Pricing Credit Default Swaps on Mortgage Backed Securities

Thomas, Drayton M. and Dobelman, John A.

Department of Statistics
Rice University, 6100 Main Street, Houston, TX 77005, USA

Technical Report TR-2013-02
Department of Statistics, Rice University
Abstract

One of the principal roots of the 2007-09 U.S. financial crisis was limited regulation and oversight of the secondary housing market. While blatant fraud in the pooling and securitization of mortgages was rampant, this paper examines the potential impacts of more “creative” forms of dishonesty and naive assumptions regarding pooled mortgage loans. Spread pricing for credit default swaps on underlying mortgage backed securities is the primary metric used to gauge the financial implications. This paper works with theory, various simulations, and logistic regression analysis to answer questions of interest. The primary finding is that increased variance of certain loan characteristics in underlying pools of mortgages, as well as increased dependency among default events, drastically increases the fair and correct pricing of a CDS contract for those assets.
## Contents

1 Introduction  

2 Study Design and Data Collection  
   2.1 Simulation  
   2.2 Data  

3 Methods for Statistical Analysis  
   3.1 Estimation and Inference  
   3.2 Theoretical Concepts  
      3.2.1 Valuation of a Credit Default Swap  
      3.2.2 Modelling Default for a Collateralized Mortgage Obligation  
      3.2.3 Relaxing the Independence Assumption  
      3.2.4 Generating Default Regressor Variables to Specified Distributions  

4 Results  
   4.1 Logistic Regression Model for Default Probabilities  
   4.2 Underlying Distributions of FICO and LTV  
   4.3 Deviation from Independence  

5 Discussion
1 Introduction

A credit default swap (CDS) is a class of derivative security on an underlying asset, typically some form of bond. A CDS is perhaps best considered as an insurance contract between two parties, the buyer and the seller. The buyer makes periodic premium payments to the seller to insure against the event of the underlying asset defaulting. Each period, if the underlying does not default, then the buyer makes the premium payment to the seller. If the asset never defaults, then the payments are made for the duration of the contract, which is typically the time to maturity of the underlying asset. If in any given period the underlying asset does default, then the seller pays the buyer the notional amount of the asset, the contract is concluded, and the buyer makes no further payments. CDS prices are quoted as the spread, the percentage of the notional amount that the buyer pays to the seller in each period as the “insurance premium.” The units are typically basis points (bps), and one bp is (1/100)%. CDS were originally used for corporate and sovereign debt instruments, but soon emerged in the secondary mortgage market as well.

A mortgage loan is a bond, and so a CDS contract can be created for it. First, many mortgages are packaged into investment vehicles for sale to banks, investment funds, etc. The resulting package is referred to as a mortgage backed security (MBS), which is an asset giving its owner the risks and returns (in this case, payments) of the underlying mortgages. In order to give different levels of risk and reward, Wall Street financial engineers came up with collateralized debt obligations (CDO), which were essentially MBSs with imposed structure. Investors did not buy the MBS but rather tranches of the MBS, which are best thought of as floors in a building. Buy a lower floor for a better returns, but when the building floods, you suffer the first losses. If you buy a higher floor, then the returns are lower, but the risk is lower because your floor won’t flood until the water level gets much higher. Here, flooding is defaulting. You lose money when a certain percentage of the underlying loans default. Thus a higher tranche corresponds to a higher portion of loans that must default before you lose your investment. A CDO on a MBS is referred to as a CMO, collateralized mortgage obligation.

The main purpose of this project is to examine how different misrepresentations and wrongful assumptions regarding a pool of mortgages can cause mispricings for CDS securities. Outright fraud was prevalent in the 2007-2009 financial crisis, but there are several more subtle ways that misrepresentation of a CMO can occur. In his book, The Big Short, financial journalist Michael Lewis outlines several of these methods. One that is of particular interest, which can be analyzed using statistical methodologies, is the effect of reporting only the average FICO in an MBS prospectus. FICO is the prominent credit report score that financial institutions use to evaluate the creditworthiness of a loan applicant. The idea is that since lower FICOs default at much higher rates than higher FICOs, and do so in a stronger than linear fashion, the variance of FICOs in the underlying pool of mortgages can be a form of misinformation. If the average FICO is 700, then a pool with 1/2 600 and 1/2 800 scores is much riskier than one where all FICO scores are 700. This project also extends this “average dilemma” to loan-to-value (LTV) statistics, another very crucial metric for determining a mortgage’s probability of default. A “semi-fraudulent” financial agent could construct pools of mortgages carefully so as to include many bad loans but still have average metrics that are adequate enough for the MBS to be given a good rating by agencies such as Moody’s and Standard and Poor’s. Additionally, parties valuating CDS contracts on these assets would misprice them due to this lack of full information. The final, and most critical, assumption examined in this project is that of independence of default events. Basic statistical inference typically assumes independent observations, so failing to be careful about this assumption when it does not hold can lead to drastic errors.

While examining the above is the end goal of this project, reaching this requires several intermediate objectives pertaining to general valuations of CDS securities, algorithms to combine individual probabilities into cumulative and total probabilities, and more. These will be discussed in detail in later sections.
2 Study Design and Data Collection

2.1 Simulation

This project seeks to simulate the effects of wrongful assumptions regarding the underlying mortgages in a CMO. The effect is measured in terms of the pricing on a CDS security on the underlying asset. The design for this simulation experiment is one of many layers.

This simulation tests the effects on CDS pricing for underlying CMO securities of deviations from certain assumptions. These assumptions can be derived from either inaccurate or incomplete information about the underlying mortgages. The assumptions, when considered, are fairly unrealistic (but frequent prior to 2007) and so are referred to as “naive assumptions.” Deviation from an assumption is the input in the simulation to be varied.

These are the naive assumptions to be examined:

- If the average FICO score in a mortgage pool is $\mu$, then the FICO score of each mortgage in that pool is $\mu$.
- If the average LTV for each mortgage in the pool is $\mu$, then the LTV of each mortgage in that pool is $\mu$.
- Defaults within a mortgage pool are events independent of one another.

Below is the full, ordered workflow of the simulation design. Further explanation and methodology for this workflow can be read in Section 3.2.

1. Translate an assumption (or deviation from it) into statistical parameters for input. For a pool of $n$ loans, create the default probability vectors for each loan under the assumptions.
2. Combine the default probabilities into a single default probability vector that represents the probability of default in the context of a CMO tranche.
3. Use the resulting single probability vector to solve for the spread price of a CDS contract on the CMO. Vary the inputs to examine how they affect the pricing.

A question in this simulation is essentially a test of a naive assumption. For the two assumptions regarding FICO and LTV, the input parameter is the underlying distribution of those metrics given $\mu$ and a range of possible values. In the case of FICO, the givens are $\mu = 680, [L, U] = [620, 740]$. In the case of LTV, the givens are $\mu = 80, [L, U] = [40, 120]$. The average values are chosen as such because they are extremely realistic for a pool of mortgages. The ranges are chosen such that they are symmetric about the mean and the bad end (120 for LTV and 620 for FICO) is the worst case for an originated mortgage provided the originator is neither completely incompetent or fraudulent (most responsible lenders today would not originate $> 100$ LTV or $< 680$ FICO, but the given ranges would not have seemed completely irresponsible prior to the 2007-09 financial crisis).

For the assumption of independence, the input parameter is $\kappa \geq 0$, which this project will refer to as the “contagion parameter.” For each pool of loans, the sample size (number of loans in that pool) is $n = 100$. This number is chosen because it is realistic. It is desirable to keep $n$ reasonably low since the method for modelling default for a CMO is computationally very taxing.

Most stages of the simulation workflow are deterministic, so therefore the replication sizes are 1 in those cases. Variation of inputs is the source of repeated trials. The default modelling for CMO tranches employs Monte Carlo simulation using a replication size of $K = 2000$, and is discussed in much more detail later in this paper.

This simulation design also allows the experimenter to begin later in the workflow. For example, one could assume certain default probabilities either of individual loans or for the CMO tranche and enter the workflow at step 2 or 3, respectively.
2.2 Data

Data is a minor aspect of this project, and is used only within the first stage of the simulation workflow. This stage requires data with which to build a model to predict default probabilities of a mortgage loan for given FICO and LTV metrics.

This project uses Fannie Mae’s Single Family Loan Level Acquisition and Performance data, which is publicly available to users who sign up for a free account. There are two distinct datasets, Acquisition and Performance. The Acquisition data includes loan characteristics such as FICO and LTV. The Performance data includes flags that can be used to determine if and when a loan defaulted. Using a unique loan identifier, the two datasets can be joined by loan.

After substantial data cleaning, the resulting file for analysis has 31,364 loans. One can model default using the characteristics, FICO and LTV. Whether or not a loan defaults is a binary response variable, and FICO and LTV are continuous predictor variables. FICO is a unitless measure ranging from 300 to 850. LTV is measured in percentage points.

A point to note is that this data is available for any quarter of origination dating back to 2000Q1. The only data used is the data from 2000Q1. There are two primary reasons for this selection. First, it is desirable for all loans in the data to be originated around the same time because this is the most realistic in the context of a securitized pool of mortgages. Therefore, only one dataset should be used. Since the data chosen has a sample size of more than 30,000, we do not expect to have issues of low sample size. Second, by choosing the earliest data set, loans have more opportunities to default. This is better for inference, and allows one to also take into consideration the age of the loan at default. Using data from 2013Q3, for example, would produce extremely biased results as many loans that will default have not yet done so.

3 Methods for Statistical Analysis

3.1 Estimation and Inference

From the loan level dataset, we wish to estimate default probabilities using FICO and LTV characteristics. For this a logistic regression model is used. Maximum Likelihood is the method to estimate the coefficients, which are the odds ratios for a unit increase in the corresponding parameters. Letting \( p \) be the probability of default (at any point over the life of the loan), the model is specified:

\[
\text{logit}(p) = \beta_0 + \beta_1(FICO) + \beta_2(LTV) \tag{1}
\]

It is important to discuss the assumptions of logistic regression in the context of this data:

- The true conditional probabilities are a logistic function of the independent variables: there is no theory to suggest one way or another the validity of this assumption. However, logistic regression models are commonly used in the mortgage industry, so this appears to be a safe assumption.

- No important variables are omitted: this assumption is almost certainly invalid. However, the model is limited by what data is publicly available (as opposed to confidential and/or proprietary data that lenders have). LTV and FICO are typically considered to be by far the most influential variables for this prediction, however, so the assumption should be relatively harmless. Furthermore, the goal of this model is to be illustrative and realistic, not to be used as a credit modelling tool.

- No extraneous variables are included: this is certainly valid, as LTV and FICO are often the first two characteristics lenders examine.

- The independent variables are measured without error: this is valid because LTV is determined exactly assuming accurate home appraisals (and the value of a home essentially is the appraised value) and FICO is observable directly in that it is reported by agencies.
• The observations are independent: as discussed elsewhere in this paper, this is often invalid. However, a later stage in the simulation workflow accounts for deviance from independence, so it is safe to make this assumption in this stage.

• The independent variables are not linear combinations of each other: this is definitely valid. FICO and LTV have no linear relationship.

Thus it is safe to use this regression model, keeping in mind a few imperfections. However, this model only predicts probabilities of default over the entire course of the loan, and we need the probability in each time period. Looking at the data and from literature, defaults occur more often in the early periods in a way that fits well into an exponential distribution. Therefore, we can extrapolate the scalar probability to a vector of exponential decay such that the total, cumulative probability of default along that vector is equal to the original scalar probability.

3.2 Theoretical Concepts

In order to answer the questions posed by this project, we need to derive several additional methods and formulas.

3.2.1 Valuation of a Credit Default Swap

There are many differing methodologies to valuate a credit default swap. This project uses a risk-neutral, arbitrage-free probability model. The assumption of risk neutrality may not be realistic in an applied context, considering that many buyers purchase a CDS to insure against risk (i.e. they are risk-averse). Risk-neutrality is however realistic in the application that both the buyer and seller are large financial institutions. For CDS on MBS, this is almost always the case. The probability model component requires that we can completely partition the sample space. This is true as this model partitions outcomes into no default and default for any period. The inputs are as follows:

• $T$: the number of periods. For this valuation approach, a period need not be a year, though throughout this paper periods are years.

• $\delta$: a vector of discount factors corresponding to the time periods. It’s length is $T$. This input should be adjusted so that the factors take into account the length of each period. For the purposes of this paper, each period is one year. The discount rate used to create this vector in the applications in this project is the Fannie Mae 30-year Mortgage Commitment for delivery within 30-60 days (FNMA 30-yr Mtg Com del 60 days). This rate is chosen specifically for the applications in this paper.

• $R$: the recovery rate. For simplicity’s sake, we assume $R = 0$. While this may not be the most practical assumption, it simplifies the problem at hand without reducing the usefulness of the results.

• $p$: a vector where the $t$th element is the probability of default in period $t$.

• While they are not inputs, the variables $c$ (the spread) and $N$ (the bond amount) need to be considered. This setup assumes $N = $10,000 so that the output can be given as $cN$ and represent the spread quoted in basis points.

With these inputs, we can calculate the present value (PV) of a CDS contract:

$$PV = p_1 N (1 - R) \delta_1 + \sum_{t=2}^{T} \left( p_t \left( \prod_{j=1}^{t-1} (1 - p_j) \right) \left( N (1 - R) \delta_t - cN \sum_{j=1}^{t-1} \delta_j \right) \right) - cN \left( \prod_{t=1}^{T} p_t \right) \left( \sum_{t=1}^{T} \delta_t \right)$$

(2)

The above formula is a probability weighted PV based on the PV of each occurrence and its corresponding probability. The way is interpret the above formula is to consider the right hand side of the equation term-by-term:
1. The first term is the PV when the bond defaults (and therefore the CDS contract pays out) in the first period, prior to the first premium payment, multiplied by the probability of this event occurring.

2. The second term corresponds to the bond defaulting in any period \( t > 1 \). The PVs are multiplied by the event probabilities and summed to give this term.

3. The third main term is the PV associated with the bond never defaulting. In this case, the buyer has paid the premium at each period and never received any payments in return. Obviously, this is the worst case for the buyer, and best case for the seller. The sign is negative since PV is calculated here from the perspective of the buyer.

The key now is to set \( PV = 0 \) and solve for \( c \). There is not an easy analytical solution to this, especially considering that the middle term is composed of an unspecified number of sub terms. Therefore, we look for a numerical method.

Observation reveals that while the PV function is not linear in all terms, it is linear in \( c \). This means that there is exactly one root with respect to \( c \). Given this, a simple bisection algorithm will yield a solution, when provided with an interval that contains the solution. Basic logic gives \( 0 \leq c \leq N \), so by starting with the interval \([0, N]\), the algorithm is guaranteed to produce the solution.

We now have a method to price a CDS from these inputs. For the purposes of this paper, all of the inputs save for \( p \) will be fixed. Thus, given the probability of default for each period over the life of the bond, we can solve for the spread of a CDS on that bond.

### 3.2.2 Modelling Default for a Collateralized Mortgage Obligation

Above we have determined how to solve for a CDS spread given a vector of default probabilities for each period. For a CDS on a single bond, we only need the default probability for that bond. However, MBS are typically pooled into CMO structures, which means our problem is more complex.

We must first define what exactly is a default in the context of a CMO. The answer is that an investor loses money when enough of the individual mortgages default to wipe out his tranche. For example, if an investor purchases the bottom 20% tranche, then a default from his perspective is when at least 20% of the underlying mortgages default. Thus is the definition of default in the context of a CMO, and it becomes clear what the vector \( p \) is in this case. It is the vector corresponding to the probabilities in each period that at least \( \alpha \% \) of the underlying mortgages have defaulted. Since a mortgage cannot “un-default,” we know that the vector must be increasing. More precisely:

\[
P_s \leq P_t \quad 1 \leq s \leq t \leq T
\]

\( P \) denotes the CMO default probability vector. Assuming \( n \) mortgage bonds in a CMO, we seek to solve for

\[
P = f(p_1, p_2, \ldots, p_n)
\]

To clarify, in this case each \( p_i \) is a vector with \( T \) elements. This is not a problem with a clear analytical solution. If all entries were equivalent (i.e. \( p_{it} = k \forall i,t \)), then one could imagine some solution derived using the binomial distribution. However, this is not the case in general, and for the applications this project explores, this certainly does not hold.

Given that analytical methods fail, we again search for numerical solutions. While computationally burdensome, a Monte Carlo approach will solve for \( P \). Here we present the algorithm to solve for \( P \):

1. Organize the individual vectors into an \( nxT \) matrix, \( \Lambda \).
2. Initialize a default counter vector, \( x \), with \( T \) elements. Specify \( \alpha \) as the percentage of defaults needed to wipe out a tranche.
3. Iterate through each column. Within each column, iterate along the column. At each element, generate a standard uniform random variable, \( r \). If \( r < \Lambda_{it} \), then that is a default. Otherwise, nothing happens.
4. Whenever a default occurs, set $x_t = x_t + 1$. Since a mortgage cannot default more than once, set the elements $\Lambda_{i,t+1} \ldots \Lambda_{i,T}$ to 0. When moving onto a new column, set $x_{t+1} = x_t$, to make the counter vector cumulative.

5. After having iterated through the entirety of $\Lambda$, create a vector, $a$, with length $T$. $a_t = 1$ if $x_t \geq \alpha n$ and 0 otherwise.

6. Repeat $K$ times and take the mean of $a$. As $K \to \infty$, $\bar{a} \to E[a]$, by the Law of Large Numbers. Since $E[a] = P$, $\bar{a} \to P$.

Thus we arrive at $P$. From repeated trials, it is observed that $K = 2000$ well balances the needs of small variance and reasonable computation time. Observation and trial also confirms that $P_t$ is increasing in $t$.

### 3.2.3 Relaxing the Independence Assumption

A critical, implicit assumption in the above method for creating the CMO default vector, $P$, is that mortgage defaults are independent events. This, however, is not necessarily true. In fact, hypothesized as one of the main contributors to the housing market crisis was that this assumption was invalid. Broad, macroeconomic factors can raise or lower the probabilities of default for all mortgage loans. The way we seek to account for this is by raising default probabilities in a pool when any of the mortgages in that pool default.

To that end, we seek a transformation of default probabilities. In the absence of data or theory to suggest a particular transformation, one will be imposed. This should be simple, realistic, and illustrative. Keeping with that, the transformation $f$ should have the following properties:

1. If $0 \leq p \leq 1$, then $0 \leq f(p) \leq 1$. This implies that a valid probability is mapped to another valid probability.

2. If $p \leq q$, then $f(p) \leq f(q)$. It is desirable that this transformation of the probability be monotonically (weakly) increasing.

3. In the event of a default, this transformation should be applied uniformly to the probabilities of default of the remaining mortgages. Additionally, as $n$ increases, the magnitude of this transformation should decrease.

Given these desirable properties, the following transformation is proposed. Given a pool of $n$ mortgages and a contagion parameter, $\kappa \geq 0$:

$$f_n(p) = \min(p(1 + \kappa)^\frac{1}{n}, 1)$$

Note that $\kappa = 0$ is the case when we assume independence. It is easily verified that $f_n$ meets all three of the desired properties. Instead of attempting to solve for an optimal or most accurate $\kappa$, the contagion parameter will be another simulation input for experimentation.

As a point of clarification, the case of $\kappa > 0$ is not a statement that person A defaulting on his mortgage causes person B to be more likely to default. This is merely a framework to accommodate the common empirical observations that surges in mortgage defaults can be heavily correlated. Again correlation does not imply causation. Rather, macroeconomic changes and other outside factors can increase probabilities of default all across the country. To merge this simulation design with real world applications, consider $\kappa$ to be some measure of outside factors that increase the likelihoods of default for all mortgages.

While there is logic and evidence to support the relevance of $\kappa > 0$, there is none to suggest cases where $\kappa < 0$ (that one default makes others less likely). Therefore, for the purposes of this study, we restrict examination to $\kappa \geq 0$. 

7
3.2.4 Generating Default Regressor Variables to Specified Distributions

A key point of interest for the simulations in this project is the underlying distributions of FICO and LTV in a pool of mortgages. In either the case of FICO or LTV, we assume several inputs:

- \( \mu \): the average mortgage owner’s FICO or LTV among the mortgages in a pool.
- \( L \): the lowest possible value for this metric, and \( U \): the highest possible value for this metric. The interval \([L, U]\) could be taken to be the entire possible range of these values (e.g. \([300, 850]\) in the case of FICO). More realistically, for the purposes of this project, these ranges will be restricted to more reasonable ranges. For example, 620 is the bottom end of the subprime FICO range, so a mortgage lender would only originate with a customer with a lower score than that if that lender was either incompetent or fraudulent.

The variance of these metrics within a pool of mortgages is the key point of consideration. We must construct distributions such that the minimum/maximum values correspond to the interval \([L, U]\) and the mean is \( \mu \). The case with the minimum variance is where all values are \( \mu \). The case with the maximum variance is where 50% of values are \( L \) and 50% are \( U \), with \( \mu = \frac{L+U}{2} \). We however additionally seek a framework to include those cases in between the extremes.

The Beta distribution provides a nice framework for analysis. The two parameters for the distribution are \( \alpha, \beta > 0 \), and \( E[X] = \frac{\alpha}{\alpha + \beta} \), \( Var(X) = \frac{\alpha \beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \) \( \Rightarrow \) the variance decreases as either of \( \alpha, \beta \) increases. Setting \( W = U - L \), we can calculate:

\[
\mu = \frac{\alpha}{\alpha + \beta} W + L
\]  

(4)

Taking \( \mu, L, U, W = U - L \) as given, and \( \alpha \) as a simulation input to be varied, we can solve for \( \beta \):

\[
\beta = \alpha \frac{W}{\mu - L} - 1
\]

(5)

Let \( Y \) be the default regressor variable of interest. We can now use the transformation, \( Y = XW + L \). If we further restrict \( \mu \) and \([L, U]\) to \( \mu = \frac{L+U}{2} \), then we can use \( \alpha = \beta \) as a simpler input.

This yields:

\[
E[Y] = E[X]W + L = \mu, Var(Y) = Var(XW + L) = W^2 Var(X) = W^2 \frac{1}{8\alpha + 4}
\]

(6)

\( \alpha = \beta \to 0 \Rightarrow Var(Y) \to \frac{W^2}{4} \). This is the maximum variance, and is the case where 50% of loans are at the lower limit and 50% are at the upper limit. In the limit as both parameters go to 0, the beta distribution converges to the binomial distribution. As \( \alpha = \beta \to \infty, Var(Y) \to 0 \). As the parameters tend to infinity, the beta distribution tends to a degenerate random variable, i.e. a constant.

With this in hand, we can use the givens and an inputted \( \alpha \) to vary the underlying distribution of loan characteristics in a mortgage pool.

4 Results

4.1 Logistic Regression Model for Default Probabilities

Running logistic regression to model default based on FICO and LTV, the following formula is estimated, with standard errors reported in the parenthetical subscripts:

\[
\text{logit}(\hat{p}) = 4.970_{(0.382)} + 0.007_{(0.002)} LTV - 0.013_{(0.001)} FICO
\]

(7)

Two-sided t-tests confirm that all estimated parameters are statistically significant. The positive coefficient for LTV and negative coefficient for FICO show that higher LTVs and lower FICOs are more likely to default. This is to be expected. Using this model, we can predict default probabilities for various FICO and LTV characteristics (see Figure 1).
Figure 1: (Left) Predicted default probabilities for different FICO scores, given LTV = 80%. Note that while the curve is very convex over the [500,800] range, it is not as convex in the [620,740] range. (Right) Predicted default probabilities for different LTVs, given FICO = 680. Note the curve over the entire range is less convex than the FICO curve, and is even less so over the [40,120] range. The red lines represent the intervals of interest for each characteristic, with the blue lines as the average values to be simulated.

4.2 Underlying Distributions of FICO and LTV

Here we simulate the resulting spread prices for CDS contracts as the distributions of FICO and LTV on the underlying mortgages vary. As previously discussed, the variance shifts as we vary $\alpha = \beta$ in the beta distribution.

Table 1 shows how the prices change under the different distributions. It is important to note that these pools are constructed in an order that is sorted. Specifically, the mortgage with the worst (lowest) FICO also has the worst (highest) LTV and so forth throughout the pool. It is not true in practice that FICO and LTV are perfectly correlated like this for mortgages. In fact, there is little correlation between them. However, an economic agent can selectively bundle mortgages into pools with these characteristics. Recall that these distributions are chosen, not random. As we alter the variance, we are not modelling a change in the population of mortgages, but rather in the extent to which an economic agent or financial institution is instilling variance within the pool. Given this, it is completely reasonable, and in fact desirable, to impose the perfect correlation between FICO and LTV within the pool.

Examining Table 1, we see that in both the FICO and LTV dimensions, higher variance yields higher prices.

<table>
<thead>
<tr>
<th>(alpha = beta)</th>
<th>LTV:</th>
<th>0</th>
<th>1</th>
<th>Inf</th>
</tr>
</thead>
<tbody>
<tr>
<td>FICO:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>3600</td>
<td>474</td>
<td>160</td>
<td>93</td>
</tr>
<tr>
<td>1</td>
<td>1200</td>
<td>84</td>
<td>44</td>
<td>39</td>
</tr>
<tr>
<td>Inf</td>
<td>0</td>
<td>30</td>
<td>29</td>
<td>26</td>
</tr>
</tbody>
</table>

Table 1: Calculated spread prices (quoted in bps) for CDS contracts on underlying mortgage pools given different distributions of FICO and LTV characteristics within the bundles. Along the rows are different FICO distributions, and along the columns are different LTV distributions. Variances shown are based on the $\alpha = \beta$ input. The three main cases shown are: $\{0, 1, \infty\}$, which correspond, respectively, to transformations of the binomial, uniform, and degenerate (constant) distributions. For these calculations, average FICO was 680 in the range [620,740] and average LTV was 80 in the range [40,120]. The underlying for the CDS is the bottom 10% tranche of a CMO.
4.3 Deviation from Independence

We also examine the effect of the magnitude of the contagion parameter, $\kappa$, on the CDS spread. Figure 2 shows the effect on the CDS price as $\kappa$ grows. We can use a linear regression model to measure the influence of $\kappa$: $\hat{c}_i = \beta_0 + \beta_1 g(\kappa_i) + \epsilon_i$. $g(\kappa_i) = \sqrt{\kappa_i}$ provides the best fit (see Equation 8, standard errors in parenthetical subscripts):

$$\hat{c} = 54.5(8.6) + 4.8(0.22)\sqrt{\kappa}, \ p-value = 2.4 \times 10^{-10}, R^2 = 0.98$$

A two-sided t-test with the resulting p-value reported above determines this model is significant. The high $R^2$ suggests an excellent fit. This suggests that the spread price tracks extremely closely with $\sqrt{\kappa}$.

![Figure 2](image)

**Figure 2:** (Left) The CDS spread in bps plotted against the contagion parameter, $\kappa \geq 0$. (Right) The spread plotted against $\sqrt{\kappa}$. The right plot has a great linear fit. For these calculations, the average FICO was 680, the average LTV was 80, and the degenerate/constant case was used for the underlying distributions. The underlying for the CDS is the bottom 10% tranche of a CMO.

5 Discussion

The results very much confirm initial expectations. The objectives of this study were to determine financial implications (as measured by the CDS spread price) of deviations from the aforementioned naive assumptions regarding the variances of FICO/LTV and independent default events.

Observing the results above, increases in $\text{Var}(\text{FICO}), \text{Var}(\text{LTV})$, and $\kappa$ (more precisely, $\sqrt{\kappa}$), definitely increase the valuations of CDS spreads on underlying pools of mortgages. To put a dollar amount on these findings, assume a financial institution (AIG, for example) sells CDS contracts at a 26bps spread, when the true valuation should be a 474bps spread. If the bond value is $100 million, which is not uncommon, the differential of 448bps results in AIG collecting $4.48 million fewer per year in premium payments than they should for that contract.

There are limitations to this study. It is difficult to apply these findings in practice for the same reason they are analyzed in the first place: reports of loan characteristics for an MBS are summaries as opposed to detailed accounts of each loan booked. Additionally, the credit model to determine defaults could be more sophisticated if more detailed data were available to the public. Still, if the prospectuses for MBS securities do not also have this additional detail, then it is not helpful.

There are two particularly interesting areas of further exploration. First, it would be informative to incorporate the yields of the underlying bonds in question into this analysis. This could allow for more complex trading strategies to develop as a result. Second, it is very interesting that the CDS spread price fits so well with the square root of the contagion parameter. As $K \to \infty$, the Monte Carlo simulation to create a single probability vector will converge to a constant. This means that the spread, while a complex and compound function of the contagion parameter, is deterministic in the limit (i.e. $R^2 \to 1$ for a correctly specified regression model). The linear model built strongly suggests that the spread is a deterministic, linear...
function of $\sqrt{\kappa}$ in the limit. An analytical result to explain why this is the case would be extremely informative and useful, and thus is an excellent area for further research.

In conclusion, the findings support the original hypothesis mentioned by Michael Lewis in The Big Short, that the variance of FICO scores in the underlying pool of mortgages creates a mispricing of the CDS spread. The results additionally affirm the other two hypotheses examined: that a similar effect happens with LTV variance, and that deviation from independence also skews the pricing. Thus, we can conclude that these three factors indeed played large roles in the financial crisis. Of course, there were other influences, such as blatant fraud, at work. Nevertheless, this analysis explains a sizeable portion of the causes within the housing market.
References

