various interest rates, with the spread fixed at 2.70%, as estimated via the MDE approach using weekly data from January 1987 to May 1994.

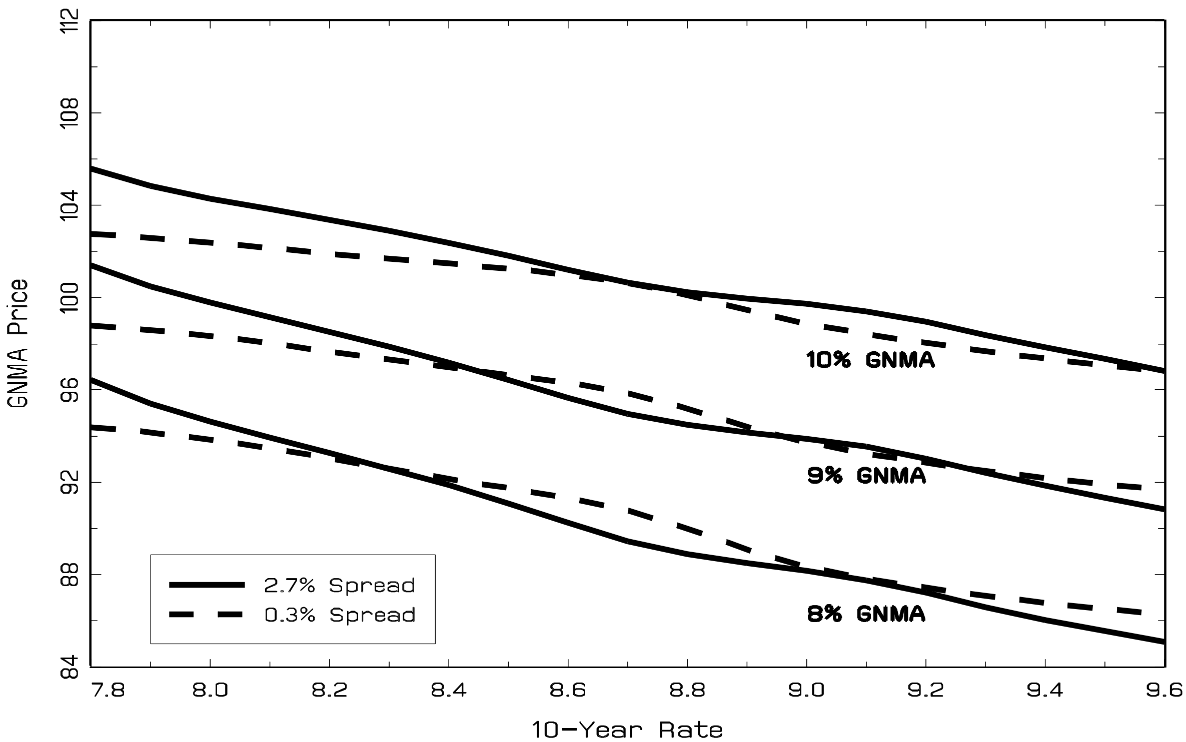

Figure 10: Prices of 8%, 9% and 10% GNMAs for various interest rates, with the spread fixed at 2.70% and 0.30%, as estimated via the MDE approach using weekly data from January 1987 to May 1994.

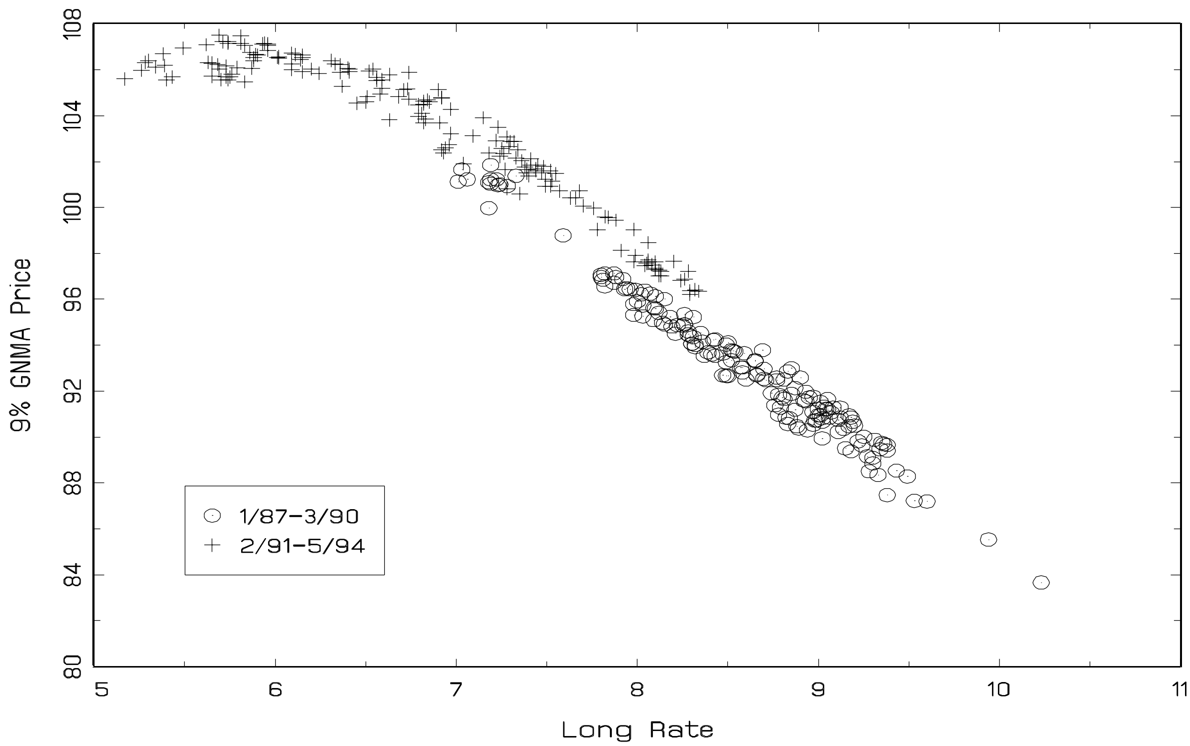

Figure 11: Observed weekly prices for a 9% GNMA for two distinct subperiods, 1/87-3/90 and 2/91-5/94.

References

[1] Aït-Sahalia, Y., 1996, "Nonparametric Pricing of Interest Rate Derivative Securities," *Econometrica*, 64, 527–560.

[2] Bartlett, W. W., 1989, *Mortgage-Backed Securities: Products, Analysis, Trading*, New York Institute of Finance.

[3] Boudoukh, J., M. Richardson, R. Stanton, and R. F. Whitelaw, 1995, "A New Strategy for Dynamically Hedging Mortgage-Backed Securities," *Journal of Derivatives*, 2, 60-77.

[4] Breeden, D. T., 1991, "Risk, Return and Hedging of Fixed Rate Mortgages," *Journal of Fixed Income*, 1, 85-107.

[5] Brennan, M. J., and E. S. Schwartz, 1985, "Determinants of GNMA Mortgage Prices," *AREUEA Journal*, 13, 209–228.

[6] Cox, J. C., J. E. Ingersoll, and S. A. Ross, 1985a, "An Intertemporal General Equilibrium Model of Asset Prices," *Econometrica*, 50, 363-384.

[7] Cox, J. C., J. E. Ingersoll, and S. A. Ross, 1985b, "A Theory of the Term Structure of Interest Rates," *Econometrica*, 53, 385-407.

[8] Dunn, K. B. and J. J. McConnell, 1981, Valuation of GNMA Mortgage-Backed Securities, *Journal of Finance*, 36, 599-616.

[9] Dunn, K. B., and C. S. Spatt, 1985, "An Analysis of Mortgage Contracting: Prepayment Penalties and the Due-On-Sale Clause," *Journal of Finance*, 40, 293-308.

[10] Epanechnikov, V., 1969, "Nonparametric Estimates of Multivariate Probability Density," *Theory of Probability and Applications*, 14, 153-158.

[11] Fabozzi, F., 1993, *The Handbook of Mortgage-Backed Securities,* Probus Publishing.

[12] Harvey, C., 1991, "The Specification of Conditional Expectations," working paper, Fuqua School of Business.

[13] Johnston, E., and L. Van Drunen, 1988, "Pricing Mortgage Pools with Heterogeneous Mortgages: Empirical Evidence," working paper, University of Utah.

[14] Litterman, R., and J. Scheinkman, 1991, "Common Factors Affecting Bond Returns," *Journal of Fixed Income*, 1, 54-61.

[15] Pagan, A., and Y.S. Hong, 1991, "Non-parametric Estimation and the Risk Premium," in W. Barnett, J. Powell and G. Tauchen (eds), *Semiparametric and Nonparametric Methods in Econometrics and Statistics*, Cambridge University Press, Cambridge.

[16] Pearson, N. D. and T.-S. Sun, 1994, "Exploiting the Conditional Density in Estimating the Term Structure: An Application to the Cox, Ingersoll, and Ross Model," *Journal of Finance*, 49, 1279-1304.

[17] Richard, S. F., and R. Roll, 1989, "Prepayments on Fixed-Rate Mortgage-Backed Securities," *Journal of Portfolio Management*, 15, 74-82.

[18] Schwartz, E. S., and W. N. Torous, 1989, "Prepayment and the Valuation of Mortgage-Backed Securities," *Journal of Finance*, 44, 375-392.

[19] Scott, D. W., 1992, *Multivariate Density Estimation: Theory, Practice and Visualization,* John Wiley & Sons, Inc.

[20] Stambaugh, R., 1988, "The Information in Forward Rates: Implications for Models of the Term Structure," *Journal of Financial Economics*, 21, 41-70.

[21] Stanton, R. H., 1992, "Essays in Financial Economics: Contingent Claims Analysis and Mortgage Prepayment," unpublished dissertation, Stanford University, Graduate School of Business.

[22] Stanton, R. H., 1995, "Rational Prepayment and the Valuation of Mortgage-Backed Securities," *Review of Financial Studies*, 8, 677-708.

[23] Timmis, G. C., 1985, "Valuation of GNMA Mortgage-Backed Securities with Transaction Costs, Heterogeneous Households and Endogenously Generated Prepayment Rates," working paper, Carnegie Mellon University.

[24] Waldman, M., 1992, "Beyond Duration: Risk Dimensions of Mortgage Securities," *Journal of Fixed Income*, 2, 5-15.