Mortgage Default and Mortgage Valuation*

John Krainer
Federal Reserve Bank of San Francisco

Stephen F. LeRoy
Federal Reserve Bank of San Francisco

Munpyung O
University of California, Santa Barbara

September 22, 2009

Abstract

We study optimal exercise by mortgage borrowers of the option to default. Also, we use
an equilibrium valuation model incorporating default to show how mortgage yields and lender
recovery rates on defaulted mortgages depend on initial loan-to-value ratios when borrowers
default optimally. The analysis treats both the frictionless case and the case in which borrowers
and/or lenders incur deadweight costs upon default. The model is calibrated using data on
California mortgages. We find that the model’s principal testable implication for default and
mortgage pricing—that default rates and yield spreads will be higher for high loan-to-value
mortgages—is borne out empirically.

*We are indebted to Fred Furlong, John Williams, Paul Willen, and seminar participants at UCLA, UCSB, the
China Economics and Management Academy, the Federal Reserve Bank of San Francisco, the Federal Reserve Board,
the University of Adelaide, the University of Melbourne and Victoria University (Wellington), and the SED Conference
in Istanbul, for comments. We also presented the paper at the Federal Reserve System Conference on Housing and
Mortgage Markets. The views expressed are those of the authors and not necessarily those of the Federal Reserve
Bank of San Francisco or the Board of Governors of the Federal Reserve System.
It is generally agreed that the mortgage valuation models in use on Wall Street underpredicted the default frequencies observed over the past several years by a wide margin. There exist two possible explanations. The first is that the models are correct, but the values taken on by exogenous variables lay in the extreme tails of their assumed distributions. In this story, practitioners were simply unlucky; even a forecast generated by a correct model will be wide of the mark if the shock—in this case, a shock to house prices—is two or three standard deviations from its mean. The second possibility is that the valuation models did not function correctly, either because they were not calibrated properly or because they somehow failed to capture the tradeoffs faced by actual participants (borrowers and lenders) in this market. In this paper we develop a structural model of mortgage default in the attempt to shed light on this question. In our model, lenders condition explicitly on the borrower’s assumed optimal default, so ex ante loan pricing and expected default are tightly linked. ¹ The main contribution is an equilibrium characterization of the way mortgage loan yields adjust to compensate for differences in borrower leverage. We also consider the way this pricing schedule is altered by the presence of borrower and lender default costs. We then calibrate the model and compare the predicted relationship between loan yield and loan-to-value to the empirical relationship observed in the data.

Mortgage valuation models used by practitioners are generally “reduced-form” models—models that correct for changes in the environment using regression-based methods. Such models are unlikely to give accurate predictions in the presence of the pronounced changes in real estate values and lending practices that we have seen in the past several years.

“Structural” mortgage valuation models specify a setting in which the financial consequences to a borrower of defaulting or not defaulting on a mortgage are explicitly modeled. They incorporate the assumption that the borrower defaults when it is optimal to do so, and only then. We believe that in the current environment structural models are likely to produce more accurate default predictions than reduced-form models, given the dramatic changes that we have experienced. In this paper we restrict our attention to structural models.

Many observers of recent financial events take it as obvious that mortgages were egregiously

¹Other authors that have studied related models include Deng, Quigley, and VanOrder (2000), Downing, Stanton, and Wallace (2005), Kau and Keenan (1999), and Vandell (1995).
mispriced during the real estate bubble period. Certainly with hindsight this is true, but we find that the evidence that it is also true ex ante is surprisingly thin. Given our modeling procedure, our failure to find convincing evidence of ex-ante mispricing occurs almost by construction: our model of default and mortgage pricing is richly parametrized, and does not produce much in the way of over-identifying restrictions. As a consequence, we can calibrate the model to real-world data with reasonable success, but the same would be true under a wide variety of possible data sets. Therefore we cannot interpret successful calibration as implying the validity of the model.

It is true that there is some weak evidence of ex ante mispricing of mortgages. For example, from our calibration exercise it appears as if mortgage pricing reflects subjective estimates of house price volatility that are lower than we see in the data. We do not make too much out of this, however, because of the difficulty of measuring volatility. Our major conclusion is that for the most part we do not see sharp contrasts between the real-world data and the predictions from theory. This outcome raises strong doubts about the claims we hear of gross mispricing in mortgage markets. What is the evidence that such mispricing has occurred? It is, of course, easy to produce oversimplified theoretical models that generate empirical predictions that conflict with the data (for example, we will see below that real-world mortgage pricing is inconsistent with optimal default and efficient pricing of default in the absence of default costs). This inconsistency, of course, should be interpreted as rejecting the oversimplification (in this case, the oversimplifying assumption that default costs are zero), not as implying that mortgages are mispriced.

1 Default and Mortgage Valuation

We study the behavior of borrowers and lenders in a stylized housing market where borrowers must decide how to finance their housing purchases. To make the analysis tractable, we make three simplifying assumptions. First, we assume that mortgages are perpetuities that have no scheduled payment of principal. Thus, all changes in homeowner equity are due to changes in house prices. Since almost all the early payments on 30-year mortgages consist of interest, this specification does not involve a major distortion. Second, we assume that the borrower can always terminate the mortgage by paying the lender its fair market value, even when that value is less than book value
(as will occur when house prices drop). As we will discuss below, this specification eliminates the possibility of optimal defaults triggered by adverse life events that affect the borrower’s ability to repay the mortgage. Finally, we assume that interest rates are deterministic. Obviously, mortgage prepayment speeds depend on interest rates and, hence, so too will the level of mortgage rates depend on the level of interest rates. Our focus in this paper is on the difference in mortgage rates for loans with different loan-to-value ratios (LTVs).

We also consider the effect of giving the borrower the option to prepay the loan at its book value subject to a prepayment penalty. Existence of this option affects mortgage pricing even though we abstract away from the interest rate fluctuations that are usually taken as the major reason for prepayment. In our model prepayment occurs because increases in house values decrease the value of the default option, implying that borrowers whose homes have appreciated are overpaying for the default option. Borrowers are motivated to reduce their payment for the default option to its economic value by refinancing. Like Downing, Stanton, and Wallace (2005), we find that the existence of the default option affects the value of the prepayment option, and vice versa.

We will take house prices to be exogenous, thereby shutting down any link between mortgage underwriting practices and house prices. This is a major shortcoming: it appears likely that increasingly lax underwriting standards were a major contributor to the price runup that occurred in the decade prior to 2006. Determining the impact of changes in mortgage underwriting standards on house prices is an important research topic, but one best left for further work.

1.1 The Model Setting

Our model, like all existing structural mortgage valuation models that incorporate default, is based on Merton’s classic paper (Merton (1974)), adapted to the housing market. We assume that housing services $x$ follow a geometric Brownian motion:

$$dx = \alpha xd t + \sigma xd z,$$  \hspace{1cm} (1.1)

where $\alpha$ is the average proportional increase over time in the level of housing services, $\sigma$ is the associated volatility and $z$ is a standard Brownian motion. As applied to investment property,
$x$ can be interpreted as the difference between rental income and costs to the owner of utilities, repairs, and so on.

It is assumed that the price of housing equals the expectation of future housing services discounted at rate $\rho$:

$$P(x(t)) = \int_{\tau=t}^{\infty} e^{-\rho(\tau-t)} \mathcal{E}_t(x(\tau)) d\tau = \frac{x(t)}{\rho - \alpha}. \quad (1.2)$$

Here $\mathcal{E}_t$ is the conditional expectations operator. As a scale choice, we will set $x(0)$, the value of housing services at the date the mortgage is initiated, equal to 1 henceforth. Thus mortgages are initiated when the price of housing equals $1/(\rho - \alpha)$. The valuation model (1.2) can be justified either by interpreting the probability measure in eq. (1.1) as the risk-neutral measure or more simply by directly assuming that agents are risk neutral. The latter assumption is acceptable here because we are not interested in the effects of risk aversion on mortgage valuation. As noted, mortgages are assumed to be perpetuities (conditional on nonoccurrence of default), and to be indexed by the level of the periodic payment $c$, which we take as exogenous.

The cash supplied by the lender toward the purchase equals the present value of the mortgage payments, making appropriate allowance for default as described below. Thus low (high) LTV mortgages can be generated by setting $c$ at a low (high) level. The borrower is assumed to supply the remainder of the purchase price as a down payment.

We will see that, depending on the evolution of $x$, the fair values of mortgages may fall short of their book values or exceed them. In the latter case borrowers are motivated to pay off the mortgages. For the present it is assumed that they cannot do so.

1.2 Costless Default

As noted, it is assumed in this subsection that default is costless to both the borrower and the lender. Thus the borrower is assumed to be able to terminate the mortgage at any time by turning over the house to the lender. He incurs no costs due, for example, to reduced access to credit or moving costs. Correspondingly, the lender suffers no loss of value when taking over a foreclosed property.
Borrowers seek to maximize the value of their home equity, \( E(x) \), or, equivalently, to minimize the value of their mortgage liability, \( M(x) \). Using dynamic programming it can be shown that maximizing the value of equity implies that \( E(x) \) satisfies the stochastic differential equation

\[
\rho E(x) dt = \mathcal{E}_x [dE(x)] + (x - c) dt,
\]

(1.3)

where \( dt \) is the increment in time. Thus the discounted gain on \( E(x) \) is a martingale, as implied by absence of arbitrage. Here \( x - c \) is the analogue of dividends in a stock valuation expression, except that \( x - c \) can be either positive or negative, and \( dE(x) \) is the capital gain. Using Ito’s lemma, (1.3) leads to the (deterministic) differential equation

\[
\frac{1}{2} x^2 \sigma^2 E''(x) + \alpha x E'(x) - \rho E(x) = -x + c.
\]

(1.4)

This differential equation is similar to those arising in finance and also in the real options literature (see Dixit and Pindyck (1994), Ch. 7 for detailed and very clear exposition of a model that contains the model here as a special case). The boundary conditions are the value-matching condition

\[
E(x_*) = 0
\]

(1.5)

and the smooth pasting condition

\[
E'(x_*) = 0.
\]

(1.6)

The smooth pasting condition is a necessary condition for optimal exercise of the default option since default is a free boundary. It requires that \( M(x) \) be tangent to \( x/(\rho - \alpha) \) at \( x_* \).

The value of the mortgage equals the value of the house less the borrower’s equity. Thus the value-matching condition implies that the value of the mortgage at \( x_* \) equals the value of the house, \( x_*/(\rho - \alpha) \), and the smooth pasting condition implies that the slope of the mortgage value as a function of \( x \) is \( 1/(\rho - \alpha) \) at \( x = x_* \). These boundary conditions allow solution for the constant of integration \( e \).
For \( x > x^* \), the solution to the differential equation is

\[
E(x) = e x^m + \frac{x}{\rho - \alpha} - \frac{c}{\rho},
\]

where \( e \) is the constant of integration:

\[
e = \frac{1}{m(\alpha - \rho)} \left[ \frac{c}{\rho} \left( \frac{m(\rho - \alpha)}{m - 1} \right) \right]^{1-m}
\]

and \( m \) is the negative root associated with the differential equation (1.4):

\[
m = -\frac{(\alpha - \frac{1}{2} \sigma^2) - \sqrt{(\alpha - \frac{1}{2} \sigma^2)^2 + 2 \sigma^2 \rho}}{\sigma^2}
\]

(the constant of integration associated with the positive root equals zero).\(^2\)

The value of the mortgage is given by

\[
M(x) = P(x) - E(x) = \frac{c}{\rho} - e x^m,
\]

from

\[
P(x) = \frac{x}{\rho - \alpha}
\]

and eq. (1.7). Because \( m \) is negative, when \( x \) is high the mortgage value approaches \( c/\rho \) asymptotically, agreeing with the value it would have in the absence of default. This makes sense: when \( x \) is extremely high default in the foreseeable future is improbable, so it does not materially affect valuation of the mortgage. However, for values of \( x \) that are not high, \( M(x) \) depends positively on \( x \). It follows that the intuition supplied above—to the effect that when default is costless existence of default should not affect the value of the mortgage—is incorrect.

Having calculated \( M(x) \), one can compute the borrower’s down payment \( E(1) \) as \( P(1) - M(1) \), the initial LTV ratio \( M(1)/P(1) \), the initial mortgage yield \( c/M(1) \) and the recovery rate

\(^2\)For \( x < x^* \), \( E(x) \) does not follow the solution (1.7) to the differential equation (1.4). The simplest way to see this is to observe that for \( x < x^* \), immediate default is optimal. That being so, the value of the mortgage equals the value of the house, implying that \( E(x) = P(x) - M(x) = 0 \).
1.3 A Numerical Example

We now provide a numerical example of the equity values and mortgage values as a function of house services that are implied by this model. Our calibration assumes that \( \rho = .07, \alpha = .03, \sigma = 0.1, \) so that the average (real) proportional gain on both the equity and the mortgage is 7\%, and the initial price-rent ratio is \( 1/(\rho - \alpha) = 25 \) (real-world price-rent ratios are closer to 10 or 15 than 25, but recall that here we are abstracting away from operating costs such as repairs and utilities expense). The standard deviation of the change in housing services, \( \sigma, \) is taken to be 10\%.

Under these parameter values a security similar to a prime mortgage can be generated by specifying that the mortgage payment \( c \) is 1.25. We term this mortgage the “low-LTV” mortgage. This is the specification used to compute \( E(x) \) shown in Figure 1. Even though the initial LTV is 71.1\%, default does not occur until the house value drops to 62.5\% of its original value. At that price the recovery rate (ratio of collateral value at the default point to the origination value of the mortgage) is 87.9\%. The yield on the mortgage at its origination date, equal to \( c/M(1) \), is 7.03\%. The difference between this rate and the expected return on the mortgage, 7\%, represents compensation for the possibility that the collateral value will drop, possibly leading to default.
A high-LTV mortgage results if \( c \) is increased from 1.25 to 1.75. This increase in the periodic payment results in an increase in the initiation value of the mortgage; the initial LTV ratio is 95.1%. Default occurs if the house value drops to 87.5% of its original value. Thus, as would be expected, the magnitude of the price drop that is just sufficient to induce default is much smaller under the high-LTV mortgage than under the low-LTV mortgage. Under the low-LTV mortgage the down payment protects the lender from all but the most drastic price drops, whereas under a high-LTV mortgage the lender has much less protection. Correspondingly, the recovery rate, 92.0%, is higher under the high-LTV mortgage than under the low-LTV mortgage.

The fact that default is likelier under the high-LTV mortgage than under the low-LTV mortgage implies that a higher yield, 7.4%, is needed to compensate the lender for the higher default risk. It is seen that the higher default probability under the high-LTV mortgage relative to the low-LTV mortgage more than outweighs the fact that the recovery rate is higher under the high-LTV mortgage than under the low-LTV mortgage.

The example demonstrates an underappreciated feature of the option-based default models: optimal default does not occur until the borrower is substantially “underwater” on the mortgage. The reason lies in the distinction between the economic value and the book value of the mortgage. A default calculation based on book value would suggest that borrowers should default when house price falls by the amount of the borrower’s equity at the origination date (falls 71.1% in this example). This decision rule is also depicted in Figure 1, labeled “\( E(x) \) incorrect.” To see that this decision rule is not optimal, assume that price drops to exactly 71.1%. Note that if prices were to fall below 71.1%, the borrower could default instantly and avoid any further costs. Thus the borrower would strictly prefer not to default. In contrast, if house prices were to rise, then the borrower would instantly have strictly positive home equity. For a default point to be optimal, the borrower must be indifferent between defaulting and staying in the mortgage at the default point. In our setup, this condition is met by imposing the smooth pasting condition on the solution, which guarantees that the asset value (the house value) and the mortgage value are not just equal, but also tangent to each other.

To gain further insight into the characteristics of the model solution, we show how mortgage
yields vary with the different assumptions for the value of house price volatility. The specified volatility of house prices of 10% used in the example may be too low. We chose it to provide a conservative estimate of the effect that theory leads us to expect of variations in LTV ratios on mortgage yields. Further discussion is provided below. In order to get an idea of how the volatility affects equity values, we computed $E(x)$ against $x$ for various values of volatility that bracket 0.1. The computed curves are shown in Figure 1.3. Standard options analysis applies here: the higher the volatility of house price changes, the greater is the value of the default option for given $x$, as reflected in $E(x)$. Also, higher volatilities are associated with lower default points.

Figure 2: Market Value of Homeowner Equity as Function of Housing Services, House Price Volatility

<table>
<thead>
<tr>
<th>$\sigma$</th>
<th>$E(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.06</td>
<td>0.487</td>
</tr>
<tr>
<td>0.10</td>
<td>0.568</td>
</tr>
<tr>
<td>0.14</td>
<td>0.625</td>
</tr>
<tr>
<td>0.20</td>
<td>0.674</td>
</tr>
</tbody>
</table>

To get a different view of the effect of different levels of price volatility on mortgage default and mortgage pricing, we also computed the tradeoff between mortgage yield and LTV for different values of volatility. These tradeoffs were obtained by varying the periodic payment $c$, and are shown in Figure 3.

1.4 Involuntary Default

The model as specified so far implies that default is always voluntary: agents will default when $x < x_*$, and they can and will avoid default when $x > x_*$. This conclusion reflects the assumption
Figure 3: Mortgage Yield as Function of LTV

\[
\sigma = 0.06 \\
\sigma = 0.10 \\
\sigma = 0.14 \\
\sigma = 0.20
\]
that when housing services equal \( x \) lenders will be willing to accept the market value of the mortgage, \( M(x) \), as payment for extinguishing the mortgage. Lenders are willing to do so even though when \( x < 1 \), \( M(x) \) is less than the book value of the mortgage because of the increased risk of default when \( x < 1 \) relative to when \( x = 1 \). Short sales, under which the lender voluntarily accepts repayment of the mortgage at less than its book value, reflect the fact that \( M(x) \) is less than the book value of the mortgage when \( x < 1 \). Thus in the model as specified life events such as job loss or medical expenses do not affect optimal default.

Contrary to our specification, in many cases lenders will accept mortgage repayment only at full book value even when full book value exceeds market value. This is so because, first, the borrower cannot generally credibly threaten to default if the short sale request is denied. Second, most mortgages are securitized, in which case the servicer often does not have authority to restructure mortgages to avoid default. In any case he almost always lacks incentive to do so. In such cases if the book value of equity is negative the borrower who cannot continue to pay the mortgage may be unable to refinance or sell profitably even when the market value of the equity is positive. In such settings the assertion above that life events like unemployment do not affect mortgage defaults no longer applies.

1.5 Second-Lien Mortgages

In practice, borrowers who need a high LTV mortgage often obtain two mortgages: a first-lien mortgage, usually at 80%, and a second-lien mortgage for the difference, which is typically on the order of 10% or 15%. The first-lien mortgage is senior to the second-lien mortgage, meaning that in the event of default the recovery is applied to it first. There are two institutional reasons for dividing up the mortgage in this way. First, until recently the government agencies declined to insure high-LTV mortgages, rendering them difficult to securitize. Accordingly, dividing mortgages into two pieces made it possible to securitize the senior mortgage. Second, many lenders insisted on mortgage insurance when the LTV exceeded 80%, so dividing the mortgage into two securities allowed borrowers to avoid the insurance expense on the senior security.

The model set forth in this paper can easily accommodate dividing a mortgage into two or
more pieces. It is reasonable to assume that the borrower’s default point is that implied by the LTV associated with the sum of the values of the two mortgages. The question now is how the decomposition of the recovery into payments to the senior and junior debt-holders affects their pricing. Surprisingly, in plausible cases the model set out here reduces default risk for the first lien-holder even relative to the case in which there is no second lien. In such cases the existence of a second lien would be expected to reduce the equilibrium yield on the first lien.

To see this, recall the “low-LTV” (71.1%) and “high-LTV” (95.1%) mortgages discussed above. Assume that the high-LTV mortgage is decomposed into the low-LTV mortgage and a second lien with an initial value of 24.0% of the purchase price. The default point of the high-LTV mortgage is $x_\ast = 87.5\%$, implying that the division of that mortgage into two securities results in the senior mortgage having a 100% recovery rate in the event of default. In contrast, in the absence of the second lien, default would occur at $x_\ast = 62.5\%$, implying a partial loss for the first lienholder. Thus existence of the second lien ensures default that is timely (from the point of view of the first lien-holder). As a result the first lien would be priced as if it were risk-free in the presence of the second lien, but not otherwise.

2 Costly Default and Prepayment

2.1 Costly Borrower Default

In the model of the preceding section the default option is a device that allows sharing between the owner and the lender of the risk of house price decreases. Without a mortgage the owner obviously bears 100% of the risk, whereas when a lender supplies part of the purchase price, the fact that the value of the mortgage drops as house value drops implies that the lender provides partial insurance to the borrower against the risk of loss of value. The proportions in which the loss is shared between the borrower and lender depend on the volatility of house price changes and the terms of the mortgage. The associated transfer of risk entails no deadweight loss, due to the assumption of zero default costs. Accordingly, high-LTV loans imply no inefficiency.

To the extent that the preceding analysis, including the assumption of zero default costs, applies
to the US crisis of the last several years, the interpretation is that mortgage markets exhibited
nothing that can be called an institutional failure. Along these lines, the story incorporated in
the model just presented is that house values dropped several trillion dollars, but these costs were
unaffected by the extent of mortgage default. On this interpretation the use of subprime mortgages
did not contribute to the magnitude of the loss, and the losses associated with default were a pure
transfer involving no deadweight loss.

There is now general agreement in the literature that it is not possible to explain the empirical
patterns of mortgage defaults while maintaining the assumption of costless default (e.g., Deng,
Quigley, and VanOrder (2000)). Indeed, it appears that some subprime loans were granted that
had negative equity at the date of origination, and this is so under either the book or market
definition of equity. Existence of such mortgages is possible only in the presence of costs of default
that induce borrowers to be willing to make payments on the loan even when their equity is negative.
In the absence of such costs borrowers would default immediately upon closing a negative-equity
mortgage. Lenders, anticipating this, would never grant negative-equity mortgages. Thus, the
existence of mortgage loans with negative economic equity is evidence in favor of positive default
costs.

Default costs to the borrower include the expense of moving necessitated by eviction, and
also the costs associated with impaired access to credit markets in the future. Costs to the lender
include expenses incurred repairing, maintaining and reselling foreclosed properties, and also include
foregone income during the period (which may be a year or more) between when the borrower
terminates mortgage payments and when the lender can gain possession of the property. 3

To determine the effects of default costs on pricing and expected default behavior we assume
the existence of constants $k_b$ and $k_l$ that capture the default costs of the borrower and lender,
respectively. It is assumed that these costs are not received as revenue by any party (of course, one
can model costs paid by the lender and received by the borrower, or vice-versa, by setting $k_l$ equal
to $-k_b$). We also assume that default costs are proportional to the initial value of the house at the

---

3The press has reported many instances when defaulting borrowers stole appliances and vandalized the properties
upon vacating foreclosed houses. To forestall such behavior, some lenders have offered payments of several thousand
dollars to owners of foreclosed properties who turned them over to lenders in satisfactory condition.
origination date. We did this so as to be able to carry over the assumption of scale-invariance, which implies that the convention that \( x = 1 \) at the mortgage origination date is a harmless numeraire choice. Default costs are assumed to be the same for all borrowers, implying unrealistically that all mortgages with the same contract terms will default, and at the same time, if any of them does. Similarly, all lenders have the same costs.

As in the absence of default costs, it is assumed that there is free entry into, and exit from, the lending industry. Accordingly, lenders adjust loan terms so as to pass the discounted expected value of default costs on to borrowers.

Existence of costly default implies that it is necessary to distinguish between the value of a mortgage to the borrower as a liability and its value to the lender as an asset. The easiest way to see this is to observe that when \( x = x_* \) the mortgage has liability value of \( x_*/(\rho - \alpha) + k_b \) to the borrower and asset value of \( x_*/(\rho - \alpha) - k_l \) to the lender. Thus the deadweight loss resulting from default at the time of default equals \( k_b + k_l \). We denote the values of the mortgage to the borrower and the lender as \( M_b(x) \) and \( M_l(x) \). For any \( x \) we have \( M_b(x) > M_l(x) \); this difference is a decreasing function of \( x \) because for high values of \( x \) the possibility of default is minor contributor to the value of the loan to either borrower or lender. Both mortgage values converge to \( c/\rho \) as \( x \) goes to infinity, which equals the value of the mortgage when there is no default.\(^4\)

In this subsection it is assumed that \( k_l = 0 \), so that borrowers, but not lenders, have positive default costs.

\(^4\)The presence of deadweight losses due to costly default implies an inconsistency in the formulation of the model. The model specifies that the discounted value at \( t \) of expected housing services equals \( x_t/(\rho - \alpha) \) regardless of whether or not a mortgage is used. This specification makes no allowance for the present value of the loss due to default risk when a mortgage is used. Thus when either \( k_b \) or \( k_l \) is greater than zero, house buyers using mortgages would find that houses are overpriced relative to the expected value of subsequent housing services.

As we will discuss below, this shortcoming of the model has an unfortunate implication for the modeling of prepayable mortgages. As will be noted, borrowers whose homes have appreciated have a motive to refinance because default risk has decreased, implying that they are overpaying for the default option. It would be natural to assume that borrowers will refinance whenever the value of the mortgage liability exceeds the initiation cost plus the prepayment penalty. If the prepayment penalty is less than \( M_b(1) - M_l(1) \), however, this specification implies that borrowers would repay their loans immediately upon incurring them. This bizarre implication results from the fact that the model gives no explanation for why individuals use mortgages in the first place. Therefore in the presence of default costs the model cannot give a consistent explanation of when agents refinance as it stands.

One could eliminate the problem just noted by setting the initial price at a level lower than \( 1/(\rho - \alpha) \). However, the indicated amount of the decrease depends on the loan-to-value ratio. This is different for different values of \( c \), the determination of which we are not modeling, so it is unclear how great a decrease to specify. Thus this respecification does not provide an easy resolution of the problem. We prefer to avoid the major respecifications that would be required to deal with this difficulty.
Incorporating borrowers’ default costs into the model presented above is achieved by replacing the value-matching condition (1.5) with

\[ E(x_*) = -k_b, \] 

so that borrowers default when their equity equals \(-k_b\). The smooth pasting condition remains \(E'(x_*) = 0\). The values of the mortgage to the borrower and the lender as a function of \(x\) are then calculated by using the boundary conditions as modified to derive the constants of integration of the differential equations representing \(M_b(x)\) and \(M_l(x)\).

For each specification of default costs we computed the loci of pairs of LTV ratios and mortgage yields in the presence of default costs. Whenever default costs are strictly positive there are two loci, one corresponding to the initial value of the loan to the borrower and one corresponding to the initial value of the loan to the lender. In Figure 4 it is assumed that \(k_b = 8\), so that \(k_b\) equals 32% of the assumed purchase price of 25. The figure shows LTV-yield loci associated with both the borrower’s loan value and the lender’s loan value. For comparison we also show the corresponding locus when default costs are zero. As above, we assumed \(\sigma = .1\).

Figure 4 illustrates the consequences of the fact that default is less probable when it is costly to the borrower than when it is costless. For any LTV ratio the equilibrium mortgage yield is lower in the presence of default costs. As would be expected, this effect is more pronounced for high-LTV mortgages.

In the presence of borrower’s costs as just specified the maximum coupon is \(c = 2.82\), so that the mortgage payment is almost three times the initial service yield of the house. Accordingly, the value of the mortgage as a liability to the borrower equals 132% of the value of the house. At that value of \(c\) the lender would supply 100% of the value of the house. The borrower would default virtually immediately (with probability 1), incurring default costs equal by assumption to 32% of the value of the house.

Now consider lower values of \(c\). In that case the borrower would not default immediately. Since the lender would receive high payments for some strictly positive interval of time, he would be willing to supply more cash than when \(c = 2.82\). This explains the backward-bending LTV-yield
Figure 4: Mortgage Yield as Function of LTV, Borrower Default Costs

\[ k_b=8, k_l=0 \]
locus based on the lender’s mortgage value (indicated in Figure 4).

We would not expect to see values of $c$ between 2.38 and 2.82 in equilibrium. For these high values of $c$ lenders would be willing to supply more cash if at the same time they decreased the coupon, so as to decrease the probability of default. Borrowers would insist that they do so. The maximum mortgage payment that is not dominated in this way is that which maximizes the lender’s LTV. As the diagram indicates, this is $c = 2.38$, and the LTV associated with that value of $c$ is about 110%. At this coupon the lender supplies 110% of the value of the house, and the liability value of the mortgage to the borrower is 125%. This value of $c$ produces a yield of 8.75%. For successively lower coupons the lender is willing to supply lower amounts of cash, and mortgages are priced to generate yields less than 8.75%.

2.2 Costly Lender Default

In this section we set the borrower’s default cost $k_b$ equal to zero and set the lender’s default cost $k_l$ equal to 8. Nonzero $k_l$ does not affect the default point relative to the case in which default costs are zero; the borrower has no incentive to take any account of the seller’s default costs. Figure 5 shows that for any yield the lender will supply less cash in the presence of default costs compared to the zero default cost case, for the obvious reason that the mortgage is costlier to him in the event of default.

With lenders’ default costs the maximal mortgage payment is 2.26, at which point default is immediate with probability 1. The existence of a 32% default cost implies that the lender would supply 68% of the purchase price.

As with borrowers’ costs, the lenders’ LTV-yield locus is backward-bending, and the upper portion is dominated. The maximal value of $c$ that is not dominated is $c = 1.83$, at which value the lender will supply 78% of the value of the house. At this value of $c$ the yield on the mortgage is 9.40% and the value of the mortgage as a liability to the borrower is 93%.
Figure 5: Mortgage Yield as Function of LTV, Lender Default Costs

\( k_b = 0, k_l = 8 \)
2.3 Costly Prepayment

Up to now it has been assumed that mortgages are terminated only by default, so that, conditional on no default, mortgages are perpetuities. This is obviously a simplification: borrowers have the option to prepay mortgages at their book values, possibly subject to a prepayment penalty. In the real world they elect to prepay and refinance when prevailing interest rates drop, since otherwise the present value of the mortgage payments would exceed the mortgage’s book value. Under the present model this feature is absent since we are taking interest rates to be deterministic and constant.

As noted in the introduction, in the present model increases in house prices generate a motive for prepayment even when interest rates are deterministic. This is so because mortgage values increase with \( x \) due to the decreased probability of default in the economically relevant future. For \( x > 1 \) borrowers are paying a higher price for default risk than is justified actuarially, since mortgage payments are based on the default probability appropriate at the origination date, when \( x = 1 \). This, of course, is the reason we have \( M(x) > M(1) \) for \( x > 1 \). Thus borrowers have an incentive to prepay their mortgages when the value of the collateral rises above \( 1/(\rho - \alpha) \).

The model predicts that borrowers will be willing to make high periodic payments for high LTV mortgages because they expect to be able to refinance on better terms after generating additional equity due to capital gains. Further, they are protected from large losses due to decreases in property values because of the default option and the low level of their equity. Thus lenders are subjected to default losses if house values drop, but, due to prepayment, are unlikely to benefit from large gains if they rise. We therefore expect that lenders will require high periodic payments for high LTV mortgages that are prepayable, particularly if (1) borrowers’ default costs are low, so that they will default at high values of \( x \), (2) lenders’ default costs are high, and (3) prepayment penalties are low. For the same reasons, borrowers will accept such terms. As emphasized in the financial press, in the past several years we have seen exactly the pattern just described.

In the absence of prepayment costs, borrowers would be motivated to repay continuously whenever \( x > 1 \). To avoid this unrealistic prediction, we assume that prepayment is subject to a penalty \( k_p \), so that borrowers will prepay only when the values of their mortgages exceed their origination
values $M(1)$ by at least $k_p$.\(^5\) It is assumed that prepayment is subject to penalty $k_p$ regardless of when the prepayment occurs. In reality prepayment penalties apply only during the first several years of the mortgage. Our specification has the advantage that it makes possible application of the stationary setting adopted in earlier sections.

Define $x^*$ as the level of $x$ at which borrowers elect to prepay. To derive an upper bound for $k_p$, note that if $k_p > c/\rho - M(1)$, the amount the borrower would have to prepay exceeds the maximum value of the mortgage. In that case the borrower will never prepay ($x^* = \infty$). Thus in the absence of prepayment costs we have $1 < x^* < \infty$ only when $0 < k_p < c/\rho - M(1)$. We assume that these inequalities are satisfied.

As shown above, the differential equation defining $E(x)$ has the solution

$$E(x) = e_1 x^{m_1} + e_2 x^{m_2} + \frac{x}{\rho - \alpha} - \frac{c}{\rho}. \quad (2.2)$$

Here, as opposed to the derivation of $E(x)$ when no prepayment was allowed, both constants of integration $e_1$ and $e_2$ are nonzero. The roots $m_1$ and $m_2$ are given by

$$m_1, m_2 = \frac{-(\alpha - \frac{1}{2}\sigma^2) \pm \sqrt{(\alpha - \frac{1}{2}\sigma^2)^2 + 2\sigma^2 \rho}}{\sigma^2}. \quad (2.3)$$

The boundary conditions are the value-matching conditions

$$E(x_*) = 0 \quad (2.4)$$

and

$$E(x^*) = \frac{x^* - 1}{\rho - \alpha} + E(1) - k_p. \quad (2.5)$$

\(^5\)Note that by deleting the subscript on $M(x)$ we are excluding default costs. There is a conceptual problem in modeling prepayment in the presence of default costs. If the prepayment penalty is less than the deadweight loss due to default costs at $x = 1$, then default will occur at $x = 1$. To see this, suppose that $k_p < M_b(1) - M_l(1)$. Then by immediately prepaying their mortgages, borrowers can replace the liability $M_b(1)$ with the cost $k_p + M_l(1)$, which by assumption is lower. The bizarre implication that borrowers would prepay a mortgage that they had just incurred reflects the exogenous specification that homeowners borrow despite the deadweight costs of doing so (when $k_b$ or $k_l$ exceeds zero), but are free to terminate the loan at any time upon payment of the penalty $k_p$ (see the discussion in note 2). To ensure that the default point is strictly higher than 1 it is sufficient to restrict attention to values of $k_p$ that exceed $M_b(1) - M_l(1)$.

We sidestep this problem by allowing borrowers to have a prepayment option only when there are no default costs.
The smooth pasting conditions are

\[ E'(x) = 0 \quad \text{and} \quad E'(x^*) = \frac{1}{\rho - \alpha}. \]  

(2.6)

Define \( x_{**} \) as the default point when there is a prepayment option. Then higher values of \( k_p \) induce higher values of both \( x^* \) and \( x_{**} \). To see the latter effect, suppose that there is no prepayment, so that the borrower is indifferent between defaulting and continuing to pay the mortgage when \( x = x^* \). Now suppose that prepayment is permitted. Existence of the prepayment option increases the equity value of the mortgage, implying that \( E(x^*) > 0 \). It follows that \( x_{**} < x^* \).

Figure 6 shows mortgage yield as a function of LTV under \( \sigma = .15 \) and \( c = 1.5 \), assuming several prepayment regimes. The solid line shows mortgage values in the absence of a prepayment option. This specification yields the lowest mortgage yield for each value of LTV, as would be expected because then the lender benefits from dropping default risk as \( x \) increases. The other extreme, generated by \( k_p = 0.1 \) and indicated by the dotted line, produces a high yield for each LTV since the lender will benefit only slightly from a capital gain. The cases \( k_p = 0.5 \) and \( k_p = 0.25 \) and \( k_p = 0.5 \) are intermediate.

As Downing, Stanton, and Wallace (2005) observed, the prepayment option interacts with the default option. Most obviously, in the present model borrowers would have no motive to prepay in the absence of the default option. Conversely, the optimal default point depends on the prepayment regime: the cheaper prepayment is, the lower the optimal default point. Figure 7 shows the interdependence of prepayment penalty and default points.

3 Implementation of the Model

3.1 Data

We used data from LPS Applied Analytics (LPS). LPS collects loan-level data on mortgage characteristics and performance from nine of the top ten US mortgage servicers, accounting for about 60 percent of the entire U.S. mortgage market. The overall historical data set contains about 107
Figure 6: Mortgage Yield as Function of LTV, Prepayment
Figure 7: Mortgage Value with and without Prepayment Penalty
million loans, with about 33 million active loans as of the end of 2008.

In order to reduce the data set to a more manageable size we restricted attention to loans made on properties in California. The California housing market, of course, is extremely diverse, and in many respects exhibits the same within-market variation present in the national housing market. Price appreciation was generally strong in most markets until 2006, and some markets, such as those in the Inland Empire and the Central Valley, benefited from so-called “innovations” in mortgage lending that increased the supply of credit flowing to previously marginal borrowers. As is well known, these trends reversed after 2006.

Our sample contains loans originated in California between 2000 and 2007. We limited our attention to conventional first lien mortgage loans originated for home purchases. We adopted this restriction because refinancing decisions are governed largely (but not entirely, as discussed above) by interest rate changes, which are suppressed in the model presented above.\(^6\)

<table>
<thead>
<tr>
<th>Table 1: Summary statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Full sample</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>2002</td>
</tr>
<tr>
<td>mean   std. dev.</td>
</tr>
<tr>
<td>Origination amount ($000)</td>
</tr>
<tr>
<td>392       265</td>
</tr>
<tr>
<td>Term of Loan (months)</td>
</tr>
<tr>
<td>360       41</td>
</tr>
<tr>
<td>Loan-to-value</td>
</tr>
<tr>
<td>.75       .14</td>
</tr>
<tr>
<td>Subprime</td>
</tr>
<tr>
<td>.05       .03</td>
</tr>
<tr>
<td>Jumbo</td>
</tr>
<tr>
<td>.42       .29</td>
</tr>
<tr>
<td>Fixed-rate</td>
</tr>
<tr>
<td>.45       .67</td>
</tr>
<tr>
<td>Adjustable-rate</td>
</tr>
<tr>
<td>.54       .32</td>
</tr>
<tr>
<td>Securitized</td>
</tr>
<tr>
<td>.82       .84</td>
</tr>
<tr>
<td>Non-agency securitization</td>
</tr>
<tr>
<td>.44       .27</td>
</tr>
<tr>
<td>FICO at origination</td>
</tr>
<tr>
<td>724       54</td>
</tr>
<tr>
<td>10yr Treasury spread</td>
</tr>
<tr>
<td>1.78%      .52%</td>
</tr>
<tr>
<td>ARM margin over benchmark</td>
</tr>
<tr>
<td>2.84%      1.24%</td>
</tr>
<tr>
<td>Observations</td>
</tr>
<tr>
<td>993,943     25</td>
</tr>
</tbody>
</table>

Even with these restrictions, the sample consists of nearly 1 million mortgages. Summary data for the sample are shown in Table 1. The mean loan amount was $360,000 and the mean loan term is 30 years. About 5 percent of the loans were labeled by the originator as being subprime.\(^7\)

---

\(^6\)Gerardi, Shapiro, and Willen (2008) called attention to the tendency for borrowers to extract equity and thereby alter their leverage through a cash-out refinance. We abstract from this phenomenon, as borrowers in our model face no credit constraints.

\(^7\)Mayer, Pence, and Sherlund (2008) reported that subprime loan originations increased from about 10% of total
About 42 percent of the loans were nonconforming jumbo loans, a much higher figure than would obtain nationally. This statistic represents the fact that California housing prices are higher than those in other parts of the country, implying that a larger proportion of California loans exceed the upper bound for conforming loan amounts than is true nationwide. A full 84 percent of the loans were securitized. More than half of the securitized loans were nonagency, or private-label, securitizations.

Many of the variables in Table 1 are interpreted naturally as risk factors. As such, they do not enter directly into the formal model developed above: in the model borrowers are treated as homogeneous and are assumed to be able to refinance whenever the market value of equity is positive. As noted, that assumption is unrealistic empirically. As a result of borrowers’ inability to prepay mortgages at market values when these values have dropped due to housing price decreases, one would expect default rates to reflect borrowers’ risk factors as well as house price changes. Accordingly, risk factors such as the credit score and other variables speaking to ability to repay mortgages would be expected to influence mortgage prices.

The credit score (FICO) is the main demographic control used in the empirical analysis. The mean FICO score in the sample is 724, which would generally indicate a prime borrower. Despite the fact that the theory presented above deals only with fixed-rate mortgages (FRMs), we will frequently make comparisons between FRMs and adjustable-rate mortgages (ARMs). We do this because it turns out that the two classes of mortgages have different default behavior as well as pricing characteristics.

The fact that our formal model does not incorporate borrower risk variables would not be a major concern if these risk variables were uncorrelated with LTV. However, this uncorrelatedness property is not satisfied in the data. Table 2 shows a correlation matrix between LTV ratios and several risk variables. As expected, high LTV ratios are associated with risky borrowers; that is, borrowers with low FICO scores. Borrowers with high LTVs are more likely to have an adjustable-rate mortgage than a fixed-rate mortgage. To deal with these associations we will include controls for borrower risk in many of the tabulations.

---

mortgage originations nationwide in the early 2000s to a peak of 35% of total originations. Thus, subprime loans appear to be underrepresented in the LPS data.
Table 2: Correlation matrix for risk factors

<table>
<thead>
<tr>
<th></th>
<th>LTV</th>
<th>FICO</th>
<th>ARM</th>
<th>Jumbo</th>
</tr>
</thead>
<tbody>
<tr>
<td>LTV</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FICO</td>
<td>-0.24</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARM</td>
<td>0.12</td>
<td>-0.12</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Jumbo</td>
<td>0.05</td>
<td>0.09</td>
<td>0.34</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 1 also displays summary statistics for mortgages originated in two years, 2002 and 2006. Since 2002 followed a recession and 2006 coincided with the peak of the housing boom, comparison of these two years provides a useful summary of mortgage market changes over time. First, note that the run-up of housing prices in the first half of the current decade resulted in a sharp increase in average loan size: from $290,000 in 2002 to $461,000 in 2006. One might expect that average LTVs would drop over the period since individuals changing homes in 2006 could carry over major capital gains, which was much less true in 2002. However, there was no change in average LTV ratios, pointing to the fact that individuals were apparently using debt to finance higher levels of consumption than would otherwise be possible. It is important to remember that the LTV ratios used here are computed from the first lien mortgage. As we discuss below, true leverage was almost surely higher in aggregate than what is reported in Table 1.

The proportion of the sample labeled subprime almost tripled, from 3% to 8%, between 2002 and 2006. The proportion of loans that exceeded conforming loan limits set by the GSEs (the jumbo share) increased from 29 per cent to 43 per cent. Significantly, in 2002 about two-thirds of the loans were FRMs, with the remainder being adjustable-rate, while these ratios were reversed in 2006. Again, the increase in loan amounts necessitated by rising housing prices led rising numbers of borrowers to use adjustable rate mortgages, for which qualifying is easier. Finally, securitization rates were much higher in 2006 than in 2002.

Figure 8 shows that the distribution of borrower FICO scores shifted down somewhat between 2002 and 2006. That FICO scores dropped at all over this period is surprising: housing prices had risen sharply over that period, implying that most homeowners in financial difficulty could readily

---

8FRMs and ARMs make up about 99% of the loans in the sample in any given year. Interest-only and option-ARMs are classified as adjustable-rate mortgages for these purposes. The remaining 1% of the sample consists of various alternative mortgages, such as graduated payment mortgages.
refinance or sell in order to avoid defaulting on financial obligations. The fact that FICO scores dropped despite this fact is consistent with other evidence of declining underwriting standards during this time period (see Demyanyk and van Hemert (2008)).

### 3.2 LTV and Default

The theoretical model presented above focused on initial LTV and changes in house prices as the major determinants of default. It is not difficult to link house prices to default behavior in the data.\(^9\) House prices in California doubled between 2000 and the peak in 2006, and then proceeded to fall by more than 25% through year-end 2008. In the early part of this time period delinquency rates were very low and stable, before starting a dramatic upward surge at about the same time as the peak in house prices. Generally speaking, California housing markets that experienced large price declines were also characterized by high default rates (see Figure 9).

Theory suggests that default rates for borrowers with high LTVs will be more sensitive to house price declines than those for low LTV borrowers. Unfortunately, directly testing this implication

\(^9\)See Doms, Furlong, and Krainer (2008), Gerardi, Shapiro, and Willen (2008), and the references therein for a discussion of the house price-default correlation.
of the model is not straightforward. The problem is that, as noted above, high-LTV borrowers frequently combine an 80% first-lien mortgage with a second lien rather than borrow the full amount using a single mortgage. We have no information about which first-lien mortgages are linked to a second-lien mortgages, but have every suspicion that such linkages exist. To see this, consult Figure 10, which shows the distribution of LTV ratios in 2002 and 2006 for either FRMs and ARMs. In 2002 a considerable fraction (22%) of the first-lien FRMs had LTVs greater than 80%. This fraction diminished to about 6% in 2006. As Figure 10 suggests, the reshaping of this distribution resulted in a strong concentration of loans with exactly 80% LTV. Given the sharply rising house prices over this time period that generally outpaced increases in income, the presumption is that a substantial fraction of the 80% mortgages were associated with second-lien mortgages, raising the combined LTV above the level indicated by the data by an amount that we cannot directly measure.

Obviously the optimal default point on an 80% first-lien mortgage will depend on whether or not there exists a second mortgage on the property. Indeed, in our data, the default rate on 80% LTV mortgages is higher than the default rate for both 79% and 81% LTV mortgages. 10 Away

10Our model suggests that the optimal default point rises in the presence of a second mortgage. This higher default probability need not accompany a higher loan price for the first mortgage, however. In essence, the second mortgage provides protection by raising the recovery value to the first lien holder, as noted in Subsection 1.5.
Figure 10: Kernel density estimates of LTV distribution

Fixed-rate loans

Adjustable-rate loans

Solid=2002, Dash=2006
from 80%, however, we have more confidence that the leverage implied by the first lien on the property is an accurate measure of total leverage.

Figure 11 displays the cumulative default frequencies plotted against the time since origination for loans in our data set. Mindful of the potential problems interpreting the default behavior for 80% LTV loans, we defined a low LTV mortgage to be one where $0.65 < LTV < 0.75$, and a corresponding high LTV mortgage by $0.85 < LTV < 0.95$. The high-LTV mortgages default at about twice the rate of low-LTV mortgages. Note that even though the relative rates of default between high and low LTV mortgages are about the same for FRMs and ARMs, the level of the default rate on ARMs is much higher than on FRMs.

So far all these results, following the theoretical model, ignore borrowers’ risk factors. In the discussion above we noted that because borrowers are often unable to repay mortgages at their
economic values when these have dropped, default frequencies would be expected to depend on borrowers’ risk variables as well as LTV and house prices. This dependence is seen in the data. The two right-most panels of Figure 11 show the differential performance of high FICO score mortgages compared to low FICO score mortgages. FICO score is negatively correlated with LTV in our data. However, holding LTV constant, we still see that low FICO score borrowers default much more frequently over time than high FICO score borrowers. For loans with fixed-rate mortgages and \(0.85 < LTV < 0.95\), the default rate 40 months after origination was about 5% for borrowers with FICO scores below 700, as compared to 2% for borrowers with FICO scores above 700. For adjustable rate loans, the corresponding default rates were 20% and 8% for high and low FICO score borrowers, respectively.

We can also see the clear effects of differential house price appreciation on the cumulative default rates. For all classes of mortgages—fixed and adjustables—and for all ranges of LTV—the 2002 vintages performed much better than the 2006 vintages (see Figure 12). Of course, virtually all mortgage borrowers in the 2006 vintage experienced house price depreciation over the following three years. Even so, it is remarkable to see the disastrous performance of the high LTV ARMs in the 2006 vintage. Within 24 months, nearly one-half of this vintage was in default.

In summary, the data provide clear evidence that LTV is a risk factor; for both FRMs and ARMs, high LTV loans have significantly higher default rates than low LTV loans. Thus, it will be important to control for LTV and other risk factors when estimating the yield premium for high LTV loans.

### 3.3 LTV and Mortgage Pricing

In the theoretical model it was assumed that equilibrium mortgage pricing takes account of default risk. Because initial LTV governs default risk, the model predicts that mortgage spreads (the risk premium on mortgages, equal to the difference between the coupon rate on mortgages less the risk-free return) depend positively on initial LTV (because the optimal default point is higher for high-LTV mortgages, implying that default is more probable with high-LTV mortgages).

Our measure of spread varies by mortgage type. For FRMs with loan terms of 15 years or more,
Figure 12:

Cumulative Default Rates by Vintage

FRMs
Low LTV is .65<LTV<.75

ARMs
Low LTV is .65<LTV<.75
we used the difference between the fixed mortgage rate and the average 10-year constant maturity Treasury yield in the 30 days leading up to the date the mortgage was originated. For ARMs, we used the so-called “margin rate”, which is the difference between the mortgage rate and the benchmark rate to which the ARM is tied. To estimate the spread-LTV relationship, we defined a set of dummy variables $d_j$ that take on value 1 if $j\% < LTV \leq j + 5\%$, and zero otherwise. We then regressed the spread, $s_{i,j,k}$, of mortgage $i$ with LTV in interval $j\%, j + 5\%$ in market $k$ on the dummy variables and a set of risk controls,

$$s_{i,j,k} = a_k + \sum_j d_j + X_i \beta + \epsilon_i.$$  

The market-specific intercept is meant to capture possible variation in expectations for future house price behavior across markets which, as highlighted in the model section, are crucial inputs to the loan pricing problem. In this analysis, the definition of a market is a metropolitan statistical area (MSA). All mortgages on houses not located in an MSA are coded as “rural.” The set of risk factors, $X_i$, are meant to control for sources of risk that were not considered in the theory, but are in fact priced by lenders and possibly correlated with LTV. To this end, we included credit score, the log of loan amount, and indicators for whether the loan is a subprime loan, exceeds the conforming loan limit (i.e. is a jumbo loan), is securitized in a nonagency securitization, has fully documented income, and whether or not it has a prepayment penalty.

We also entered year dummies to capture the behavior of systematic economic factors over time. The loan term was taken to be a risk factor to proxy for interest rate risk. When considering ARMs, we also included indicators for whether the mortgage is interest-only, and option-ARM, as well as dummy variables recording the benchmark index that the ARM rate is based on.\textsuperscript{11} In all of the results to be presented, we dropped mortgages that had mortgage insurance. Finally, the $\epsilon_i$ in the above specification is the random disturbance.

If we omit all regressors apart from the dummy variables denoting LTV and drop the constant term, then the estimated coefficients on the dummy variables will equal the average spread for that LTV bucket. In our estimation, we included a constant term and excluded the dummy variable for

\textsuperscript{11}In our data, the vast majority of the ARMs are indexed to either the 1-year T-bill or 6-month LIBOR.
the lowest LTV bucket included in the sample—45-50%. Thus, a coefficient estimate on a particular LTV bucket—say 85-90% LTV—is the estimate of the excess spread that a particular borrower in a particular market paid to raise the LTV from the 45-50% range to 85-90%.

Figure 13 shows estimated mortgage spreads against LTV for FRMs and ARMs, separately estimated. The solid line shows coefficient estimates for the LTV dummies without correction for risk factors other than LTV, while the dashed line plots the coefficient estimates when the risk factors were included in the specification. Generally, the inclusion of the risk factors acts to depress the estimated spreads at all LTV points, although for FRMs the difference is slight. High LTV mortgages are characterized by spreads that are about 40 basis points higher than for low LTV mortgages. Evidently, in the data, the premium on high-LTV FRM loans does not depend much on whether or not one corrects for risk factors. This is contrary to our prior expectation, which was that the spread-LTV locus would be flatter when risk controls were included than when they were not. The risk factors considered here do have an effect on loan pricing: low credit score borrowers tend to have higher mortgage rates, all else held equal. But controlling for these risk factors does not appear to have much of an effect on estimates of the difference between the rates on high and low-LTV mortgages, respectively.

For ARMs, the estimated spread-LTV relationship is more complex. First, the estimated difference between spreads on high LTV mortgages and low is much greater at about 90 basis points, compared to FRMs. Thus, even controlling for a rich set of risk factors, high LTV borrowers with ARMs were perceived by lenders to be much riskier than low LTV borrowers with ARMs. In the previous section we saw that these beliefs were justified in view of the fact that default rates have been much higher for ARMS than FRMs. The second point to notice about the estimated relationship in Figure 13 for ARMs is that yield spreads vary considerably depending on whether or not one corrects for the other risk factors, in contrast to the case for FRMs. For ARMs, the correlation between these other factors and LTV is much more important. Without the risk adjustments, we would believe that the premium for high LTV borrowers is nearly 150 basis points.

4 Calibrating the Model
Figure 13:

Spreads on FRMs and ARMs as function of LTV
The final exercise is to attempt to calibrate the model and compare actual mortgage pricing to model predictions. We will set the housing service flow drift at \( \alpha = 3\% \) and the rate of time preference equal to \( \rho = 7\% \); these values produce empirically realistic values for the real return on housing and mortgages, and a realistic value for the price-rent ratio, as discussed above. With these maintained specifications in place, the key determinants of mortgage rates and expected default behavior in the model will be the volatility of changes in housing services \((\sigma)\) and the assumptions on borrower and lender default costs \((k_b \text{ and } k_l)\).

We will focus on mortgages with LTV around 95%. Excluding mortgages with LTV equal to 80% is appropriate because, as discussed above, many of these mortgages are associated with second-lien loans. This difficulty probably does not occur in mortgages with LTV less than 80%, but another problem arises. Mortgages with LTV below 80% are likely to be eligible for guarantee by Fannie Mae and Freddie Mac. Consequently, whether or not such mortgages are actually sold to Fannie Mae and Freddie Mac upon origination, mortgage lenders have no incentive to incorporate default costs in pricing except insofar as these costs are implicitly reflected in the prices at which the mortgages can be sold. Whether or not this is the case is far from clear. Also, default risk is a minor factor in the pricing of low-LTV mortgages, so in any case such mortgages are of secondary interest given our focus on default and its consequences for mortgage pricing. We are left with mortgages with LTV over 80%; for concreteness, we focus on mortgages with LTV around 95%.

In view of the fact that we do not have a theoretical model for adjustable-rate mortgages, our attention in this section is restricted to fixed-rate mortgages.

We assumed that mortgages are not prepayable \((k_p = \infty)\). Relaxing this assumption and incorporating a prepayment penalty of \(k_p\) has the same qualitative effect on the spread-LTV plot as imposing strictly positive lender’s costs (doing either makes the spread-LTV plot more convex; compare Figures 4 and 5). Therefore, having allowed for strictly positive \(k_l\), we cannot infer the effects of \(k_p\) with any accuracy.

Calibrating the model will consist of specifying parameters that generate empirically accurate values of the spread on mortgages with LTV around 95%, and that generate default behavior that appears to be reasonable. We showed above that, empirically, the spread for 95% mortgages is
around 30-40 basis points, so we prefer parameter values that generate approximately that value. Reproducing default behavior is more difficult. Ideally, we would back out the default point from the recovery rate and the other parameter values. Our data set contains the recovery rate as a variable, but it is extremely sparsely populated, so this estimator is not open to us. Using actual default rates will not allow a precise calibration because default rates are extremely variable, even among mortgages issued at the same time with equal LTVs. We are reduced to imposing on the data an arbitrary prior specification of default points and recovery rates. We assumed that the recovery rate on a 95% mortgage is around 60-70%, and calibrate to this value.\footnote{If \( k_l = 0 \), calibrating from the recovery rate is equivalent to calibrating from the default point; recall that if \( k_l = 0 \) the recovery rate is the quotient of the default point and LTV. If \( k_l > 0 \), for given default point the recovery rate relevant for the lender will be lower than that resulting from this calculation.

It is worth noting from this that if one had separate priors about the default point and the recovery rate, one could in principle directly calibrate \( k_l \).}

We begin by considering the specification of \( \sigma \). In the empirical housing literature estimated price volatility is around 15%. For example, using data from the PSID, Flavin and Yamashita (2002) estimated the standard deviation of the real return on housing to be about 14%. Case and Shiller (1989) report similar estimates for the return on individual houses (14-15%).\footnote{The Flavin and Yamashita and the Case and Shiller estimates of the standard deviation of housing returns are based on data sampled from across the United States. Our sample is drawn solely from California, arguably implying higher values for the volatility parameter.} After aggregating individual units to the city or the national level, most house price indices display much less volatility. The annual price change in the Case-Shiller Composite 10 City index has a standard deviation of about 9% between 1987 and 2008. The annual price change in the FHFA house price index is a little over 3% over the same time period. Since the loan pricing that we are studying corresponds to rates on individual mortgages, we take the higher estimates of \( \sigma \) to be more realistic.

We will start by specifying \( \sigma = .15 \), but we will also consider \( \sigma = .1 \).

Table 3 reports computations from the theoretical pricing model that are useful in the calibration exercise. Each panel displays the solutions for parameters of interest under a particular specification of \( \sigma, k_b \) and \( k_l \). The rows of the table show the values of LTV, the yield, the recovery rate and the default point under various specifications of \( \sigma, k_b \) and \( k_l \). The columns are associated with values of \( c \) (not shown) that increase moving toward the right; as noted, in each panel we will focus on the value of \( c \) that produces the LTV closest to 95% (indicated in boldface).
Table 3: Model-calibrated yields and LTV

<table>
<thead>
<tr>
<th></th>
<th>(c) 0.5</th>
<th>0.75</th>
<th>1.00</th>
<th>1.25</th>
<th>1.50</th>
<th>1.75</th>
<th>2.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>LTV(%)</td>
<td>28.54</td>
<td>42.65</td>
<td>56.38</td>
<td>69.36</td>
<td>81.04</td>
<td>90.69</td>
<td><strong>97.40</strong></td>
</tr>
<tr>
<td>Yield(%)</td>
<td>7.01</td>
<td>7.04</td>
<td>7.10</td>
<td>7.21</td>
<td>7.40</td>
<td>7.72</td>
<td><strong>8.21</strong></td>
</tr>
<tr>
<td>Recovery(%)</td>
<td>77.69</td>
<td>77.98</td>
<td>78.65</td>
<td>79.91</td>
<td>82.07</td>
<td>85.56</td>
<td><strong>91.05</strong></td>
</tr>
<tr>
<td>Default point</td>
<td>0.22</td>
<td>0.33</td>
<td>0.44</td>
<td>0.55</td>
<td>0.67</td>
<td>0.78</td>
<td><strong>0.89</strong></td>
</tr>
</tbody>
</table>

**Notes:**
- \(\sigma = 0.15, k_b = k_l = 0\)
- \(\sigma = 0.1, k_b = k_l = 0\)
- \(\sigma = 0.15, k_b = 4, k_l = 0\)
- \(\sigma = 0.15, k_b = 8, k_l = 0\)
- \(\sigma = 0.15, k_b = 0, k_l = 4\)
- \(\sigma = 0.15, k_b = 4, k_l = 4\)
- \(\sigma = 0.1, k_b = 4, k_l = 4\)
The first project is to determine whether one can produce an empirically accurate model under \( k_b = k_l = 0 \). Panel 1 shows that the spread on a 97% mortgage is 121 basis points, the recovery rate is 91% and the default point is 89%. These values are all much higher than we see in the data, so this specification must be rejected.

A lower value of \( \sigma \) would be expected to render default less probable (at least under the assumption that the effect of reducing \( \sigma \) on the default point is minor). Panel 2 shows that the effect of reducing volatility to \( \sigma = 0.1 \) is to produce a spread to 36 basis points on a 95% mortgage, approximately equaling that seen in the data. However, the effects of specifying the lower value of \( \sigma \) on the default point and the recovery rate are minor, implying that the predicted values for these variables remain much higher than the values we are looking for. We conclude that, despite the simplification that would result from excluding default costs, doing so does not produce a satisfactory calibration.

We have seen that when borrowers incur costs upon default, the spread-LTV relation becomes flatter; for given LTV, default costs decrease the default point and decrease the effects of default on the spread. Such a change would be expected to move the model predictions in the direction of the data. Panel 3 shows \( \sigma = 0.15, k_b = 4, k_l = 0 \). Making this change reduces the spread on a 92% mortgage to 60 basis points, only slightly higher than what we see in the data. The recovery rate and default point are 71% and 65%, which are also close to the values we seek.

One could make a case for an even higher value for \( k_b \); if \( k_b = 8 \), so that default costs equal 32% of the purchase price of a house, then the yield spread on a 94% loan is 38 basis points, the recovery rate is 56% and the default point is 53% (Panel 4). These are reasonable values.

On prior grounds it is unclear whether borrowers’ default costs are positive or negative. However, there is not much doubt that lenders’ default costs are positive and high. Therefore we might want to impose \( k_l > 0 \) on any candidate set of parameter values. However, we have already seen that introducing lenders’ costs increases the curvature of the spread-LTV curve, so setting \( k_l > 0 \) together with \( k_b = 0 \) is not likely to produce an accurate calibration. Indeed, as Panel 5 shows, under these parameter values the model cannot produce LTV anywhere near 95% for any value of \( c \). For LTV=87% default occurs at an unrealistically high value of \( x_\ast (x_\ast = 0.89) \), and lenders would
require a spread of 221 basis points to compensate them for expected default costs. These values are a long way from those seen in the data.

It follows from this that $k_l > 0$ cannot produce an accurate calibration in conjunction with $k_b = 0$, but this is not necessarily the case if $k_b$ is also strictly positive. Borrowers’ default costs decrease the default point, therefore attenuating the effect of default on LTV (for given $c$) and the spread. In Panel 6 we show $\sigma = .15$, $k_b = k_l = 4$. Under these values the lender is willing to fund a 93% loan, but only under a spread of 159 basis points, which again is much higher than in the data. The recovery rate is 65% and the default point is 76%, which are also probably too high.

In Panel 7 we continue to assume $k_b = k_l = 4$, but set $\sigma$ equal to .1. On a 95% loan this specification produces a spread of 36 basis points, a recovery rate of 60% and a default point of 74%. All of these values are plausible, although as noted the specification $\sigma = .1$ is itself probably too low.

In sum, we conclude that Panels 3 ($\sigma = .15$, $k_b = 4$, $k_l = 0$) and 7 ($\sigma = .1$, $k_b = 4$, $k_l = 4$) produce the best results. The specification of Panel 3 has the drawback that it unrealistically sets lender costs at zero, while the Panel 7 run specifies a value of $\sigma$ that is probably unrealistically low.

5 Conclusions

In the introduction we observed that the high mortgage default rates of the past several years, together with the resulting huge losses to holders of mortgage-backed securities, could reflect either drastic mispricing of mortgages or extreme realizations of shocks. We assumed initially that it would not be difficult to find evidence that mispricing is the major culprit. However, that has not been the case: we are able to calibrate a model of optimal mortgage default and mortgage pricing that has reasonable success in reproducing the patterns in the data. Assuming Wall Street firms were using models similar to that presented here (granted, it is not clear that this was the case), it follows that mispricing is of secondary importance relative to extreme realizations of house price shocks. Of course, with hindsight it is easy to criticize lenders and borrowers for not predicting that the housing price bubble would eventually be reversed, and acting accordingly. Our point is only that models that rule out foresight of house price changes—a realistic specification, in our
opinion—can be calibrated to the data with fair success.

Let us be a bit more specific about the successes and failures of our modeling exercise. Most obviously, the model here does not predict the event of default with anything like complete accuracy. As noted, the model predicts that all the mortgages with a given LTV initiated at a given date will all default at the same time if any of them defaults. In fact there is wide variation in the default points. This variation reflects the importance of life events in motivating default. The assumption here that borrowers can always buy back their mortgages at prices equal to economic values had the effect of shutting down his line of causation. It should not be surprising that considerable loss of explanatory power results.

The qualitative predictions of the model for default and pricing turn out to be largely substantiated in the data. High-LTV mortgages are more prone to default than low-LTV mortgages, as the model (or common sense) predicts. Further, the dependence of mortgage pricing on LTV conforms to the prediction of the model (more precisely, we can find parameter values for which this is the case).

One striking finding is that the effect of LTV on yield spreads appears to be separable from the other risk factors; specifically, the difference between the yields on two otherwise identical mortgages with different LTV is the same whether or not one corrects for risk factors other than LTV. This equality is remarkably accurate for fixed-rate mortgages; the same result holds to a lower degree of approximation for adjustable-rate mortgages. The interpretation is that the effect of LTV on default does not interact to a major extent with other risk factors. Having this result greatly simplifies the analysis of the effect of LTV on default and mortgage pricing.

Two other caveats made in the text above bear repeating here. First, the equilibrium depends on parameters like default costs that are extremely difficult to measure directly. As a result of our inability to calibrate these parameters credibly, we take the view that model is not readily testable. Second, we noted that our data set does not allow us to link first- and second-lien mortgages, make it difficult to measure LTVs on 80% mortgages. We proceeded despite this problem, but there is no denying its importance.

One of the most important empirical conclusions stated above is that fixed-rate and adjustable-
rate mortgages are very different: high-LTV adjustable-rate mortgages are much more prone to default than fixed-rate mortgages with the same LTV. Further, this dramatic difference in default behavior is reflected in pricing in exactly the way the model predicts: the yield premium on high-LTV adjustable-rate mortgages appears to be much larger than the yield premium for otherwise similar fixed-rate mortgages. This difference reflects the fact that adjustable-rate mortgages are easier to qualify for than fixed-rate mortgages, implying that highly qualified borrowers make greater use of fixed-rate mortgages and less-qualified borrowers specialize in adjustable-rate mortgages. This observation, while correct, leaves open the question why lenders set the standards for qualification so differently in the two cases. We are unable to answer this question.

6 References


