DEFAULT PREDICTION FOR COMMERCIAL MORTGAGE BACKED SECURITIES

by

JAMES SCOTT DUDLEY

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE IN REAL ESTATE

THE UNIVERSITY OF TEXAS AT ARLINGTON

August 2005

ACKNOWLEDGEMENTS

I would like to thank Mr. William Hosey for his guidance and patience on this project and many others. Without his help many of my accomplishments over the past year would not have been possible.

Thank you to Dr. Matt Cypher and Dr. Fred Forgey for their help in completing this massive undertaking.

Finally, thank you to my family and friends for supporting me and not forgetting about me through this last year of stress and fatigue.

February 16, 2005

ABSTRACT

DEFAULT PREDICTION FOR COMMERCIAL

MORTGAGE BACKED SECURITIES

Publication No. ______

James Scott Dudley, M.S.

The University of Texas at Arlington, 2005

Supervising Professor: Fred Allen Forgey

Commercial mortgage default has become a topic of interest for a large number of parties due to the emergence and continued growth of the secondary mortgage market. With the multitude of parties holding a vested interest, it is important to develop a highly efficient method of monitoring collateral performance and ultimately be able to confidently predict or anticipate default.

This study shows the correlation between appearance of a loan on a Watchlist and its potential to become delinquent in the future. While testing this hypothesis, a model is created that incorporates several other variables readily used to predict

collateral performance for commercial mortgages and specifically commercial mortgage backed securities.

Implementing the use of logistic regression, two models are created to show the level of correlation and significance with delinquency. Also, a model with a high level of explanatory power from a selected group of variables is created. The results are provided and analytical commentary on their impact is discussed in detail.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION

1.1 Statement and Definition of Problem

The focus of this research is to establish whether a correlation exists between individual Commercial Mortgage Securities Association (CMSA) Watchlist items (defined below) and commercial mortgage default. In addition to testing Watchlist items for correlation, this study will also measure and evaluate other variables shown in prior studies to be highly related to default probability. The topic of mortgage default has been thoroughly researched taking into consideration a wide spectrum of components, scenarios, and methods resulting in varying results and conclusions. For example, Vandell, Barnes, Hartzell, Kraft, and Wendt (1993) confirmed the equity effect (impact of the amount of borrower equity, using loan-to-value ratios as a proxy, on default probability) as proxied by contemporaneous market loan-to-value ratios in influencing default among commercial mortgages. Archer, Elmer, Harrison, and Ling (2002), on the other hand, found that the original Loan to Value ration or "LTV" does not contribute significantly to the explanatory power of the regression model. In contrast, the expected inverse relation between incidence of default and the Debt Coverage Ratio or "DCR" was found to be strong. Property characteristics generally do not exhibit significant individual effects, with the exception of year built, which is strongly inversely related to default. Changes in loan defaults follow changes in the flow of commercial mortgage credit (Mejia 1999). Riddiough and Thompson (1993) established that the characteristics of the property and mortgage contract design, versus the characteristics of the interest-rate environment, have the greatest impact on default. (Mortgage contract design, especially those pertaining to loan and rate of the mortgage amortization, are of particular interest since the lender easily controls these factors.) Harrison, Noordewier, and Yavas (2004) found that while conventional wisdom within the mortgage underwriting industry posits that loan-to-value ratios should be positively associated with the probability of default for a given loan, previous empirical evidence has not uniformly supported this contention.

The purpose of the CMSA servicer Watchlist is to provide commercial mortgage backed securities (CMBS) investors, rating agencies, and special servicers with information pertaining to the credit risk associated with commercial mortgage loan portfolios. In establishing its standards the CMSA has attempted to create a uniform system for predicting future default and transfer events (transfer of problematic loans to the more flexible special servicer) based on items deemed to be correlated to monetary default and/or potential transfer to special servicing. In conforming to the CMSA standards, a servicer monitors specific triggers related to one of the following six categories: (1) financial conditions; (2) borrower issues; (3) property condition issues; (4) lease rollover, tenant issues, and vacancy; (5) loan maturity, and (6) other servicer discretionary items. Though all of the variables seem to be logically related to default probability, relatively new uniform standards created for Watchlist by CMSA and the fairly new nature of the CMBS industry have prevented the thorough testing of the statistical relevance of those variables as they relate to CMBS. Servicers (master, primary, and sub) need to be knowledgeable of the correlation between Watchlist items and default so that more focused information can be disseminated to investors, rating agencies, and special servicers. At the same time, investors, rating agencies, and special servicers also need to be aware of this correlation and the likelihood that prescribed circumstances will lead to default so they can request detailed and statistically meaningful information from the servicers and, therefore, increase efficiency by avoiding having to sift through superfluous information.

1.2 Research Objectives

Three primary objectives have been established to address this problem. This study seeks to:

1. Assemble a literature review focused on the impact of certain property performance components on commercial mortgage default.
2. Construct a logistic regression model to isolate and measure the correlation of specific CMSA Watchlist items, other default prevalent variables, and commercial mortgage default.
3. Produce a research statement that evaluates the relevance of CMSA guidelines and other variables in predicting default.

1.3 Significance of Research

General variables that affect default probability have been thoroughly researched and well documented for all property types over many different chronological periods.

The commercial mortgage backed securities market, however, has a relatively short history and some industry specific correlations have not been as thoroughly explored. In an effort to develop servicing standards for the CMBS industry, CMSA has created guidelines that are regularly updated, most recently in January of 2005, and refined for many servicing practices. These guidelines are meant to hold master servicers to a certain uniform standard of reporting, as to loan level and collateral level information related to CMBS securities, to provide investors, rating agencies, and special servicers with timely and meaningful information. As the demand for information from rating agencies, special servicers, and investors grows, additional resources are needed to keep up and servicers are requiring additional employees and often expensive market research tools to meet the standards set out by the industry. With the recent issuance volume growth of the commercial mortgage backed securities market and the highly competitive bidding for servicing rights resulting in narrowing profit margins, it is highly likely that in the near future some of the tasks being made industry standard may become cost prohibitive. This occurrence will make it essential that servicers focus on the most useful and important information, while abandoning other time intensive items that offer little additional benefit.

1.4 Data Source for Research

Data on specific assets and their performance related to loan level specifics, default, and Watchlist inclusion was sought from 11 currently issued CMBS pools including 1,310 loans with data spanning in vintage from 1997 to 2002. This data included historical Watchlists with related Watchlist codes, comments, and definitions.

Historical Watchlists provide a record of poor collateral performance that can be used to trace the beginnings and changes in performance for collateral deemed to be at risk of default. Other loan level information was obtained from loan level periodic files and setup files created to report data to investment trusts established by each mortgage pool's pooling and servicing agreement (PSA). These files are remitted to the Trust on a monthly basis, however, this study pulled data from the January 2005 remittance files for each respective mortgage pool.

1.5 Research Hypothesis

This research investigates the correlation between CMSA Watchlist codes, other variables of interest, and commercial mortgage defaults. A total of eight hypotheses will be tested to determine the level of correlation.

1.5.1 a. Hypothesis (Null) #1: Inclusion on the Watchlist will increase the commercial mortgage default probability.

b. Hypothesis (Alternative) #1: Inclusion on the Watchlist will not increase the commercial mortgage default probability.

1.5.2 a. Hypothesis (Null) #2: The loan-to-value level at origination is not positively correlated with commercial mortgage default.

b. Hypothesis (Alternative) #2: The loan-to-value level at origination is positively correlated with commercial mortgage default probability.

1.5.3 a. Hypothesis (Null) #3: The debt service coverage ratio at origination is not positively correlated with commercial mortgage default.

b. Hypothesis (Alternative) #3: The debt service coverage ratio at origination is positively correlated with commercial mortgage default.

1.5.4 a. Hypothesis (Null) #4: The occupancy level at origination is not positively correlated with commercial mortgage default.

b. Hypothesis (Alternative) #4: The occupancy level at origination is positively correlated with commercial mortgage default.

1.5.5 a. Hypothesis (Null) #5: Property type is not positively correlated with commercial mortgage default.

b. Hypothesis (Alternative) #5: Property type is positively correlated with commercial mortgage default.

1.5.6 a. Hypothesis (Null) #6: Loan size is not positively correlated with commercial mortgage default.

b. Hypothesis (Alternative) #6: Loan size is positively correlated with commercial mortgage default.

1.5.7 a. Hypothesis (Null) #7: Market size is not positively correlated with commercial mortgage default.

b. Hypothesis (Alternative) #7: Market size is positively correlated with commercial mortgage default.

1.5.8 a. Hypothesis (Null) #8: The state foreclosure process will not impact commercial mortgage default probability.

b. Hypothesis (Alternative) #8: The state foreclosure process will impact commercial mortgage default probability.

1.6 Overview of Research Methodology

The primary research methodology for this project is logistic regression. The selection of this methodology resulted from a desire to isolate the default and Watchlist variables so that a correlation measurement between the two can be taken. Many categorical response variables have only two values, for example: yes/no, good/bad, pass/fail, etc. In models having these characteristics for the dependent variable it is suitable to use logistic regression. Logistic regression allows one to predict a discrete outcome, such as group membership, from a set of variables that may be continuous, discrete, dichotomous, or a mix of any of these. There are two main uses of logistic regression. First is the prediction of group membership. Since logistic regression calculates the probability of success over the probability of failure, the results of the analysis are in the form of an odds ratio. For example, logistic regression is often used in epidemiological studies where the result of the analysis is the probability of developing cancer after controlling the other associated risks. Additionally, logistic regression provides knowledge of the relationships and strengths among the variables (Connor 2005).

The logistic function was invented in the 19th century for the description of the growth of populations and the course of autocatalytic chemical reactions (Cramer 2002). Logistic regression has since been used in a wide range of fields to predict probabilities of occurrence.

For this study, one model will be created with the dependent variable of monetary default. Occurrences of default will be measured against the presence of that loan on the Watchlist prior to default and specifically which code or criteria caused its appearance on the Watchlist.

Taking into consideration the implications on the results, a quantitative variable was created by subtracting the date a loan was added to the Watchlist from the last distribution date. This variable will be representative of the seasoning involved prior to Watchlist loans becoming defaulted. The use of independent variables for loan-to-value at origination, loan balance, property type, debt service coverage at origination, state foreclosure proceedings, market size, and occupancy at origination will also be included in order to increase the explanatory power of the model and provide insight to the influence of these variables on default probability. The data will then be entered into SPSS and analyzed for pertinent correlations.

1.7 Limitations

The number of Watchlist observations that ultimately defaulted was limited, however the percentage of Watchlist and defaulted loans was consistent with historic levels exhibited by the CMBS market. The implications of this low level of occurrences create a study in which correlations are based on a relatively small number of defaults. Further limitations are caused because of the development of the CMBS industry itself. All of the data collected was prior to the new CMSA standards being created in April of 2003, which significantly increased the number of Watchlist loans and Watchlist criteria. The implications of this limitation are that results from the study may not be representative of the predictive power of current Watchlists. More recent data was not used due to the lack of seasoning present in pool vintages created after

April 2003. Other variables used in the study should not be affected like the Watchlist criteria because these variables are equally representative across all time periods.

CHAPTER II

LITERATURE REVIEW

2.1 Commercial Mortgage Default

Commercial mortgage default in the CMBS market can take two major forms, non-monetary and monetary. The focus of this study will be on monetary default, which relates the non-payment of principal, interest, and escrow by the borrower prior to their defined payment date. Monetary default is usually associated with a cash flow shortage. Properties can run into a cash flow shortage in a variety of ways which include, but are not limited to: bankruptcy, waste to the property, tenant rollover, loss of franchise, loss of license, increase in expenses, decline in overall economic conditions, and erosion of market conditions.

Investors in CMBS attempt to mitigate these credit risk situations by requiring originators to provide securitizations with diversification of property type, geography, and loan size. However, because many of these situations cannot be predicted from securitization, it is important for investors to have a system in place for monitoring the underlying collateral so that workout and buy/sell strategies can be administered.

Parkus (2002) conducts and analyzes a default study on 35,000 CMBS loans in order to attempt to find the major factors of default. The study finds that default cannot be estimated by any single variable, but must be evaluated through the combination of variables. This study found the eight most important default factors to be (1) relative mortgage rate spreads; (2) relative cap rate spreads; (3) tenant concentration; (4) real estate market size; (5) cross default provisions; (6) state foreclosure laws; (7) loan vintage; and (8) property type. The relative mortgage rate spread is calculated as the difference between the mortgage rate on the loan and the average mortgage rate on similar loans originated during the same month. The idea behind the relative mortgage rate spread is that a lower than average spread indicates that the lender deemed the loan to be less risky than average, and a higher than average spread indicates that the lender found the loan more risky than the average. The relative mortgage rate spread can be used as a proxy for the lender's view of the riskiness of the loan at origination.

The cap rate spread is calculated the same as the relative mortgage rate spread but instead using capitalization rates. The cap rate spread is calculated by taking the difference between the cap rate of a loan and the average of the cap rates of similar loans originated during the same month. The cap rate spread is a proxy for the appraiser's view of riskiness of the property's cash flow relative to similar properties. The same relationship to default probability exists as with the relative mortgage rate spread with higher than average spreads increasing default probability and lower than average spreads decreasing default probability.

A major risk to a property's cash flow is the number of tenants and type of tenants a project has. A project with a large number of similar size (by occupied square footage) tenants that is diversified from tenant rollover as opposed to a single tenant property is at a far lower risk of default. This variable rates tenant concentration and finds that increased tenant concentration equates to a higher default probability.

Real estate market size alludes to property location relative to the surrounding market size. This study showed that default risk was enhanced if a property was in a small market rather than a major market. The findings were the same across all property types.

This study also found that cross-defaulted loan structures exhibited much higher default rates than non-crossed loans. The author found the reason for this to be two-fold. First, cross-defaulted loans tend to be viewed by originators as being particularly risky. Second is that the results may be driven by contrived defaults, in other words instances where borrowers let loans go delinquent in hopes that the lender will accelerate the loan.

State foreclosure laws can also substantially impact default rates for loans. This can be seen relative to the two differing philosophies on state foreclosure proceedings. The two foreclosure procedures are known as judicial and non-judicial foreclosure. The impact of each on default rates corresponds to the amount of time it takes to resolve a foreclosure under each philosophy. Judicial foreclosure states incur higher rates of default due to the lengthy foreclosure process and the borrower's awareness of rights of redemption and timing. Non-judicial foreclosure states are much more lender friendly and do not incur the same amount of defaults because foreclosures can be performed much more quickly.

Loan vintage refers to the risk associated with the year in which a loan was originated. As underwriting practices have changed and volume of origination has fluctuated the quality of the overall origination varies from year to year. The author has noted that some years have a much stronger correlation with default due to the practices and volume at the time of origination.

The final major factor included by the author in influencing default was property sector. Each of the major property sectors has inherent factors that influence their ability to perform. In extracting these risk factors and evaluating property types as a whole, it is evident that some types are riskier than others.

Rutherford (2004) performs a study to analyze the correlation between the relative mortgage spread and default risk. This study used a linear regression model with variables for the original mortgage rate, property type, treasury rate, original amortization, metropolitan statistical area or "MSA" dummy variable, and cutoff capitalization rate. This study incorporated data from 4,542 loans in 22 CMBS contracts. Of the data used 4.5 percent of the loans were defaulted mortgages. This study found that as relative mortgage rate spreads increased so did the default probability. The implications of this study can be seen in the need to address the relative mortgage spread as a default risk indicator. Using this method of analysis provides insight into the originators perception of the borrower and can identify loans likely to be of above average default risk.

Gordon and Kizer (2004) analyze the aggregate performance of commercial mortgage backed securities and examine variables that increase or decrease expected loss through probability of default. This study provides the following findings. Generally, delinquency increases as loans enter into their middle age and the default probability is once again decreased typically starting in the later years stated as years six to seven. Healthcare and lodging properties show increasing default over time and do not exhibit the same settling that is often present among other property types. Another finding of the study that refers to the modeling of the default cycle stated that only the retail sector followed a linear default trajectory, and that all other core property types (office, multifamily, and industrial) follow a hill-shaped pattern with delinquency falling of in the later years. The underwritten debt coverage ratio can be misleading in estimating loan default risk when not accompanied by property type. Typically, however, the debt coverage ratio will be higher for properties with a higher perceived risk, therefore default rates are often higher for properties with higher Debt Service Coverage Ratios or "DSCRs". Higher loan to value ratio loans tend to experience two to three times the delinquency rate of low loan to value ratio loans. Loans on properties in smaller metropolitan statistical areas typically under-perform and have increased default rates from those of larger areas. Additionally, properties in areas that do not have the population to be considered an MSA further under-perform and pose a higher default risk than their MSA counterparts.

Titus, Betancourt, Grossman, and Kozel (2003) researched 315 liquidated loans with a balance of $987 million from 118 CMBS transactions and provide a review of the characteristics found. The authors' findings are as follows: The subject loans had an average cumulative loss severity of 40.11 percent over the 9 year period studied.

The number of liquidated loans increased year over year, most significantly from 2001 to 2002. Market perceptions about the relative risk characteristics of the different property types were confirmed. The correlation between resolution time and loss severity were found to be statistically significant. The average time from default to liquidation was 19 months. Liquidated multifamily projects accounted for the lowest loss severity of the property types at 22 percent, and healthcare properties were the worst with an average loss severity of 66.5 percent. Of all 50 states, Kansas was the worst performer relative to loss severity averaging 109.32 percent, and Hawaii was the top performer averaging a loss severity of 4.38 percent. By year of origination, 1995 was at the bottom with a loss severity of 53.8 percent and 1992 was the top performer with a loss severity of 26.7 percent. The degree of loss severity was also highly correlated with the condition of the property markets. Loss severities have shown that they increase with the deterioration of property market fundamentals. Finally, loss severities vary depending on which special servicer performs the workouts. Loss severities range between 21 and 62 percent for special servicers throughout the industry.

Berger and Sternin (2003) describe the functions of Watchlists and how they are used to monitor credit performance. According to the authors, Watchlists have historically been used as an internal mechanism to monitor the credit risk of commercial mortgage portfolios. Watchlists were developed to evaluate both qualitative and quantitative criteria that are believed to be indicators of a loan's credit risk for internal use. Common inclusions are measures of the debt coverage ratio, loan to value ratio, occupancy percentage, net operating income, lease rollover, and months to loan maturity. Because clear benefit to special servicers, rating agencies, and investors of having an early warning mechanism for CMBS transactions, servicer Watchlists have been expanded from an internal risk management tool to an investor reporting device. Watchlists provide more specific information than the standard delinquency report and are also more comprehensive. Watchlists include loans that have significant future events that may increase their credit risk profile, such as pending maturity or major lease rollover.

Vandell, Barnes, Hartzell, Kraft, and Wendt (1993) discuss commercial mortgage default theory related to loan term, borrower, property, and economic/market conditions. The authors use a data set of 2,899 loan histories provided by an insurance company to analyze default relationships through the proportional hazards estimation technique. Results from the study found the equity effect or loan-to-value ratio to be the most significant indicator of default probability, and found cash flow and borrower effects to be significant but less important.

Vandell (1992) evaluated the relationship of commercial mortgage default incidence and characteristics such as borrower, property, market, and general economic conditions. This study showed a high proportion of variation in default incidence over time could be explained by variations in the market value of the loan-to-property value ratio.

Archer, Elmer, Harrison, and Ling (2002) explore the default experience of 495 fixed rate multifamily mortgage loans securitized by the Resolution Trust Corporation and the Federal Deposit Insurance Corporation from 1991 to 1996. This study finds that the loan-to-value ratio is not highly correlated to default incidence, but that the strongest relationship with default comes from property characteristics, specifically ZIP code location, initial cash flow as reflected in the debt coverage ratio, and post origination changes in the local economy.

Mejia (1999) assessed the presence of a credit supply effect in the commercial mortgage market. Mejia defines a credit supply effect as the effect of mortgage supply on the level of loan defaults. The author finds that changes in loan defaults are often preceded by changes in commercial mortgage supply with a lag time of four to five years. This leads the borrower to conclude that in addition to the typical loan underwriting criteria of borrower and property characteristics that underwriters should also consider the effect of flow of capital on the riskiness of the investments.

Riddiough and Thompson (1993) examine the influence of default transaction costs on a borrowers decision to follow through with their default decision. The author argues that these costs are different from borrower to borrower and hidden from the lender/investor at the time of origination. Results from the study show that characteristics of the property and loan structure have the greatest impact on default risk as opposed to characteristics of the interest rate environment.

Elmer (1992) uses cross-sectional time series data with an options based mortgage default model to simulate the default and loss characteristics of price level-adjusted mortgages and standard fixed-payment mortgages. Results from the study show that default risk is two to seven times higher for price level-adjusted mortgages than for fixed-payment mortgages.

Ciochetti, Dang, Goa, and Yao (2002) examine the factors driving a borrower's decision to terminate commercial mortgage contracts with the lender through either prepayment or default. Findings suggest that values of put and call options drive default and prepayment actions in a nonlinear fashion. The study also finds that variables used to proxy cash flow and credit conditions have a significant influence upon the borrower's termination decision.

Ciochetti and Vandell (1999) researched the characteristics of return for well-diversified commercial mortgage portfolios. This study found that mortgage returns and volatility for mortgages are comparable to other forms of fixed-income assets. Also, returns were found to vary by property type and region of origin.

Ambrose, Benjamin, and Chinloy (1996) developed a model of the market for commercial real estate loans based on variables used by investors and lenders in their decision-making. The variables chosen by the authors included: the income capitalization rate, the debt coverage ratio, and the loan-to-value ratio. This study found that underwriting is debt-coverage ratio heavy focusing on the income statement, as opposed to focusing on balance sheet items such as the loan-to-value ratio. The authors ultimately found that lending institutions and investors focus on cash flow rather than net equity in constraining lending for commercial property.

Pyhrr, Born, Robinson, and Lucas (1996) evaluate linkages in the market between urban economic cycle, real estate supply, and demand cycles, construction cycles, and property life cycles. Findings show that these cycles can have a significant impact on cash flow variables and cause dramatic alterations to market value estimates.

The authors suggest that economic and market cycles should be addressed in a valuation model.

Vandell and Thibodeau (1985) create a formal theoretical model that non-loan related influences on the default decision. The non-loan variables deemed to be of concern include payment levels relative to income, current and expected neighborhood market conditions, economic conditions, wealth, borrower characteristics proxying for variability in income or crisis events, and transactions costs incurred upon default. The model carried out by the authors found that these other factors all play a role in default risk.

Vandell (1985) develops an empirically based model to assess default risk in the commercial mortgage market. This study finds that an appropriate default model must include information about the property, its market, and broader economic conditions. To be useful the default model must be specified using loan term variables that are generalizable to other mortgage instruments in the commercial market. The model also should be estimated in a multi-period framework to account for differences in the effects of default over time. Finally, default model specification for the residential mortgage and commercial loan markets is very different from that for income properties, because of different structures of the investment, economic situations, and motivations for purchase and continued holding.

Chun and Goldberg (2003) address the issue of “imminent default” in the CMBS market. Imminent default is common language in pooling and servicing agreements throughout the CMBS industry. This language gives the servicer the judgment call of whether or not outstanding circumstances underway with the borrower pose an imminent threat of monetary default. The authors find that typically there is agreement between the servicing parties involved when a transfer for imminent default occurs, but warns that with depressed economic conditions there is likely to be more of a gray area in regards to the imminent default language.

Lans and Price (2001) found CMBS bond default rates to be significantly lower than that of the corporate bond market. The authors found that for the period of 1990-1991 non-investment grade CMBS default averaged .14 percent compared to 3.07 percent for corporate bonds. This difference was attributed to the diversification of CMBS transactions not seen in the commercial bond sector. In the CMBS market, multiple assets, multiple borrowing entities, and multiple tenants back most transactions, which provides significant diversification of risk.

Maris and Segal (1999) found that yield spreads for CMBS bonds have narrowed as the industry has developed. The yield on a particular security is determined by a variety of factors, including the overall level of interest rates in the economy, the security's term to maturity, the probability of default, and tax considerations. According to the authors there is considerable evidence that lenders have eased their underwriting standards for commercial real estate loans, which means that investors are earning a lower yield spread on securities that might have greater default risk.

Saunders (2003) discusses due diligence underwriting, which is focused on collateral credit risk: determining the competitive position of an asset within its market, assessing the quantity and quality of the collateral's cash flow, and determining an asset's ability to generate sufficient cash flow over time to service the mortgage debt. Due diligence underwriting evolved from B-piece investors who sought to fully understand the collateral credit risks they were taking by investing in CMBS. Now investors collect, inspect, analyze, and report on the collateral prior to purchase.

2.2 Logistic Regression

Garson (2001) discusses the key terms, concepts, and output related to logistic regression. The author states that logistic regression is a form of regression that is used when the dependent is a dichotomy and the independents are of any type. Logistic regression can be used to predict a dependent variable on the basis of independents and to determine the percent of variance in the dependent variable explained by the independents; to rank the relative importance of independents; to assess interaction effects; and to understand the impact of covariate control variables. According to Garson, logistic regression does not assume linearity of relationship between the independent variables and the dependent, does not require normally distributed variables, does not assume homoscedasticity, and in general has less stringent requirements.

Wuensch (2004) describes logistic regression and specifically details the steps need to complete a logistic regression model in SPSS. An example of the steps and output of an actual model are exhibited for a psychological research study. The author states that logistic regression is used to predict a categorical, usually dichotomous variable from a set of predictor variables. With a categorical dependent variable, discriminate function analysis is usually employed if all of the predictors are continuous and nicely distributed; logit analysis is usually employed if all of the predictors are categorical; and logistic regression is often chosen if the predictor variables are a mix of continuous and categorical variables and/or if they are not nicely distributed (logistic regression makes no assumptions about the distributions of the predictor variables).

Pampel (2000) presents a detailed overview of the component parts of logistic regression modeling. This material introduces the relationships of probabilities, odds, and logits discussing in detail the use of each and the differing characteristics of all three. The author also details the interpretation of logistic regression coefficients, estimation and model fit analysis, and probit analysis.

CHAPTER III

RESEARCH METHODOLOGY

3.1 Methodology

This research investigates the correlation between CMSA Watchlist items, other selected items, and commercial mortgage default through delinquency. Specifically, commercial default is being tested for correlation with presence on the Watchlist, debt service coverage ratio at origination, the difference between the date added to the Watchlist and the delinquency date, the loan amount at origination, the loan-to-value ratio at origination, the occupancy level at origination, the loan vintage, the property type, the market size, and the type of foreclosure used in the subject state. This study seeks to correlate these variables with the likelihood that a commercial mortgage will become delinquent in the future. The study employs logistic regression through the use of SPSS in order to measure and analyze these correlations.

3.2 Data: Source and Treatment

The data set consists of a sample of 1,310 loans currently in CMBS pools of varying vintage. The subject loans are backed by collateral property located throughout the United States and consisting of various property types. Upon origination these loans were placed into commercial mortgage pools and sold on the secondary market to various CMBS investors. CMBS investors typically include pension funds, insurance companies, public institutional, and private institutional investors. CMBS bonds are broken down into different traunches. These traunches are sold at different pricing levels and carry varying levels of risk. After securitization has been completed, firms bid on the servicing rights for each mortgage pool. The servicer handles the day-to-day maintenance of each loan in the pool through cash management, surveillance, and other tasks generating reports to be provided to the investing bondholders and the trust. The servicer completes these tasks in return for a fee and the right to earn float interest on many accounts where funds are held for the borrowers. In the case of the loans represented in this study the servicing rights were then acquired by General Motors Acceptance Corporation Commercial Mortgage or "GMACCM", which in turn reports on the collateral and performance of the loans to each of the trusts for the various mortgage pools.

The data collected for this study was taken during the January 2005 reporting cycle for various CMBS pools from the GMACCM website. Specifically, data was collected on the subject loans from the setup and loan periodic files on the GMACCM website, which provide large amounts of loan level information for time periods dating back to loan origination. The GMACCM website allows for data to be downloaded into an Excel spreadsheet. After downloading the data from the website it was sorted to check its integrity. The original loan count was at 1,728, but all 1996 vintage loans had to be disregarded due to the lack of Watchlist information. The absence of this information is due to the relative infancy of the CMBS market at this stage, which caused surveillance to be meager at best and resulted in lapses in certain types of data.

After the deletion of the 1996 vintage pools, a preliminary analysis was done on the data through sorting and calculating the delinquency percentage of the total, which equated to 3.36 percent. This number was then compared to other studies on CMBS default that had been completed to determine if this was a reasonable default rate. After concluding that this percentage was reasonable in comparison to previous studies the data was coded into dummy variables that would be entered into the regression. Dummy variables are used to turn qualitative variables into quantitative variables by creating a yes/no variable out of each qualitative occurrence through the use of 1's and 0's. In this study all of the variables with the exception of Watchlist difference (WLDiff) were of a qualitative nature and required coding as dummy variables. Table 3.1 provides a list of variables originally created for inclusion in the model.

Table 3.1 Variables

Variable	Abbreviation	Description
Current	Current	Is the loan current on mortgage payments
Delinquent	Delinquent	Is the loan delinquent on mortgage payments
Watchlist	Watchlist	Is the loan on the Watchlist
Watchlist difference	WLDIFF	Difference in days between delinquency and Watchlist addition
Loan greater than 20 million	LoanGrt20Mill	Loans larger than 20 million
Loan between 10 and 20 mi	LoanAmt10to20	Loans between 10 and 20 million
Loan between 5 and 10 mil	Loan Amt5to10	Loans between 5 and 10 million
Loan less than 5 million	LoanAmtLess5	Loans less than 5 million
Loan to value greater than.85	LTVGrt.85	LTV greater than .85
Loan to value between .70 and .85	LTV.70to.85	LTV from .70 to .85
Loan to value between .55 and .70	LTV.55to.70	LTV from .55 to .70
Loan to value less than .55	LTVLess.55	LTV less than .55
Debt service coverage greater than 2	DSCRGrt2	DSCR Greater than 2
Debt service coverage between 1.50 and 2	DSCR1.50to2	DSCR from 1.50 to 2

Table 3.1 - continued

Variable	Abbreviation	Description
Debt service coverage less than 1.50	DSCRLess1.50	DSCR less than 1.50
Occupancy greater than 90 percent	OCCGrt90	Occupancy greater than 90 percent
Occupancy between 80 and 90 percent	OCC80to90	Occupancy from 80 to 90 percent
Occupancy less than 80 percent	OCCLess80	Occupancy less than 80 percent
Vintage 1996	Vintage96	Loan Vintage 1996
Vintage 1997	Vintage97	Loan Vintage 1997
Vintage 1998	Vintage98	Loan Vintage 1998
Vintage 1999	Vintage99	Loan Vintage 1999
Vintage 2000	Vintage00	Loan Vintage 2000
Vintage 2001	Vintage01	Loan Vintage 2001
Vintage 2002	Vintage02	Loan Vintage 2002
Retail Property	Retail	Retail Property Type
Office Property	Office	Office Property Type
Multifamily Property	Multifamily	Multifamily Property Type
Lodging Property	Lodging	Lodging Property Type
Other Property type	Other	Other Property Type
Small Market	Smallmarket	Market Not Listed in REIS top 50
Large Market	Largemarket	Listed in REIS top 50
Judicial foreclosure state	JudicialForcl	Located in a state that practices judicial foreclosure
Non-judicial foreclosure state	NonJudicialForcl	Located in a state that practices non-judicial foreclosure

Some of the original variables were altered or combined

Upon completion of the dummy variable coding the variables were copied from Excel into SPSS for evaluation. The first step taken using SPSS was to run the descriptive statistics for the variables. A breakdown of the original descriptive statistics can be viewed below in Table 3.2 (After the completion of further analysis some of the larger categories were condensed which will be discussed below). A final list of the descriptive statistics can be viewed in Table 4.1 in the results section of this study.

Table 3.2 Original Descriptives

Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
Current	1,310	0.00	1.00	0.9664	0.18023
Delinquent	1,310	0.00	1.00	0.0336	0.18023
Watchlist	1,310	0.00	1.00	0.2511	0.43384
WLDIFF	1,310	-32.00	591.00	1.8282	30.86062
LoanGrt20Mill	1,310	0.00	1.00	0.0557	0.22948
LoanAmt10to20	1,310	0.00	1.00	0.0931	0.29073
LoanAmt5to10	1,310	0.00	1.00	0.2107	0.40795
LoanAmtLess5	1,310	0.00	1.00	0.6260	0.48406
LTVGrt.85	1,310	0.00	1.00	0.0382	0.19167
LTV.70to.85	1,310	0.00	1.00	0.6031	0.48945
LTV.55to.70	1,310	0.00	1.00	0.2756	0.44697
LTVLess.55	1,310	0.00	1.00	0.0702	0.25563
DSCRGrt2	1,310	0.00	1.00	0.0557	0.22948
DSCR1.50to2	1,310	0.00	1.00	0.3076	0.46169
DSCRLess1.50	1,310	0.00	1.00	0.6260	0.48406
OCCGrt90	1,310	0.00	1.00	0.8221	0.38254
OCC80to90	1,310	0.00	1.00	0.0878	0.28309
OCCLess80	1,310	0.00	1.00	0.0542	0.22650
Vintage96	1,310	0.00	0.00	0.0000	0.00000
Vintage97	1,310	0.00	1.00	0.1260	0.33192
Vintage98	1,310	0.00	1.00	0.3069	0.46137
Vintage99	1,310	0.00	1.00	0.2603	0.43897
Vintage00	1,310	0.00	1.00	0.0809	0.27281
Vintage01	1,310	0.00	1.00	0.1458	0.35304
Vintage02	1,310	0.00	1.00	0.0802	0.27163
Retail	1,310	0.00	1.00	0.2939	0.45572
Office	1,310	0.00	1.00	0.1603	0.36703
Multifamily	1,310	0.00	1.00	0.3061	0.46105
Lodging	1,310	0.00	1.00	0.0496	0.21724
Other	1,310	0.00	1.00	0.1901	0.39251
Smallmarket	1,310	0.00	1.00	0.6046	0.48913
Largemarket	1,310	0.00	1.00	0.3954	0.48913
JudicialForcl	1,310	0.00	1.00	0.3366	0.47274
NonJudicialForcl	1,310	0.00	1.00	0.6634	0.47274
Valid N (listwise)	1,310				

An improved grasp of the frequency of each variable was obtained after a review of the descriptive statistics. This information was used as a basis for inclusion of variables that did not represent the norm for both underwriting standards and collateral characteristics. For example the majority of the loans represented by the variables were under $5 million (62.6%). This variable creates a control group for the other less common categorical variables to be tested against.

Included in, but analyzed prior to the regression data, was the correlation matrix. This matrix is used to check for possible issues with co-linearity between independent variables.

Multicollinearity in logistic regression models is a result of strong correlations between independent variables. The existence of multicollinearity inflates the variance of the parameter estimates. That may result, particularly for small and moderate sample sizes, in lack of statistical significance of individual independent variables while the overall model may be strongly significant. Multicollinearity may also result in wrong signs and magnitudes of regression coefficient estimates and, consequently, in incorrect conclusions about relationships between independent and dependent variables. (Truszczynska 2005).

After analyzing the correlation matrix and anticipating issues from having too many variables, the categories for debt service coverage, loan-to-value, and loan balance were consolidated into two variables for each category. The variables for occupancy and market size were taken out completely. Occupancy was removed because it was thought to be adequately covered by the debt service coverage and loan-to-value at origination. The variable for market size was excluded due to inaccurate property address data.

3.3 Logistic Regression

Many categorical response variables only have two values, for example: yes/no, good/bad, pass/fail, etc. In models having these characteristics for the dependent variable it is suitable to use logistic regression. Logistic regression allows one to predict a discrete outcome, such as group membership, from a set of variables that may be continuous, discrete, dichotomous, or a mix of any of these. Many social phenomena are discrete or qualitative rather than continuous or quantitative in nature: an event occurs or does not occur, a person makes one choice but not the other, an individual or group passes from one state to another. Binary discrete phenomena usually take the form of a dichotomous indicator or dummy variable. Although it is possible to represent the two values with any numbers, employing variables with values of 1 and 0 has advantages. The mean of a dummy variable equals the proportion of cases with a value of 1, and can be interpreted as a probability. (Pampel 2000)

The standard logistic regression formula for a model with multiple independent variables is stated:

$$p(B) = \frac{\mathrm{Exp}(\alpha+\beta_1 x_1+\beta_2 x_2+\ldots+\beta_i x_i)}{1+\mathrm{Exp}(\alpha+\beta_1 x_1+\beta_2 x_2+\ldots+\beta_i x_i)}$$

or simplified

$$p(B) = \frac{1}{1+\mathrm{Exp}(-(\alpha+\beta_1 x_1+\beta_2 x_2+\ldots+\beta_i x_i))}$$

Where p(B) is the probability (p) that case "i" is a member of group B, such that p(B) = 1; Exp is a function that raises the number e exponentially to the power of the value inside the parenthesis, where the number e, Euler's number, is the irrational number whose natural logarithm is 1; α is the intercept constant; the βs are the coefficients for the independent variables; and the x's are the independent variables for the corresponding β coefficient.

Logistic regression fits the model for commercial default, because default is a qualitative variable and must be analyzed as such. Using logistic regression also allows a measurement of correlation and probability that a loan on the Watchlist for a certain trigger will move on to be a defaulted loan. Additionally, other control variables can be tested for their explanatory power in predicting default. This allows an analysis of possibly new triggers that could better explain and predict future commercial defaults and allows for an evaluation of the predictive power of variables currently utilized by the CMBS industry.

An initial regression run was completed using all of the variables in order to further evaluate their relationships and correlation with the dependent variable. A test for co-linearity was completed to assess correlation between independent variables. Co-linearity can cause issues with a model and cause less than reliable regression output. After completion of the initial run with all of the variables the Forward Stepwise method provided by SPSS was run to attempt to strengthen the model.

Stepwise refers to the steps taken by SPSS to select a set of independent variables to create a model with the best predictive fit. There are two basic types of stepwise logistic regression: forward inclusion and backward removal. In forward stepwise (method used in this study) all independent variables are initially kept out of the model, and through tests of significance (steps) are added to the model to produce the best fit. Independent variables that do not have a sufficient level of significance are left out of the model. Backward stepwise is completed in the opposite manner. Initially all of the independent variables are included, and as the same tests of significance are completed variables are removed from the model producing a model with independent variables of a certain level of significance in regard to the dependent variable. Below is the basic procedure for a forward stepwise regression.

STEP 0

1. Fit the "intercept only" logistic regression model and compute the log-likelihood.
2. Fit each independent variable into the model separately.
    a. Compute the log-likelihood values for each independent variable.
    b. Perform the likelihood ratio tests on each variable.
    c. Compute p values for each independent variable.
3. Select the independent variable with the smallest p value.
4. Proceed to Step 1 if the independent variable has a p value less than the threshold.

STEP 1

1. Fit the selected independent variable with the intercept logistic regression model and compute the log-likelihood for the model.
2. Fit the remaining independent variables into the model.
    a. Compute the log-likelihood values for each independent variable.

   b. Perform the likelihood ratio tests on each variable.

   c. Compute p values for each independent variable.

3. Select the independent variable with the smallest p value.

4. Proceed to Step 2 if this value is below the threshold.

STEP 2

1. Fit the two selected independent variables and the intercept into the logistic regression model and compute the log-likelihood for the model.

2. Check for removal of independent variables in the model.

   a. Compute the log-likelihood values for each independent variable.

   b. Perform the likelihood ratio tests on each variable.

   c. Compute p values for each independent variable.

3. Select the independent variable that when removed yields the largest p value.

4. If that p value is greater than the threshold remove it otherwise leave it.

5. Fit each remaining independent variable into the model.

   a. Compute log-likelihood values for each independent variable.

   b. Perform likelihood ratio tests for each independent variable.

   c. Compute p values for each independent variable.

6. Select the independent variable with the smallest p value.

7. Proceed to Step 3 if the p value is below the threshold.

STEP 3

This procedure it identical to Step 2. The program checks for backward elimination followed by forward selection. This is repeated until the Stop Step.

STOP STEP

1. All variables have been entered into the model, or
2. All variables in the model have p values below the threshold

All variables remaining in the model after this step are deemed to be significant.

Additional testing and modeling was completed. Due to the over abundance of variables, the number of variables was lowered and the regression was run with various combinations in an effort to increase the predictive power of the model. As mentioned previously the lowering of the number of variables was accomplished through the consolidation of the original groups and also the exclusion of some variables for different models.

CHAPTER IV

EMPIRICAL RESEARCH AND RESULTS

4.1 Non-Regression Analysis

This research utilizes Binary Logistic Regression to test the aforementioned research hypotheses. The hypotheses estimate that various CMBS industry related variables have a significant correlation and, therefore, have the capability of increasing the probability that a monetary default will occur.

The first step taken in analyzing the data was to evaluate the descriptive statistics (Table 4.1). The dependent variable of delinquency occurred 3.36 percent of the time, which is relatively low but representative of the percentage of defaulted loans in similar studies. The descriptives were reviewed to make sure that all of the variables included 1,310 pieces of data, then the control sets for each variable type were chosen. For each of the variable types one control category was created and left out of the regression. This allows the model to have something to compare each variable to. The control variables excluded were loan balances of less than $10 million, loan-to-value ratios of less than .75, the 1998 loan vintage, other property types, and non-judicial foreclosure.

Table 4.1 Model Descriptives

Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
Delinq	1,310	0.00	1.00	0.0336	0.18023
Watchlis	1,310	0.00	1.00	0.2511	0.43384
Wldiff	1,310	-32.00	591.00	1.8282	30.86062
loang10	1,310	0.00	1.00	0.1542	0.36128
loanl10	1,310	0.00	1.00	0.8458	0.36128
ltvg.75	1,310	0.00	1.00	0.3832	0.48635
ltvl.75	1,310	0.00	1.00	0.6168	0.48635
dcr1.5	1,310	0.00	1.00	0.3740	0.48406
dcrl1.5	1,310	0.00	1.00	0.6260	0.48406
vint97	1,310	0.00	1.00	0.1260	0.33192
vint01	1,310	0.00	1.00	0.1458	0.35304
vint00	1,310	0.00	1.00	0.0809	0.27281
vint99	1,310	0.00	1.00	0.2603	0.43897
vint02	1,310	0.00	1.00	0.0802	0.27163
Retail	1,310	0.00	1.00	0.2939	0.45572
Office	1,310	0.00	1.00	0.1603	0.36703
Multif	1,310	0.00	1.00	0.3061	0.46105
Lodging	1,310	0.00	1.00	0.0496	0.21724
Other	1,310	0.00	1.00	0.1901	0.39251
Judic	1,310	0.00	1.00	0.3366	0.47274
Nonjud	1,310	0.00	1.00	0.6634	0.47274
vint98	1,310	0.00	1.00	0.3069	0.46137
Valid N (listwise)	1,310				

A correlation matrix was created to assess the correlation between the independent variables. The detailed matrix is represented below in Table 4.2. This matrix is used to see if there are any variables that can possibly be removed because they are explaining the same thing as another variable in the study. Generally, independent variables should not be correlated over .85 percent or co-linearity issues will ensue. This data was analyzed between data not included in the same categories. As should be expected, extremely low and often negative correlations will be present for variables in the same categorical type because of the binary nature of the data. For example, a loan would only have been originated once, and therefore would have a low correlation with all other loan vintages. The correlation matrix produced by the data set appears to be logical and free of extremely high correlations between independent variables.

Table 4.2 Correlation Matrix

Correlation Matrix																
	Constant	watchlis	wldiff	loang10	ltvg.75	dcrl1.5	vint97	vint01	vint00	vint99	vint02	retail	office	multif	lodging	judic
Constant	1.000	-0.053	0.065	-0.006	-0.142	-0.125	-0.440	-0.453	-0.287	-0.432	-0.209	-0.619	-0.614	-0.595	-0.546	-0.350
Watchlis	-0.053	1.000	-0.276	0.053	0.029	-0.064	0.014	-0.045	0.017	0.022	0.000	0.008	0.000	-0.013	-0.071	-0.012
Wldiff	0.065	-0.276	1.000	0.026	-0.049	-0.068	-0.018	0.005	-0.001	-0.013	-0.310	0.020	0.004	-0.041	-0.008	-0.042
loang10	-0.006	0.053	0.026	1.000	-0.027	-0.105	-0.111	-0.089	-0.027	0.010	-0.046	-0.052	-0.093	0.033	-0.149	-0.077
ltvg.75	-0.142	0.029	-0.049	-0.027	1.000	-0.252	0.141	0.022	0.091	0.019	0.024	0.021	0.067	-0.148	0.018	-0.037
dcrl1.5	-0.125	-0.064	-0.068	-0.105	-0.252	1.000	-0.015	0.018	-0.162	-0.173	0.025	-0.152	-0.028	-0.077	0.154	0.037
vint97	-0.440	0.014	-0.018	-0.111	0.141	-0.015	1.000	0.479	0.360	0.445	0.262	0.030	0.103	-0.004	-0.100	0.087
vint01	-0.453	-0.045	0.005	-0.089	0.022	0.018	0.479	1.000	0.338	0.429	0.264	0.087	0.028	0.034	0.083	0.104
vint00	-0.287	0.017	-0.001	-0.027	0.091	-0.162	0.360	0.338	1.000	0.336	0.179	0.076	0.029	0.063	-0.049	0.002
vint99	-0.432	0.022	-0.013	0.010	0.019	-0.173	0.445	0.429	0.336	1.000	0.232	0.146	0.155	0.132	0.022	0.054
vint02	-0.209	0.000	-0.310	-0.046	0.024	0.025	0.262	0.264	0.179	0.232	1.000	-0.016	-0.005	-0.025	0.014	0.027
Retail	-0.619	0.008	0.020	-0.052	0.021	-0.152	0.030	0.087	0.076	0.146	-0.016	1.000	0.607	0.662	0.534	0.055
Office	-0.614	0.000	0.004	-0.093	0.067	-0.028	0.103	0.028	0.029	0.155	-0.005	0.607	1.000	0.609	0.514	0.033
Multif	-0.595	-0.013	-0.041	0.033	-0.148	-0.077	-0.004	0.034	0.063	0.132	-0.025	0.662	0.609	1.000	0.534	-0.005
Lodging	-0.546	-0.071	-0.008	-0.149	0.018	0.154	-0.100	0.083	-0.049	0.022	0.014	0.534	0.514	0.534	1.000	0.087
Judic	-0.350	-0.012	-0.042	-0.077	-0.037	0.037	0.087	0.104	0.002	0.054	0.027	0.055	0.033	-0.005	0.087	1.000

4.2 Full Logistic Regression Model

The initial model created from the subject data forced all of the variables originally selected. This approach was used to evaluate the correlation and significance of all of the independent variables prior to attempting to create a model with a high level of explanatory power. Logistic regression models use several types of output to explain the strength of a model and the components a model. These component parts include the Wald Statistic, the Sig. or Significance, Chi-square, Cox and Snell R Square, Nagelkerke R Square, and the overall percentage. A detailed listing of these statistics for the model can be seen in Tables 4.3, 4.4, and 4.5.

The Wald Statistic is generally used to test the significance of individual coefficients for each independent variable (in other words to test the null hypothesis that the particular effect of a coefficient is zero). The Wald Statistic calculates a Z statistic ($Z= B/SE$) for each coefficient in the model. The significance of each variable is also measured by the p statistic (column titled sign). The significance is used to measure the strength of the correlation between each independent variable and the dependent variable. In this model the level of significance was set at .05, which results in a variable resulting in a p score of less than .05 is statistically significant. According to both the Wald Statistic for the model coefficients and the corresponding p statistics, five of the model's independent variables were statistically significant. These five variables include (1) Watchlist, (2) Watchlist difference, (3) loan vintage 2001, (4) judicial foreclosure, and (5) lodging. All of variables chosen by the model as significant are logical choices for default predictability, and the results are a strong indicator that the hypothesis, of a loan appearing on the Watchlist being correlated with future delinquency is correct.

First, the type of foreclosure proceedings has been found to be significant in other CMBS studies. The basic reasoning for this is that borrowers are more likely to default due to the lengthy process of foreclosure in judicial states. The timing of foreclosure presents many problems for lenders in regards to obtaining their funds through payment or foreclosure. A prolonged period of non-payment, especially in the secondary mortgage market, requires funds be advanced to pay the bondholders, and a slow foreclosure process can present other problems with the condition of the collateral asset left to the care of a non-paying borrower. All of this gives savvy borrowers leverage for negotiation, potential for forbearance, and time for recovery or burning off excess cash. This effect is amplified by the non-recourse nature of most loans included in CMBS portfolios.

Second, lodging properties appearing as significant is probably the easiest correlation to picture. Ever since September 11, 2001 the majority of the lodging industry has been in a deep decline. Resulting from a large decline in travel from the business and tourism sectors, a large increase in problem lodging properties has been seen across the nation. Ultimately, the market has caused the default of many properties in this industry.

Lastly, the loan vintage of 2001 was found to be significant.

Table 4.3 Variables in the Equation

Variables in the Equation									
								vint97	
		B	S.E.	Wald	df	Sig.	Exp(B)	0	0
Step 1(a)	watchlis	-1.699	0.673	6.371	1	0.012	0.183	0.014	-0.045
	wldiff	0.063	0.021	8.729	1	0.003	1.065	-0.018	0.005
	loang10	0.099	0.434	0.052	1	0.820	1.104	-0.111	-0.089
	ltvg.75	-0.132	0.401	0.108	1	0.742	0.876	0.141	0.022
	dcrl1.5	-0.631	0.385	2.679	1	0.102	0.532	-0.015	0.018
	vint97	0.884	0.501	3.114	1	0.078	2.421	1.000	0.479
	vint01	0.985	0.495	3.967	1	0.046	2.679	0.479	1.000
	vint00	0.354	0.702	0.254	1	0.614	1.425	0.360	0.338
	vint99	-0.011	0.553	0.000	1	0.984	0.989	0.445	0.429
	vint02	-0.556	0.906	0.376	1	0.540	0.574	0.262	0.264
	retail	0.308	0.585	0.278	1	0.598	1.361	0.030	0.087
	office	0.502	0.614	0.667	1	0.414	1.651	0.103	0.028
	multif	0.477	0.564	0.713	1	0.398	1.610	-0.004	0.034
	lodging	1.522	0.674	5.097	1	0.024	4.583	-0.100	0.083
	judic	0.893	0.340	6.884	1	0.009	2.442	0.087	0.104
	Constant	-4.077	0.624	42.633	1	0.000	0.017		

a. Variable(s) entered on step 1: watchlis, wldiff, loang10, ltvg.75, dcrl1.5, vint97, vint01, vint00, vint99, vint02, retail, office, multif, lodging, judic.

Two other output components of a logistic regression are the Cox and Snell R-Square and the Nagelkerke R- Square. These are two attempts at creating an R squared variable for logistic regression similar to that for a linear regression.

Garson (2005) states there is no widely accepted direct analog to ordinary least squared or "OLS" regression's R squared. This is because an R squared measure seeks to make a statement about the "percent of variance explained," but the variance of a dichotomous or categorical dependent variable depends on the frequency distribution of that variable. For a dichotomous dependent variable, for instance, variance is at a maximum for a 50-50 split and the more lopsided the split, the lower the variance. This means that R squared measures for logistic regressions with differing marginal distributions of their respective dependent variables cannot be compared directly, and comparison of logistic R squared measures with R squared from OLS regression is also problematic.

The Cox and Snell R Squared and the Nagelkerke R Squared for this model were .056 and .219 respectively. The goal of this research was not to create a model with high explanatory power so these two statistics were not given heavy consideration. The complexities in evaluating these attempts at an R squared was also prohibitive, however, a stepwise regression was completed and will be detailed later. The results from the stepwise model add further difficulty to interpretation of these two components.

Table 4.4 Full Model Summary

Model Summary			
Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	310.022	0.056	0.219

According to Garson (2005), the Hosmer-Lemeshow Goodness-of-Fit Test tests the null hypothesis for the data entered into the model created by the researcher. The test divides subjects into deciles based on predicted probabilities then computes a Chi-square from observed and expected frequencies. A probability (p) value is then computed from the Chi-square distribution with eight degrees of freedom to test the fit of the logistic model. If the Hosmer-Lemeshow Goodness-of-Fit Test statistic is .05 or less, we reject the null hypothesis that there is no difference between the observed and model-predicted values of the dependent. If the H-L Goodness-of-Fit test statistic is greater than .05, as we want for well-fitting models, we fail to reject the null hypothesis that there is no difference, implying that the model's estimates fit the data at an acceptable level. This does not mean that the model necessarily explains much of the variance in the dependent variable; only however much or little it does explain is significant.

The H-L Goodness-of-Fit test for the subject model resulted in a quality fit supported by a p value over the H-L statistic of .05. The results of this test support the previous conclusion that there are significant variables affecting the likelihood of a dependent occurrence.

Table 4.5 Full Model Goodness-of-Fit

Hosmer-Lemeshow Test

Step	Chi-square	Df	Sig.
1	4.127	8	0.846

Contingency Table for Hosmer and Lemeshow Test

		delinq = .00		delinq = 1.00		Total
		Observed	Expected	Observed	Expected	
Step 1	1	129	129.670	1	0.330	130
	2	125	125.229	1	0.771	126
	3	130	129.786	1	1.214	131
	4	129	128.379	1	1.621	130
	5	128	128.003	2	1.997	130
	6	132	133.991	5	3.009	137
	7	130	128.238	2	3.762	132
	8	126	126.383	5	4.617	131
	9	125	125.695	8	7.305	133
	10	112	110.626	18	19.374	130

4.3 Forward Stepwise Logistic Regression

A forward stepwise regression model was created in order to try and create a model from the variables with the highest level of significance in an attempt to find the model that can be produced from the variables with the strongest goodness-of-fit. The independent variables selected by the stepwise regression were Watchlist difference, lodging, Watchlist yes/no, loan vintage 01, loan vintage 97, and judicial foreclosure. All of these variables have acceptable levels of significance and contributed to a strong goodness-of-fit. A detailed list of the variables selected can be viewed through the steps taken on Table 4.6 below.

Table 4.6 Stepwise Variables

Variables in the Equation									
								95.0% C.I. for EXP(B)	
		B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Step 1(a)	wldiff	0.042	0.018	5.731	1	0.017	1.043	1.008	1.080
	Constant	-3.494	0.164	455.381	1	0.000	0.030		
Step 2(b)	wldiff	0.043	0.018	5.960	1	0.015	1.044	1.009	1.081
	lodging	1.348	0.464	8.431	1	0.004	3.850	1.550	9.565
	Constant	-3.617	0.178	413.191	1	0.000	0.027		
Step 3(c)	watchlis	-1.719	0.669	6.608	1	0.010	0.179	0.048	0.665
	wldiff	0.057	0.020	8.269	1	0.004	1.059	1.018	1.101
	lodging	1.524	0.473	10.391	1	0.001	4.590	1.817	11.592
	Constant	-3.391	0.182	345.849	1	0.000	0.034		
Step 4(d)	watchlis	-1.698	0.671	6.397	1	0.011	0.183	0.049	0.682
	wldiff	0.054	0.018	8.850	1	0.003	1.056	1.019	1.094
	lodging	1.641	0.480	11.679	1	0.001	5.162	2.014	13.231
	judic	0.871	0.333	6.834	1	0.009	2.389	1.244	4.591
	Constant	-3.781	0.258	214.337	1	0.000	0.023		
Step 5(e)	watchlis	-1.778	0.672	7.008	1	0.008	0.169	0.045	0.630
	wldiff	0.056	0.018	9.359	1	0.002	1.058	1.020	1.097
	vint01	0.897	0.402	4.978	1	0.026	2.453	1.115	5.397
	lodging	1.758	0.488	12.961	1	0.000	5.799	2.227	15.099
	judic	0.895	0.334	7.172	1	0.007	2.448	1.271	4.715
	Constant	-3.957	0.279	200.461	1	0.000	0.019		

Table 4.6 - continued

Variables in the Equation									
Step 6(f)	watchlis	-1.758	0.673	6.826	1	0.009	0.172	0.046	0.645
	wldiff	0.058	0.019	9.758	1	0.002	1.060	1.022	1.099
	vint97	0.995	0.417	5.690	1	0.017	2.706	1.194	6.130
	vint01	1.111	0.421	6.982	1	0.008	3.038	1.332	6.928
	lodging	1.527	0.504	9.199	1	0.002	4.606	1.717	12.357
	judic	0.927	0.336	7.625	1	0.006	2.527	1.309	4.878
	Constant	-4.177	0.307	185.153	1	0.000	0.015		

a. Variable(s) entered on step 1: wldiff.
b. Variable(s) entered on step 2: lodging.
c. Variable(s) entered on step 3: watchlis.
d. Variable(s) entered on step 4: judic.
e. Variable(s) entered on step 5: vint01.
f. Variable(s) entered on step 6: vint97.

The forward stepwise regression found several strongly significant variables and produced a model based on this significance. Though the model shows significance, according to the –2 log-likelihood level achieved by the model, the likelihood of the independent variables being represented as having a log linear relationship with the dependent is unlikely. This means that though the variables in this model are significant the probability that they would be significant in a similar log linear model is unlikely. The low Cox and Snell and Nagelkerke R squareds of .052 and .205 respectively, support this finding. A summary can be view below in Table 4.7.

Table 4.7 Stepwise Model Summary

Model Summary			
Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	348.163	0.028	0.109
2	341.733	0.033	0.128
3	331.043	0.040	0.159
4	324.256	0.045	0.178
5	319.919	0.049	0.191

Table 4.7 - continued

Model Summary			
6	314.830	0.052	0.205

a. Estimation terminated at iteration number 6 because parameter estimates changed by less than .001.
b. Estimation terminated at iteration number 8 because parameter estimates changed by less than .001.

The stepwise regression also calculated the goodness-of-fit for the model. The general rule of thumb is that the higher the Chi-square significance the better the fit of the model. In the model produced by the stepwise regression a relatively high significance was produced, therefore it can be assumed that the null hypotheses for the variables included in the model can be rejected or that the independent variables do exhibit a high level of correlation with the dependent. A detailed table of the Hosmer-Lemeshow Test can be seen in Table 4.8.

Table 4.8 Stepwise Goodness-of-Fit

Hosmer and Lemeshow Test			
Step	Chi-square	df	Sig.
1	0.000	0	.
2	171.299	1	0.000
3	958.929	2	0.000
4	0.368	3	0.947
5	1.272	4	0.866
6	2.619	5	0.758

Contingency Table for Hosmer and Lemeshow Test						
		delinq = .00		delinq = 1.00		Total
		Observed	Expected	Observed	Expected	
Step 1	1	0	1.973	2	0.027	2
	2	1,266	1,264.027	42	43.973	1,308
Step 2	1	0	1.977	2	0.023	2
	2	1,208	1,205.607	30	32.393	1,238
	3	58	58.417	12	11.583	70

Table 4.8 - continued

Contingency Table for Hosmer and Lemeshow Test						
		delinq = .00		delinq = 1.00		Total
		Observed	Expected	Observed	Expected	
Step 3	1	0	1.996	2	0.004	2
	2	297	295.219	0	1.781	297
	3	24	23.353	0	0.647	24
	4	945	945.432	42	41.568	987
Step 4	1	205	205.146	1	0.854	206
	2	111	110.686	1	1.314	112
	3	598	598.355	14	13.645	612
	4	5	4.755	0	0.245	5
	5	347	347.058	28	27.942	375
Step 5	1	166	166.462	1	0.538	167
	2	126	125.800	1	1.200	127
	3	516	515.147	9	9.853	525
	4	29	28.324	0	0.676	29
	5	279	279.893	14	13.107	293
	6	150	150.374	19	18.626	169
Step 6	1	145	144.617	0	0.383	145
	2	136	136.940	2	1.060	138
	3	434	435.320	8	6.680	442
	4	38	37.181	0	0.819	38
	5	250	249.334	9	9.666	259
	6	164	162.816	6	7.184	170
	7	99	99.792	19	18.208	118

4.4 Implications and Further Research

The results from this model show a high level of significance for the Watchlist variables and for several other variables of interest. This study shows that even prior to the implementation of the CMSA Watchlist standards that the Watchlist contained significant predictive capabilities and should continue to be produced and refined. The findings of other significant and insignificant variables agree with most of the findings of the literature review. The surveillance implications of these findings lean towards improved internal risk prediction, but also detract from the power of default prediction through "industry standard" underwriting data. In other words, little dependence should be placed on the ability of loan-to-value, occupancy, and debt service ratios at origination to be significant predictors of future default. Instead focus should be placed on active monitoring of both collateral level items and market predictors both inherent in the market and appearing from changing economic conditions.

Many complications with prediction continue to be experienced in the surveillance field with complications and anomalies arising due to the non-recourse nature of these loans and the multi-party nature of the industry. As information continues to improve in quality and quickness however, the probabilities of prediction will likely improve. Intuitively, further research should be completed on seasoned CMBS originated after the 2003 CMSA Watchlist standards were implemented. A study along these lines could be compared to these findings and analyzed to see if the additional criteria and monitoring has created an environment of greater predictability.

REFERENCES

Ambrose, B., Benjamin, J., & Chinloy, P. (1996). Credit restrictions and the market for commercial real estate Loans. Real Estate Economics, 24 (1), 1-22.

Ambrose, B., Buttimer, R., & Capone, C. (1997). Pricing mortgage default and foreclosure delay. Journal of Money, Credit, and Banking, 29 (3), 314-325.

Archer, W., Elmer, P., Harrison, D. & Ling, D. (2002). Determinants of multifamily mortgage default. Real Estate Economics, 30 (3), 445-473.

Berger, S. & Sternin, J. (2003). Watching Servicer Watchlists. CMBS World, Summer 2003, 36-38.

Capozza, D., Kazarian, D., & Thomson, T. (1997). Mortgage default in local markets. Real Estate Economics, 25 (4), 631-655.

Chun, R., & Goldberg, M. (2003). How soon is imminent for CMBS defaults? Structured Finance. April 25, 1-4.

Ciochetti, B., Dang, Y., Goa, B., & Yao, R. (2002). The termination of commercial mortgage contracts through prepayment and default: A proportional hazard approach with competing risks. Real Estate Economics, 30 (4), 595-634.

Ciochetti, B., & Vandell, K. (1999). The performance of commercial mortgages. Real Estate Economics, 27 (1), 27-61.

Ciochetti, B., Youngheng, D., Lee, G., Shilling, J., & Rui, Y. (2003). A proportional hazards model of commercial mortgage default with originator bias. Journal of Real Estate, Finance and Economics, 27 (1), 19-24.

Cooperstein, R., Stevens, F., & Meyers, H. (1996). Modeling mortgage terminations in turbulent times. Journal of American Real Estate and Urban Economics Association, 19 (4), 473-494.

Cramer, J.S., "The Origins of Logistic Regression" (December 2002). Tinbergen Institute Working Paper No. 2002-119/4.

Elmer, P. (1992). PLAM default risk. The Journal of Real Estate Research. Spring 1992, 157-167.

Episcopos, A., Pericli, A., & Hu, J. (1998). Commercial mortgage default: A comparison of logit with radial basis function networks, Journal of Real Estate Finance and Economics, 17 (2), 163-178.

Garson, G. (2001). Logistic Regression. North Carolina State University. http://www2.chass.ncsu.edu/garson/pa765/logistic.htm.

Gordon, S., & Kizer, L. (2004). U.S. CMBS loan performance: Impact of seasoning, leverage and location on probability of default. structured finance. November 8, 1-30.

Harrison, D., Noordewier, T., & Yavas, A. (2004). Do riskier borrowers borrower more? Real Estate Economics, 32 (3), 385-411.

Lambrecht, B., Perraudin, W., & Satchell, S. (2003). Mortgage default and possession under recourse: A competing hazards approach. Journal of Money, Credit, and Banking, 35 (3), 425-442.

Lans, D., & Price, J. (2001). Comparing CMBS and corporate bond defaults. CMBS World. Vol. 2: 12-14.

Lee, G., Crowell, P., Fitzsimmons, P., & Serif, U. (2003). CMBS delinquencies: The devil is the denominator. CMBS Market Watch Weekly. July 18, 1-15.

Lee, G., Crowell, P., Fitzsimmons, P., & Serif, U. (2003). CMBS delinquency update: “Drilling Down.” CMBS Market Watch Weekly. May 9, 1-20.

Lekkas, V., Quigley, J., & Order, R. (1993). Loan loss severity and optimal mortgage default. Journal of American Real Estate and Urban Economics Association, 21 (4), 353-371.

Maris, B., & Segal, W. (1999). Analysis of Yield Spreads on Commercial Mortgage-Backed Securities. Journal of Real Estate Research. Vol 23: 236-252.

Mejia, L. (1999). Availability of credit and loan default: A look at the commercial mortgage supply cycle. Journal of Real Estate Research, 18 (1), 175-196.

O’Rourke, M., & Metz, M. (2004). 2004 U.S. CMBS conduit loan default and loss study. Structured Finance. August 23, 1-13.

Pampel, F. (2000). Logistic Regression. Sage Publications.

Parkus, R. (2002). Analyzing default risk in conduit CMBS. Deutsche Bank. October 4, 1-39.

Pyhrr, S., Born, W., Robinson, R., & Lucas, S. (1996). Real property valuation in a changing economic and market cycle. The Appraisal Journal. January, 14-24.

Riddiough, T., & Thompson, H. (1993). Commercial mortgage pricing with unobservable borrower default costs. Journal of the American Real Estate and Urban Economics Association, 21 (3), 265-291.

Ross, S. (2000). Mortgage lending, sample selection and default. Real Estate Economics, 28 (4),: 581-621.

Rutherford, S. (2004). A statistical approach to default risk: implications for OCM CMBS surveillance. February 27, 1-14.

Saunders, J. (2003). CMBS Due Diligence Underwriting: Helping Investors Make Informed Decisions. CMBS World, Spring 2003, 52-55.

Connor, E. (2005). Logistic Regression. San Francisco State University. http://userwww.sfsu.edu/~efc/classes/biol710/logistic/logisticreg.htm

Titus, J., Betancourt, M., Grossman, E., & Kozel, P. (2003). Real Estate Market Perspectives: Outlook By Sector for 2003. Real Estate Outlook, Winter 2003. 1-36.

Truszczynska, H. (2005). Multicollinearity in Logistic Regression. University of Kentucky http://www.uky.edu/ComputingCenter/SSTARS/Multicollinearityin LogisticRegression.htm

Tu, C., & Eppli, M. (2003). Term default, balloon risk, and credit risk in commercial mortgages. The Journal of Fixed Income, December, 42-52.

Vandell, K. (1984). On the assessment of default risk in commercial mortgage lending. Journal of the American Real Estate and Urban Economics Association,, 12 (3), 270-297.

Vandell, K. (1992). Predicting commercial mortgage foreclosure experience. Journal of the American Real Estate and Urban Economics Association, 20 (1), 55-88.

Vandell, K., Barnes, W., Hartzell, D., Kraft, D., & Wendt, W. (1993). Commercial mortgage defaults: Proportional hazards estimation using individual loan histories. Journal of the American Real Estate and Urban Economics Association, 32 (4) 451-480.

Vandell, K., & Thibodeau, T. (1985). Estimation of mortgage defaults using disaggregate loan history data. Journal of the American Real Estate and Urban Economics Association, 13 (3), 292-317.

Wuensch, K. (2004). Binary Logistic Regression with SPSS. East Carolina University. http://core.ecu.edu/psyc/wuenschk/MV/Multreg/Logistic-SPSS.doc

BIOGRAPHICAL INFORMATION

Prior to enrolling in the Master of Science in Real Estate program at the University of Texas at Arlington, the author completed his undergraduate studies at Angelo State University graduating in 2002 with a Bachelor's degree in Finance with a Real Estate Option. Since graduating from his undergraduate studies he has been employed by ORIX Capital Markets, splitting time between collateral surveillance and asset administration. His research interests include commercial mortgage default prediction and collateral tracking for CMBS portfolios.