Default Prediction of Various Structural Models

Ren-Raw Chen,*

Rutgers Business School – New Brunswick

94 Rockafeller Road

Piscataway, NJ 08854

Shing-yang Hu,

Department of Finance

National Taiwan University

Taipei, Taiwan

Ging-Ging Pan

Department of Finance

National Ping-Tung University of Sciences and Technologies

Ping-Tung, Taiwan

July 21, 2006

* We would like to thank Hsuan-Chu Lin for pointing out a critical typographical error in the Black-Cox model. Ren-Raw Chen would like to Douglas Dwyer at Moody's KMV and Roger Stein at Moody's Investors Services, and Jalal Akhavein, Greg Gupton, and Ahmet Kocagil at FitchRatings for helpful discussions and insightful inputs at an early stage of the paper. We are obliged to valuable comments by Neil Pearson. All errors are our own.

Abstract

The structural approach of credit risk modeling has gained growing attention in both academics and industry. While there has been increasing effort in testing the structural models for credit derivatives pricing, little result has been shown on the default prediction of the structural models. In this paper, we compare comprehensively various credit structural models for their default prediction capability. We select models that cover distinctly different assumptions such that we can study how and why certain models can predict default better than others. In addition to the well known existing structural models, we also introduce a non-parametric model to study the distributional characteristics underlying the structural models.

Our results indicate that the distribution characteristics of the equity returns and endogenous recovery are two important assumptions. On the other hand, random interest rates that play an important role in pricing credit derivatives are not an important assumption in predicting default.

1. Introduction

There are two approaches in pricing credit risk – the reduced form approach (represented by Jarrow and Turnbull (1995) and Duffie and Singleton (1999)) that assumes exogenous default and recovery and the structural approach (pioneered by Black and Scholes (1973) and Merton (1974)) that regards default and recovery endogenously as a result of bad operation of the firm. Proponents of the structural models contend that equity prices carry useful credit information that can be used to price credit derivatives. For example, Vassalou and Xing (2004) find size and book-to-market of a company to exhibit a strong link to the Merton model induced probability of default. Huang and Huang (2003) and Eom, Helwedge and Huang (2004) try to explain corporate bond prices using various structural models. Both studies conclude that the structural models do not price corporate bonds well.[1] In this paper, we argue that although the structural models are not accurate in predicting bond spreads, they are very effective in predicting defaults. This is because predicting defaults requires only ranking and not level.

Industry has long used the structural models to predict defaults.[2] And yet there has been little work in academics to validate such approach. The only study is recently by Bharath and Shumway (2004) that compares the simple Merton model with the Hazard model by Shumway (2001). They conclude that the Hazard model performs marginally better than the Merton model. We show in the paper that many structural models perform substantially and significantly better than the Merton model. For example, the accuracy ratio (that measures the power of a model in predicting default, to be explained fully later) of the Black-Cox model is 10% higher than that of the Merton model in predicting defaults 12 months ahead of time[3] and in a sub-sample comparison a slightly tweaked Merton model (Non Parametric

---

[1] Pricing corporate bonds with the structural models can be problematic. First of all, corporate bonds, except for very few, are highly illiquid. Most studies use matrix prices that are generated with smoothing algorithms. Using a model to test model-generated prices is tautological. For the very limited number of actual transaction prices, market participants believe that they reflect many technical components as opposed to true values of the bond and therefore cannot be priced by no arbitrage models. Moreover, market participants believe that there are high discrepancies between bond prices quoted by bond dealers and those computed from equity prices due to the segmentation in trader background and skillset.

[2] KMV (now Moody's KMV), founded in 1988, was known as the first company adopting the Merton model in predicting defaults.

[3] See Table 7.

model) can outperform the Merton model by over 19% in predicting defaults 1 month ahead of time.[4]

Forecasting default is not new. Besides rating agency's ratings, quantitative models also have a long history, e.g., Altman's z score (1968) and the Ohlson model (1980).[5] However, these models use historical information that is staled and furthermore ignore the important information embedded in volatility. Moody's KMV has found even the simple Merton model can outperform Altman's z score by a significant margin.[6] In this paper, we compare comprehensively various structural models for their default prediction capability. We select models that cover distinctly different assumptions such that we can study how and why certain models can predict default better than others. We choose Merton (1974), Longstaff and Schwartz (1997), Flat Barrier,[7] Black and Cox (1976), and Geske (1977) models. In addition to these well known structural models, we also introduce a Non Parametric model to study the distributional characteristics underlying the structural models.

Our results indicate that the distribution characteristics of the equity returns and endogenous recovery are two important assumptions. On the other hand, random interest rates that play an important role in pricing credit derivatives are not an important assumption in predicting default.

2. The Models

In this paper, we select the following structural models:

- Black-Scholes-Merton
- Flat Barrier
- Black-Cox (exponential barrier)
- Geske (2 periods)
- Longstaff-Schwartz
- Non Parametric

[4] See the Financial subsample in Table 9.

[5] Interested readers are referred to Shumway (2001) for a careful review of default prediction models.

[6] See the large number of white papers in Moody's KMV websites, for example, http://www.moodyskmv.com/research/singleObligor_wp.html.

[7] For the barrier model, see Hull (2004).

All of above models, except for the Non Parametric model, are closed form and given in the Appendix. The Merton (Black-Scholes) model and the Non Parametric model are single period and the rest are multi-period. The Non Parametric model (to be explained in details later) replaces the asset value distribution in the Merton model with an empirical distribution. The Black-Cox, Longstaff-Schwartz, and Flat Barrier models assume an exogenous default barrier. The Black-Cox model assumes a exponential barrier and the Flat Barrier and Longstaff-Schwartz models assume a flat barrier. Both the Black-Cox model and the Flat Barrier model assume a constant interest rate but the Longstaff-Schwartz model assumes the Vasicek model for the term structure of interest rates. The Black-Cox and the Longstaff-Schwartz models are similar in assuming exogenous recovery and yet the Flat Barrier model recovers the asset value as in the Merton model. The Geske model allows both recovery and default barrier to be determined endogenously.

In all of these models, the asset value and the asset volatility are the unknown inputs to the model. The equity is treated as a call option. By Ito's lemma, we obtain the equality volatility as follows:

$$\sigma_E = \sigma_A \frac{\partial E}{\partial A} \frac{A}{E} \tag{1}$$

where A is the market asset value, σ_A is the asset volatility, both of which are unobservable, E is the equity value and σ_E is the equity volatility, both of which are observable. We use the equity price and equity volatility that are observable to back out asset value and asset volatility.

The Merton (Black-Scholes) model is a single period model where the underlying company issues a zero coupon debt. However, in reality, companies issue multiple debts with coupons, making the Merton practically infeasible. One suggestion, proposed by Moody's KMV, is to convert book values of various debts into an *equivalent one-year value* and then use the Merton model. The conversion is determined empirically by data. A rough formula provided by Moody's KMV is half of long term debt and total debt, or short term debt plus half of long term debt.[8]

There are additional assumptions to be made for other models. The Flat Barrier model and the Longstaff-Schwartz model require an exogenous fixed barrier, which we set it to be the discounted value of the one-year equivalent value of debt at the risk free rate. The

[8] This formula is given in a KMV white paper. The actual formula varies with companies and times.

Longstaff-Schwartz model needs further parameter values of the Vasicek interest rate model that is estimated by the method described in the Appendix. The correlation of the interest rates and the equity returns is set to zero. The Black-Cox model requires a terminal barrier and a depreciation rate for the exponential barrier, which are set to be the same as the one-year equivalent value and the risk free rate respectively. The Geske two-period model needs two strike values which are set to be the short term debt and long term debt respectively.

A. Distant to Default

After we use the models to estimate the current asset value A and the asset volatility σ_A, we calculate the standard statistic for measuring the default likelihood, known as *distance to default*, or DD. It measures quantitatively how far the company is away from its default point for a chosen horizon. DD is defined as the number of standard deviations of the expected asset value deviating from the debt obligation of the company. To assess the DD for a future point in time, since the asset value is random, we compute the Euclidian distance between the expected value of asset and the default point from it, and then divide it by the standard deviation. Formally,

$$\frac{E[A]-K}{\sigma_A} \tag{2}$$

where $E[\cdot]$ is the expectation operation under the real measure. The default point, K, is calculated based upon the book value of debt which is at most updated 4 times a year (for example, in the Merton model, it is approximated as the average of short term debt and total debt). Consequently, even though equity prices are liquid and frequent, DDs are subject to much more staled information and less accurate.[9]

Empirically equation (2) does not guarantee DD to be positive. In fact, we do experience negative DDs in our sample. Note that the magnitudes of the DD measure means little but only the relative ranking is important.

DD is usually translated to *probability of default* (known as PD) to give a quantitative measure as to how likely a company is going to default. Both reduced form and structural models give risk neutral PDs from DDs, hence additional steps must be taken to arrive at real PDs.[10]

---

[9] Note that in the Merton model in the log space, the distance to default is precisely d_2 defined in the Appendix.

[10] References of the conversion from risk neutral PDs to real PDs can be found in a number of places, for example, Moody's KMV.

While risk-neutral PDs are useful in pricing corporate bonds and credit derivatives (e.g., see Eom, Helwedge and Huang (2004)) and carry important information of correlation and price change (Delianedis and Geske (2003)), real PDs are important in the sense that ,for practical purposes, investors and regulators still need to gauge the actual likelihood of default. Also, being able to compute reliable real PDs permits the model to be able to calibrate to Moody's or S&P's historical default probabilities, which improves the pricing capability of the models.

Luckily, in predicting default, risk neutral and real PD calculations are not necessary. All we need for default prediction is the relative rankings of the DDs. Since transforming DDs to PDs (risk neutral or real) is a monotonic transformation, the relative rankings are preserved and hence no PD calculation is necessary. This drastically reduces the problem of model mis-specification in applying the structural models to price corporate bonds and credit derivatives.

B. Our Non-parametric Model

In addition to the structural models, we also introduce a Non Parametric model. The purpose for this Non Parametric model is to capture the known effect of the skewness and kurtosis in stock returns and hence to correct the biases in the chosen structural models. As we shall see later, all models produce right-skewed DDs (median smaller than mean) and the skewness drops as the prediction period lengthens. This is similar to the equity returns that demonstrate dissipating skewness as the investment horizon becomes longer. This is the motivation for us to propose the Non Parametric model.

Furthermore, the NP model also incorporates the leverage effect more effectively than the exisiting structural models. This approach is similar to the traditional bankruptcy models (see Altman (1968) and Ohlson (1980) and the citation in Shumway (2001)) in that we adopt historical information. However, different from the traditional models, we use only the historical information of the equity prices and nothing else in order to be comparable to the structural models. A benefit of the Non Parametric model is that it does not assume a particular parametric distribution of the asset value. Rather, it lets the history of the asset returns (implied from the history of the equity returns) be the true distribution.

To implement the Non Parametric model, we need distribution of returns on asset, or ROAs, (i.e. $\ln A(T)/A(t)$ in Merton) but the market ROAs are not observable. What are observable are the returns on equity, or ROEs. Hence, we need a mapping from distribution of ROEs to the distribution of ROAs. The mapping we adopt is the WACC equation:

$$r_A = \frac{D}{A} r_D + \frac{E}{A} r_E \tag{3}$$

where r_A is ROA, r_D is cost of debt approximated by the risk free rate, and r_E is ROE (here we assume no tax). A, D, and E are market asset value, market debt value, and market equity value respectively. ROE is defined as weekly returns of equity and E is defined as the market cap of the firm. Since the market value of debt is not directly observable, many researchers use the book value. However, using the book value of debt for D poses a severe problem. When a firm is near default, both E and D are small but the book value of debt remains fixed. To resolve this problem, we apply the following algorithm.

Define $\bar{D}$ as the book value of debt and $\bar{A} = \bar{D} + E$. When the firm is far from default, the equity is deep in the money and $\partial E / \partial A \to 1$, or equivalently $D \to \bar{D}$ and $\partial D / \partial A \to 0$.[11] When the firm is near default, both D and E are near 0 but E is approaching 0 much faster than D (i.e. $\partial D / \partial A \to 1$ and $\partial E / \partial A \to 0$, again, the insight from the Merton model). At this time, we let $A \sim D$. In summary, when a firm is far from default A is approximated by $\bar{D} + E$ and when it is near default it is approximated by just D. For in-between cases, we apply $\partial D / \partial A$ as the weight. That is:

$$A = \frac{\partial D}{\partial A} D + \left(1 - \frac{\partial D}{\partial A}\right)(E + \bar{D}) \tag{4}$$

Substitute in the LHS by $A = D + E$ and get:

$$D = \bar{D} + \left(1 - \frac{1}{\partial E / \partial A}\right) E \tag{5}$$

In the implementation, we let $\partial E / \partial A$ be approximated by the Merton model. In other words, we let $\partial E / \partial A = N(d_1)$ of the Merton model defined in the Appendix. We use past 12 months of equity prices to for the ROE histogram. Then we obtain the histogram for ROA's by (3) where the debt is obtained by (5) and $A = D + E$. σ_A is the standard deviation of the 52 ROA observations.

3. The Data

The equity prices are obtained monthly from CRSP and financials are quarterly from Compustat. A historical equity volatility (standard deviation) is computed using 12 past

[11] This insight can be most easily gained by the Merton model. Note that even thought different models suggest different results for the equity and debt, this insight remains the same.

monthly discrete returns. The equity price and the volatility are used jointly to back out the asset price and asset volatility.

The default data are obtained from BankruptcyData.com which collects Chapter 11 filings with asset greater than $50 million. The period of study is from January 1985 till December 2002. It covers over 6,000 firms and 1,047,727 observations. During this period, the number of defaulted firms totaled 1,091 with 2001 reaching the peak of 166 firms, accounting for 2.81% of the total number of firms of that year.

At any given point in time, we have information of the defaulted firms. The definition of default is based on BankruptcyData.com and the bankruptcy date is either the date of Chapter 11 filing or the date of delisting, whichever comes first. We cannot use the bankruptcy date (Chapter 11) as the default date because many stocks were delisted many months prior to bankruptcy. To test if a model can forecast such default, we travel back in time to use the company's information back then. For example, to test if a model can predict default 6 months prior to default, we use the financial and market information 6 months prior to default (default filing or delisting). Table 1 presents the statistics of the sample. For example, in 1985, there were total of 18 companies that defaulted (last column). 18 of those companies had financial information 1 month prior to their defaults (column 4) and 18 of them had financial information 12 months prior to their default (next to last column). In 2001, there were 166 actual defaults but only 144 firms had financial information 1 month prior to their defaults and 149 firms had financial information 12 months prior to default. This is because many companies stopped releasing financial information once they were certain that they would file bankruptcy. So the stock market continued to trade and yet no financial information was available. The rest of Table 1 continues to show further breakdowns from 1 month forecast to 12 months forecast. Column 1 shows the number of observations studied in each year. For example, there were 48,383 observations from 4,428 firms in 1985 and 73,166 observations from 6,665 firms in 1997. The total number of observations in our sample is 1,047,727 in the whole sample. We use this sample to start our analysis. However, the sample size will reduce in other analyses (e.g. The Non Parametric model requires a history of returns and hence will reduce the sample size. Also, not all firms have agency ratings, that will reduce the sample size.)

Table 2 through Table 4 continue to show how these defaulted firms are classified by market capitalization (stock price × outstanding shares), leverage (total liabilities ÷ (market cap + total liabilities)), and SIC code.

Clearly, size has a negative impact on bankruptcy (i.e. larger the size, less likely to default) and leverage has a positive impact (i.e. higher the leverage, more likely to default). This is true regardless of the forecast period. Table 2-1 confirms the common wisdom that smaller firms are more prone to default, regardless of the prediction period. Another common wisdom confirmed by Table 2-1 is that the extremely small firms may come and go quickly. What we observe from Table 2-1 is that as we move up in size, there are (relatively proportionally) more firms that have financial information 12 months prior to default. For example, from lowest decile to the second lowest, we observe from 422 firms of 1 month and 2646 firms of 12 months (which represents 15.95%) to 162 firms of 1 month and 2002 firms of 12 months (which represents only 8.09%).

Table 3 and Table 3-1 show another dimension of how default can become a problem when leverage is high. Table 3-1 demonstrates a clear monotonic relationship between leverage and bankruptcy. We should particularly note that regardless of which default prediction period, the number of bankruptcies grows exponentially even though the leverage ratios are divided equally. Between the last decile and next to last, the number of bankruptcies grows 3 times for 1-month forecast and twice for 12-month forecast.

Table 4 demonstrates the industry breakdowns. Quite surprisingly, we would expect bankruptcy to be a macro event and hits indiscriminately across industry sectors. On the contrary, some industries (Manufacturing (Division 3), Transportation, Communications, Electric, Gas, and Sanitary Services (Division 4), and Wholesale Trade, Retail Trade (Division 5)) have experienced more defaults than others.

Lastly, Table 5 breaks down the sample by credit ratings. We note that only a fraction of the sample contains rating information. Out of total size of more than 1 million records, only about one-fifth (202,066) contain credit ratings. In this smaller sample set, the majority data belong to A and BBB rating classes. BB has an important fraction. The rest is trivial. Nevertheless, we still observe that the default likelihood is rather monotonic (except for A and A-), which brings some comfort when we test the models by ratings.

For the benchmark Treasury rates and to estimate the term structure model, we use the daily 3-month Constant Maturity Treasury (CMT) rates for the same period published by the U.S. Department of the Treasury. These rates are available from the web site of the Federal Reserve Bank of St. Louis.[12]

---

[12] http://research.stlouisfed.org/fred2/categories/22

4. Estimation Methodology

We test the default prediction power by various models in the following horizons: 1 month, 2 months, …, 12 months prior to default. As described in the Data section, our sample consists of more than 15 years of data grows from 1,256 firms in 1985 to near 5,000 firms in late 1990s and early 2000s.

A. Estimation of the Vasicek Model

The Longstaff-Schwartz model adopts the Vasicek term structure model to capture the dynamics of interest rates. The Vasicek model assumes the following mean-reverting Gaussian process for the instantaneous short rate:

$$dr = \alpha(\mu - r)dt + \upsilon dW \tag{6}$$

where α, μ, and υ are constants representing reverting speed, reverting level, and volatility respectively. This process, observed in discrete time, is effectively an AR(1) process as follows:

$$r_t = \left(1 - e^{-\alpha h}\right)\mu + e^{-\alpha h} r_{t-h} + \sqrt{\frac{\upsilon^2\left(1 - e^{-2\alpha h}\right)}{2\alpha}}\, e_t \tag{7}$$

where $h = \frac{1}{252}$ and $e_t \sim N(0,1)$. The parameters hence can be estimated by an autoregression:

α	0.7116
μ	0.0352
υ	0.0437

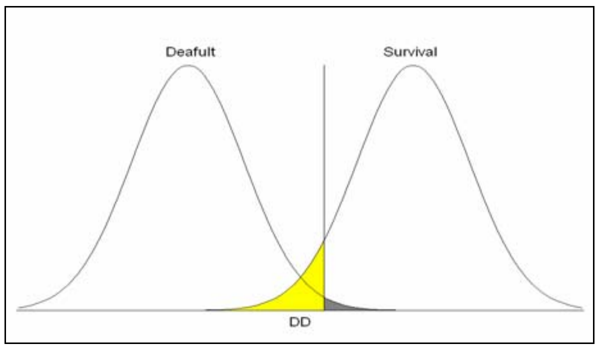
Details are provided in the Appendix.

B. ROC Methodology and AR Test

The Receiver Operating Characteristic (ROC) curves were first developed in the 1950's for radio signals contaminated by noise. Later it was introduced to medical research for testing the power of various diagnosis tests. A powerful test in disease detection should produce the least possible false positives and false negatives and the ROC curve is ideal for depicting these errors. False positive and false negative of a medical test is similar to the Type-I and Type-II errors in statistics. Hence formal statistical testing is possible when we use the ROC methodology. Similar to the medical test that tries to differentiate healthy individuals from unhealthy ones, a powerful default model should be able to successfully differentiate financially healthy firms from unhealthy firms. Yet only recently, this methodology has been introduced to evaluating credit risk. The first application of such methodology in finance appeared in a series of Moody's quantitative credit risk modeling group (e.g. Stein (2002, 2005)).[13]

Stein (2002, 2005) argues that the power of a default model to predict defaults is its ability to detect "True Positive", or true default, and the capability of a default model to calibrate to the data is its ability to detect "True Negative", or true survivals.[14] A good model should be both powerful and capable to calibrate.

The details of the ROC methodology can be found, for example, in Hanley and McNeil (1982) and the default application can be found in Stein (2002, 2005).[15] Briefly, there are two populations, default and survival, each of which is plotted against DD as follows:



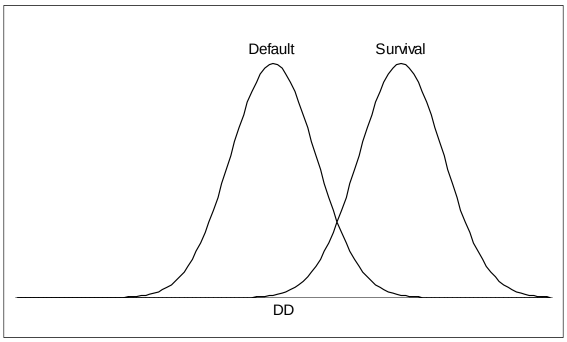
Since no model is absolutely perfect and powerful, it is always possible that some small quantity of defaulted firms are erroneously evaluated to have higher DDs than some (also small quantity) survival firms. As a result, no matter where we set the cut-off point of DD to differentiate defaulted firms from the survived ones, there will always be 'false defaults' (i.e. firms that are classified as default actually survived) and 'false survivals' (firms that are classified as survival actually defaulted). The light-shaded (yellow-shaded) area on the left of the cut-off line represents false default and the dark-shade (gray-shaded) area on the right of the cut-off line represents false survival. Setting the null hypothesis to be H_0: DD < x represents default where x is an arbitrary cut-off threshold, we are able to relate the Type-I error (α) to false survival and the Type-II error (β) to false default. To compare different models, we can control the Type-II error at a certain level and compare the Type-I errors generated by the models.

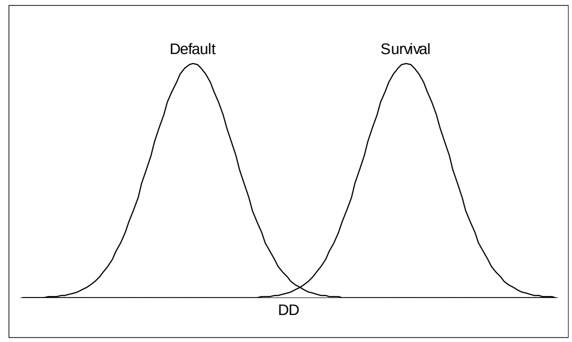
[13] The idea of ROC is similar to the Lorenz curve which is to measure income disparity. If everyone makes equal income, then it is a straight line where everywhere on the line shows equal income percentage and population percentage. For example, 25% population earns 25% income.

[14] These are also called "Sensitivity" and "Specificity" respectively in the medical research.

[15] For an intuitive introduction, see http://www.anaesthetist.com/mnm/stats/roc.

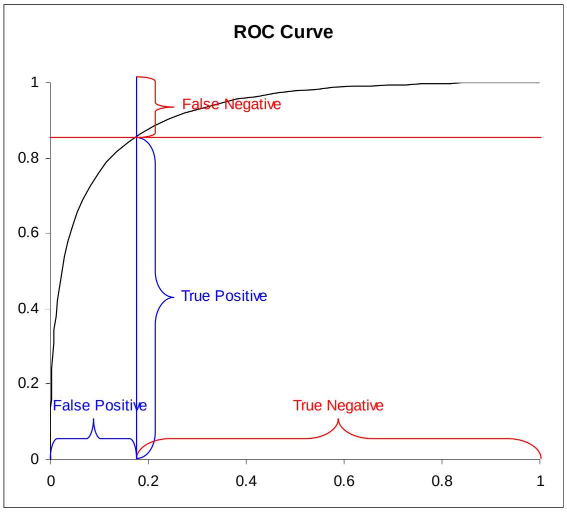
The following diagrams depict what a good and a bad model are. The model on the left generates DDs that have a large overlap between default and survival samples. On the other hand, the model on the right generates DDs that are a lot more discriminative. As a result, it is easy to conclude that the model on the right is better than the model on the left. For the same Type-II error, the model on the right has a smaller Type-I error.

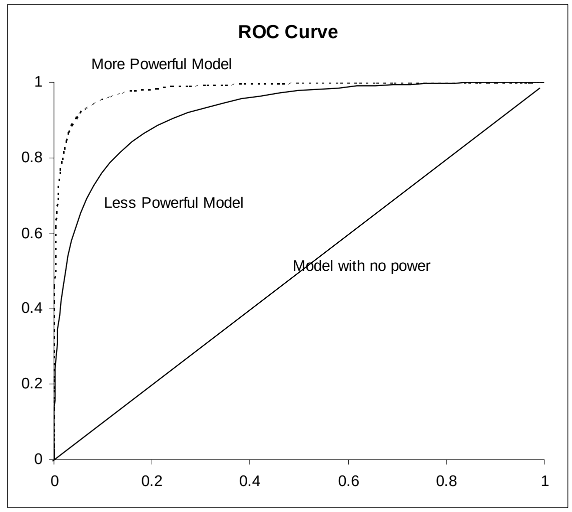




The ROC curve is a plot of probability of False Positive, or false default, and the probability of True Positive, or true default. Note that the survival sample contains true survivals and false defaults and the default sample contains true defaults and false survivals. Hence the probabilities of true survival and false default sum to unity (and similarly the probabilities of true default and false survival sum to unity.) The diagram on the left explains graphically the ROC curve.

The diagram on the right plots three different models. The more powerful model should be more concave because it predicts more defaults successfully than the less powerful model. A model that is not able to differentiate default from survival is a model that produces DDs that overlap the default sample and the survival sample completely (i.e. the two distributions are one distribution). In this case, the ROC curve is a 45 degree line. Certainly, a model can be ‘wrong’ in predicting default in that it gives defaulted firms higher DDs than survived firms. Such a model will have a convex ROC curve. In sum, the more powerful is the model that successfully set apart the two distributions, and then the more concave is the ROC curve.





The key statistic in the ROC methodology, which is also known as the Cumulative Accuracy Profile (CAP) or the Power Curve, is the Accuracy Ratio (AR) that is the area under the ROC curve. Since the ROC curve is generally concave, the more powerful is the model, the higher is the AR. At the end, we use the ARs of various models to test their prediction powers of default. We test 6 different models covering distinctly different assumptions in the paper. Note that the ROC curve is non-parametrical. Hence, it is ideal for the two distributions when they are non-parametrical.

Hanley and McNeil (1982) derive the standard error of AR. Note that their estimates for the standard error depend to a degree on the shapes of the distributions, but are conservative so even if the distributions are not normal, estimates of the standard error will tend to be a bit too large, rather than too small.

$$\sigma(A) = \sqrt{\frac{A(1-A) + (n_1 - 1)(Q_1 - A^2) + (n_2 - 1)(Q_2 - A^2)}{n_1 n_2}} \tag{8}$$

where A is the area under the curve, n_1 is the number of defaulted firms, n_2 is the number of survival firms, and

$Q_1 = \frac{A}{2-A}$

$Q_2 = \frac{2A^2}{1+A}$

To test two models if they are different from each other, the standard error of the difference of the two areas, $A_1 - A_2$, is:

$$\sigma(A_1 - A_2) = \sqrt{\sigma(A_1)^2 + \sigma(A_2)^2 - 2\rho\sigma(A_1)\sigma(A_2)} \tag{9}$$

where ρ is the quantity that represents the correlation induced between the two areas by the same study of the same set of cases. Assuming we have two models M1 and M2, that classify all firms under study into either default and survival and we have already calculated the areas under the ROC curves (A1 and A2). The procedure is as follows:

- Look at survived firms. Find how the two sets of DDs correlate for these firms, and obtain a value ρ_1 for this correlation;
- Look at the defaulted firms, and similarly derive ρ_2, the correlation between the two tests for these patients;
- Average out ρ_1 and ρ_2;
- Average out the areas A1 and A2,
- Use Hanley and McNeil's (1983) Table I to look up a value of ρ, given the average areas, and average of ρ_1 and ρ_2.

Finally, the test statistic is Gaussian:

$$z = \frac{A_1 - A_2}{\sigma(A_1 - A_2)} \tag{10}$$

In our ROC tests, we implement the following steps:

- Rank all DD's from smallest to greatest

  To calculate DD, we need the most recent data on the stock price, stock volatility, and accounting data. For example, if we want to calculate DD at the end of December 1987, we use the market capitalization at the end of December and the annualized standard deviation of stock return calculated from 52 weekly stock returns ending at the end of December. The accounting data is from the latest quarterly financial statements available to investors. We assume that financial statements will be available to investors three months after the end of a quarter. Therefore, we use financial statements from fiscal quarters ending in July, August, or September 1987. If any data items are missing, DD will be missing.

- Divide the entire region into 100 groups (z from 1 ~ 100)
- Divide the sample into default and survival groups

  If the forecast horizon is 1 month, then firms that go bankruptcy in the following month (it will be January 1988 if we use DD calculated at the end of December 1987) will fall in the default group. If the forecast horizon is 12 month, then firms that goes bankruptcy within the following 12 month (it will be January 1988 if we use DD calculated at the end of December 1987) will be in the default group.

- In the default group, z is divided into 100 percentiles and compute cumulative probability (y)
- In the survival group, z is divided into 100 percentiles and compute cumulative probability (x)
- Plot (x, y) curve
- Calculate AR

5. Results

A. General Results and the z Test

The DD results for the six models are reported in Table 6. The full sample size contains 1,047,227 observations that apply to all models but the Non Parametric model that uses 861,917 for the reasons mentioned earlier. The levels of the DDs across various models vary drastically. The distribution of the DD's for all the models are right-skewed for all the 12 forecasting periods (1 month forecast to 12 month forecast). The DDs are high for nearer forecast horizons and low for more distant forecast horizons, consistent with the fact that default is more likely for a longer period of time. The dispersions of the distributions (measured as the standard deviation) also decrease as the forecast period lengthens. All but two models demonstrate the similar pattern of dispersion. All but the Geske and the Non Parametric models start the standard deviation at low 60s for the one month forecast gradually reduce to about 17 for the twelve month forecast. For the Geske model, the dispersion is large for the short term forecast (612.81 for one month) and small for the long term forecast (17.10 for twelve month) and the Non Parametric model demonstrates very small dispersion throughout 12 forecasting periods. Note that smaller dispersion reduces the area of overlapping between the default sample and the survival sample introduced earlier, and in turn increases the ARs, and ultimately the power of the model. This demonstrates that the Non Parametric model has a higher precision in its DD measure.

The distribution of the DDs carries important information. The tails of the DD distributions show particular insights. Figure 1 plots the left 5% tails of all the models in our study. The top graph shows the distributions of the 1-month forecast and the bottom graph shows the 12-month forecast's. If the distribution has a heavier left tail (more probability weight for low DDs), it means that the model is more sensitive to the situations when the firm is vulnerable, and hence more effective in predicting defaults. From the top graph, the Non Parametric model clearly has a fatter tail than the rest of the models, with the Black-Cox model ranked the second. Then it is seen that all models become quite close at the 5% critical level. From the bottom graph, we see that the Black-Cox model substantially exceeds the other models and has the heaviest left tail. The Geske model is slightly better than the Non Parametric model at the extreme left but is overcomed by the Non Parametric model in the middle of the graph (about 2.5% level).

The main results of an ROC study include the accuracy ratio (AR), the Type-I (α) error, the Type-II (β) error, and the slope at the point where an DD percentile is chosen. Due to the extremely large amount of results as a result of our tests, we must omit many detailed numbers to conserve space, which can be obtained from the authors on request.

The plots of the ROC curves are given in Figure 2. The ROC curves for various models have little intersection, implying that the model with higher AR will have a lower Type-I error given any fixed cut-off level.[16] The ARs presented in the Table are also plotted in Figure 3. The numbers are decreasing as the forecast period lengthens to reflect the fact that farther forecasts are less accurate than nearer term forecasts. Note that while we plot all models, the Non Parametric model is estimated using a smaller sample size and strictly speaking is not comparable to other models. We can casually compare the Non Parametric model to other models, but for a formal test, we must compute the DDs and ARs in a common dataset, which we shall do later. The ARs for the Merton model are close to what have been reported in the literature. It is above 80% for short term forecasts and drops to below 80% as the forecast period lengthens. The other models improve upon the Merton model are expected to perform better. The Longstaff-Schwartz, Geske, and Flat Barrier models have ARs all above 80% and the Black-Cox model above 85%.

As mentioned earlier, a powerful model should have small Type-I errors after controlling for the Type-II error. The Type-I error is the error that we falsely reject the null hypothesis that the model successfully predicts a default (i.e. false survival) and the Type-II error is the error that we falsely accept the null hypothesis (i.e. false default). As we can see, for most of the models, if we control the Type-II error (false default) at 20%, the Type-I errors (false survival) are at least 20% (with the only exception being the Non Parametric model for short term forecasts). This provides a clear idea how in general the structural models perform in predicting defaults, even though there exists no formal test. As the Type-II error is more

[16] When two ROC curves intersect, the one with a higher AR may not be better.

tightly controlled (at 10%), the Type-I error increases, yet the general pattern remains. Figure 4 plots the Type-I errors of various models for various forecast periods under $\beta = 20\%$. Similar to the AR plots (but more pronounced), all except for the Merton and the Non Parametric models present humped shaped Type-I error curves. The Non Parametric model performs the best for the short term forecasts and the Black-Cox model performs the best for the long term forecasts.

The slope of a ROC curve at a Type-II error level also indicates the power of the model. The steeper is the slope, the more powerful is the model in that the ROC curve is more concave. Hence, we observe consistently with the ARs that the slope drops as the forecast period lengthens.

From Figure 3, it is clear that the power of the models is in the following order: Black-Cox, Geske, flat barrier, Longstaff-Schwartz, and Merton, although there is slight variation from month to month. In general, from Figure 3, we conclude that there are significant differences between Black-Cox and Geske, and Geske and the rest of the models; also that Longstaff-Schwartz is insignificantly different from the flat barrier model although the Longstaff-Schwartz model is better, and finally all of them are significantly better than the Merton model. The tables show detailed testing results, broken down by month.[17] To summarize, from 3-month forecast to 7-month forecast, we have the following ranking

(NP better than) Blkcox $\succ$ Geske $\succ$ Barrier $\approx$ Longsch $\succ$ Merton

where '$\approx$' represents insignificantly better and '$\succ$' represents significantly better, and the significance level is set at 5%. We see that the Black-Cox model significantly outperforms the Geske model (t values are 6.90, 8.50, 9.25, 10.81, 11.74 for three-, four-, five-, six-, and seven-month forecasts respectively) which in turn significantly outperforms the Flat Barrier model (t values are 7.62, 8.23, 9.29, 10.16, 10.97 for three-, four-, five-, six-, and seven-month forecasts respectively). However, the Flat Barrier and the Longstaff-Schwartz models are very close to each other, although the Flat Barrier model has a slightly higher AR (accuracy ratios are 83.65%, 83.97, 83.73, 83.47, and 83.11 for three-, four-, five-, six-, and seven-month forecasts respectively) than Longstaff-Schwartz (83.57%, 83.58, 83.48, 83.27, 83.05 for three-, four-, five-, six-, and seven-month forecasts respectively). Both the Flat Barrier model and the Longstaff-Schwartz model significantly outperforms the Merton

[17] The formal z-tests similar to Table 8 can be performed in pairs of these models and the results are omitted from the paper for the consideration of space. These results are available upon request.

model (t values are 4.58, 8.63, 11.84, 15.29, and 18.12 for three-, four-, five-, six-, and seven-month forecasts respectively).

From 8-month forecast to 11-month forecast, the Black-Cox model outperforms the Non Parametric model and the remaining order is remained as follows:

Blkcox (better than NP) $\succ$ Geske $\succ$ Barrier $\approx$ Longsch $\succ$ Merton

For the 12-month forecast, the Longstaff-Schwartz model significantly outperforms the Flat Barrier model:

Blkcox (better than NP) $\succ$ Geske $\succ$ Longsch $\succ$ Barrier $\succ$ Merton

For the 1-month forecast, the Black-Cox model is not significantly better than the Geske model and the Longstaff-Schwartz, Flat Barrier, and Merton models are close to one another:

(NP better than) Blkcox $\approx$ Geske $\succ$ Longsch $\approx$ Barrier $\approx$ Merton

For the 2-month forecast, the Black-Cox model becomes marginally significantly better than the Geske model (t value 3.21):

(NP better than) Blkcox $\succ$ Geske $\succ$ Barrier $\approx$ Longsch $\approx$ Merton

We see that 11 out of 12 months (except for January) the Black-Cox model is significantly better than the Geske model which is always significantly better than the Longstaff-Schwartz model.

The Longstaff-Schwartz model is insignificantly better than the Flat Barrier model in all months which is better than the basic Merton model in 11 out of 12 months (except for January). The fact that the Longstaff-Schwartz model is not significantly better than the Flat Barrier model (more clearly seen from Figure 3) echoes the result by Wei and Guo (1997) that the Longstaff-Schwartz model and the Merton model are not nested and the recovery assumption (richer in the Merton and barrier models) and random interest rates (richer in the Longstaff-Schwartz model) have an obvious tradeoff. In pricing credit derivatives, random interest rates are more important, as found by Eom, Helwedge, and Huang (2004). In our study to forecast default, random interest rates are not as important, and hence the Longstaff-Schwartz model is no better than the barrier model, both of which assume continuous default.

The Black-Cox model outperforms uniformly the Geske model that in turn outperforms other models, indicating that taking a more flexible default barrier improves the default prediction power. This also indicates that the specification of the default barrier is more important than the specification of the interest rate dynamics.

Note that the Non Parametric model uses a different data period than the other models. Hence the ARs under the Non Parametric model are not comparable with those of the other models. The z-test of the Non Parametric model and the other models must be performed under a common dataset. Using its own dataset, the Non Parametric model is superior to the other models substantially, except for toward the end of the year that the Non Parametric model is outperformed by the Black-Cox model.

To perform the z test on all models, we limit our data set to a common sample of 965,680, reduced from 1,047,727, due to the data requirement of the Non Parametric model. The detailed results are repetitive and hence omitted for the conservation of space (and available from the authors on request). The AR results are plotted in Figure 5 which is similar to Figure 3. However, with the smaller sample set, we can perform statistical tests to formally investigate the default prediction powers of these models.

The ARs and the z-test t statistics are reported in Table 8. They are similar to the results in the whole sample except that we find the Non Parametric model outperforms the Black-Cox model significantly under 1-month through 5-month forecasts. In the 6-month forecast, the Non Parametric model is still better than the Black-Cox model but becomes insignificant (t value 1.43). Different from the result in Table 7, in the 7-month forecast, the Non Parametric model performs worse in the common sample. Beginning from the 8-month forecast, the Black-Cox model performs better than the Non Parametric model significantly and the significance level increases as we forecast over longer horizons (t values increase from 1.99 to 3.7 monotonically.)

We also note other differences between Table 7 and Table 8. In the 4-month forecast, the Flat Barrier model is significantly better than the Longstaff-Schwartz model in the common sample and yet it becomes insignificant in the full sample.

B. Industrials versus Financials

It is commonly known that financial firms cannot be modeled well in that the financial sector is highly regulated and its financial information (e.g. high leverage ratios) is industry specific. While no formal academic literature to be found, there has been plentiful industry evidence that shows the Merton model cannot do a good job in predicting defaults in the financial industry.[18] We confirm this result in our study. Our results of the various models in studying the financial firms versus non-financial firms are as follows. The sample size for

[18] See footnote 6.

this test is 965,670 in which 811,278 are non-financial firms and 154,402 are financial firms. In Table 9, we see that the ARs of the Merton model for the Industrial sector are uniformly better than those for the financial sector. Unfortunately, due to the sample difference, we cannot perform the z test. Besides the Merton model, we find that the Longstaff-Schwartz model also performs better in the Industrial sector than in the Financial sector and the differences are substantial and possibly significant. And the Black-Cox model performs better in the Industrial sector only in the long term forecasts (9 month through 12 month forecasts).

What is interesting in our study is that we find the Non Parametric, Flat Barrier, and Geske models perform better in the Financial sector than in the Industrial sector. In the Financial sector, we find that the Non Parametric model is superior substantially to the rest of the models, in particular the Black-Cox model that it was not able to dominate in the full sample (Table 8). Also note that in the Financial sector, toward long term forecasts, the Black-Cox model becomes worse than the Geske model and the Flat Barrier model.

Note that all of the models that do better in the Financial sector (Non Parametric, Flat Barrier, and Geske models) assume endogenous recovery. This implies that if we can capture better the information implied in the capital structures of the financial firms, we have better chances to predict defaults more successfully. Those assume exogenous default barriers (Longstaff-Schwartz and Black-Cox) do not have the flexibility to capture the true information in the capital structures of the financial firms and consequently cannot work well (the Black-Cox model has an exponential barrier and hence does better in some cases.) Furthermore, the fact that the Flat Barrier model is better than the Longstaff-Schwartz model significantly in the Financial sector (but worse significantly in the Industrial sector) supports strongly the importance of endogenous recovery (and the unimportance of random interest rates.)

In the following, we summarize the z test results for both sectors without presenting the detailed t statistics. Again, "$\succ$" represents significantly better at the 5% level and "$\approx$" represents insignificantly better.

In the Financial sector, the Non Parametric model outperforms all the other models significantly. The Geske model ranks the next except for the 2 month and 3 month forecasts where the Black-Cox model is better but insignificant. The remaining three models (Longstaff-Schwartz, Flat Barrier, and Merton) rank toward the last. For the 1, 3, 4, and 5 month forecasts, the Geske model is better than Black-Cox model but insignificant:

$$\text{NP} \succ \text{Geske} \approx \text{Blkcox} \succ \text{Barrier} \succ \text{Merton} \succ \text{Longsch}$$

For the 2 and 3 month forecasts, the Black-Cox model is better than the Geske model but insignificant:

$$NP \succ Blkcox \approx Geske \succ Barrier \succ Merton \succ Longsch$$

For the 6 and 7 month forecasts, the Geske model becomes significantly better than the Black-Cox model (which is insignificantly better than the Flat Barrier model):

$$NP \succ Geske \succ Blkcox \approx Barrier \succ Merton \succ Longsch$$

For the 8 month forecast, the Flat Barrier model becomes better than the Black-Cox model but insignificant. Also for 8 month forecast the Merton model no longer significantly outperforms the Longstaff-Schwartz model:

$$NP \succ Geske \succ Barrier \approx Blkcox \succ Merton \approx Longsch$$

For the rest of the forecast periods (9 through 12 months), the Longstaff-Schwartz model outperforms the Merton model but insignificantly:

$$NP \succ Geske \approx Barrier \succ Blkcox \succ Longsch \approx Merton$$

In the Industrial sector, the results are different. The Non Parametric model no longer dominates all the other models. In particular, the Non Parametric model dominates the other models in forecasting default up to 5 months and the Black-Cox model performs most superiorly for forecast over 6 months. The Geske model is a clear #3 in performance ranking except for forecasts of 11 and 12 months where the Geske model is ranked #2. It is interesting to note that, unlike the results for the Financial sector, the Longstaff-Schwartz, Flat Barrier, and Merton models are not significantly worse than the Geske model (except for the long term forecasts of the Merton model).

For the 1 month forecast, we have:

$$NP \succ Blkcox \approx Longsch \approx Geske \approx Barrier \approx Merton$$

For the forecasts between 2 and 10 months, the Geske model becomes better than the Longstaff-Schwartz model (but insignificantly). Yet, for the 2 and 3 month forecasts, the Non Parametric model remains significantly better than the Black-Cox model while for the 4 and 5 month forecasts, the Non Parametric model is no longer significantly better than the Black-Cox model. On the other hand, the Black-Cox model starts to be significantly better than the Geske model in the 3 month forecast. In summary, for the 2 and 3 month forecasts, we have:

$$NP \succ Blkcox \approx Geske \approx Longsch \approx Barrier \approx Merton$$

and for the 4 and 5 month forecasts, we have:

$$NP \approx Blkcox \succ Geske \approx Longsch \approx Barrier \approx Merton$$

Starting from the 6 month forecast, the continually improved Black-Cox model becomes a better predictor than the Non Parametric model. For the 6 month forecast it is insignificant but after which it is significantly better than the Non Parametric model. Also noticeable is that the Geske model also improves and becomes better than the Non Parametric model starting the 11 month forecast. To summarize, for the 6 month forecast, we have:

$$Blkcox \approx NP \succ Geske \approx Longsch \approx Barrier \approx Merton$$

but the significance changes for the 7, 8, 9 and 10 month forecasts as follows. For the 7 and 8 month forecasts, we have:

$$Blkcox \succ NP \succ Geske \approx Longsch \approx Barrier \approx Merton$$

and for the 9 and 10 month forecasts, we have:

$$Blkcox \succ NP \approx Geske \approx Longsch \approx Barrier \approx Merton$$

Finally, for the 11 and 12 month forecasts, the Geske model outperforms (but insignificantly) the Non Parametric model to become the second best model next to Black-Cox.

$$Blkcox \succ Geske \approx NP \approx Longsch \approx Barrier \approx Merton$$

The ARs are reported in Table 9 and yet the results of the z test are omitted from the paper (available upon request) for the sake of space conservation.

C. Investment Grades versus Speculative Grades

Similar to the previous section, another widely accepted story when we use a structural model is that structural models are more effective in pricing the credit risk in high yield companies than in high grade companies. For high grade companies, their stock prices are believed to reflect more market risk (beta risk) than credit risk since default is remote. Since the structural models rely so heavily on the equity information, for the models to be effective in pricing the credit risk, the equity prices must reflect substantial credit information. Hence, it is expected that structural models should perform better in low equity prices. Again, the ARs are provided in Table 9 and the z-tests are omitted for the sake of space conservation and we summarize the z test below.

The sample size drops significantly once we include ratings. We have only 196,530 total sample size in which 109,146 are investment grade and 87,384 are speculative grade. Our findings confirm the previous literature that the equity prices of lower grade companies carry more credit information than those of higher grades. The ARs for investment grade firms drop substantially for all the models. All models drop to low 60% for twelve month forecast. The relative rankings are given below.

In the Investment grade subgroup, in general we cannot differentiate models in terms of their default prediction. For the most part, except for the Non Parametric model, no model is significantly better than the other model. Furthermore, in short term forecast situations, models have identical ARs. For the 1 month forecast, the Non Parametric model has an AR of 86.1% which is significantly better than the Merton, Longstaff-Schwartz, Geske, and Flat Barrier models that have identical AR of 75.3%. The Black-Cox model ranks the last but is indistinguishable from the four models (AR is 75.1%).

For the 2 month forecast, the Black-Cox model is no longer last. It performs better than the Merton model (ARs of 67.59% and 67.41% respectively). Yet none of them is insignificantly outperformed by the Longstaff-Schwartz, Geske, and Flat Barrier model that have identical AR of 67.77%. The Non Parametric model remains significantly better than then rest of the models (AR is 80.87%).

Between the 3 and 5 month forecasts, the results remain similar that the Non Parametric model outperforms significantly all the other models and the Merton ranks the last while the relative rankings of the Longstaff-Schwartz, Geske, and Flat Barrier model vary.

Starting from the 8 month forecast, the performances of the models become more distinguishable and yet still except for the Non Parametric and Merton models the relative performance is insignificant. Between 8 and 11 month forecasts, we have:

$$\text{NP} \succ \text{Blkcox} \approx \text{Geske} \approx \text{Barrier} \approx \text{Longsch} \succ \text{Merton}$$

For the 12 month forecast, the rankings of the Longstaff-Schwartz and Flat Barrier models switched:

$$\text{NP} \succ \text{Blkcox} \approx \text{Geske} \approx \text{Longsch} \approx \text{Barrier} \succ \text{Merton}$$

In the speculative grade category, models show differential predicting power of default. The Non Parametric model dominates the rest of the models significantly. The Black-Cox and Geske models perform well in the short term forecasts. The Merton model performs the worst.

For the 1 month forecast:

$$\text{NP} \succ \text{Geske} \approx \text{Blkcox} \approx \text{Barrier} \approx \text{Merton} \approx \text{Longsch}$$

For the 2 and 3 month forecasts, the Black-Cox model performs better than the Geske model but still insignificant. The difference in relative rankings for these two short term forecasts is that for the 2 month forecast the Merton model is still better than the Longstaff-Schwartz model:

$$\text{NP} \succ \text{Blkcox} \approx \text{Geske} \approx \text{Barrier} \approx \text{Merton} \approx \text{Longsch}$$

but for the 3 month forecast, the Longstaff-Schwartz model performs better (while insignificant) than the Merton model:

$$\text{NP} \succ \text{Blkcox} \approx \text{Geske} \approx \text{Barrier} \approx \text{Longsch} \approx \text{Merton}$$

From the 4 month forecast and onwards, the Merton model becomes significantly worse than the Longstaff-Schwartz model. Starting from the 5 month forecast, we have the following ranking order:

$$\text{NP} \succ \text{Blkcox} \approx \text{Geske} \approx \text{Barrier} \approx \text{Longsch} \succ \text{Merton}$$

Comparing the ARs of the Investment grade and the Speculative grade, we confirm the widely accepted story that the structural models should predict defaults more effectively for low quality companies with all the models in our study. Not only do we observe uniform better performances across all the models, the differences can be as large as over 20% (e.g. 12-month forecast, Flat Barrier model, 82.32% for Speculative grade versus 62.02% for Investment grade.) It is also important to note that the differences of any model between the two groups are larger for longer term forecasts than for shorter term forecasts. This verifies the story that the equity prices of Investment grade companies carry mostly market information and little credit information and the credit information carried in the equity price becomes even less as the forecast horizon lengthens. Lastly, unlike the previous section, the Non Parametric model is the dominant model in all cases.

D. Regression Test

While the AR test is already an out-of-sample test, it does not capture the dynamics of the DDs in regard to actual defaults. To study this dynamic effect, we perform two additional tests. One is the regression test where we regress the DDs of each model on default along with a number of control variables. The other is that we run the AR test year by year. This test allows us to see how each model performs over time.

In the regression test, we regress DD against the following independent variables:

- Bankrupt – This is the main variable we study. It is a dummy variable with 1 for default and 0 for survival. We expect to see a better model with a more significant tie to this variable,
- Size – This is market cap which is computed as outstanding shares times equity price,
- Leverage – This is computed as book value debt over book value debt and market cap,
- S1 ~ S9 – These are dummy variables representing each of the SIC code.

We perform this regression test on each default forecast horizon (i.e. 1 month to 12 months). To conserve space, we report only the results of the shortest (1 month) and longest (12 months) horizons in Table 10 while leaving other results available on request.

Because DD is closely related to important financial variables such as Leverage and Size, by controlling for these financial variables, we hope to achieve a clear relationship between DD and default (Bankrupt). In both 1 month and 12 months results, it is clear that all models can significantly relate DD and Bankrupt (except for 1 month Geske). The coefficients of the Bankrupt variable are close to one another (–5 or so for 1 month and –1.4 or so for 12 months). The significant levels are also similar to one another except for the Non Parametric model which has over ten times more t statistic than the other models (which reflects the low standard deviation of the Non Parametric model).

As for the control variables, industry sectors (SIC dummy variables) are generally insignificant, indicating that the DD variable is indiscriminant across industry sectors. On the other hand, Size and Leverage are highly significant. After controlling for these variables, the DD variable of all but the Non Parametric models shows marginal importance for the 1 month forecast (i.e. the t statistic of the Bankrupt variable is slightly above 2). However, the DD variable does show significant importance for the 12 month forecast.

The Non Parametric model clearly dominates all the other models. For the 1 month forecast, its t statistic of the Bankrupt variable is –31.53 and it is –87.22 for the 12 month forecast.

E. Robustness Test

In addition to the regression test, we also perform the AR test on the year by year basis. The purpose of this test is to examine if the result we obtain by using the entire 18 years of data holds on a year by year basis. The results of this analysis are reported in Figure 6. By examining the AR plots for 1 month and 12 month forecasts (while the other plots for 2 ~ 11 months are left out to conserve space and available upon request), we see evidently that the Non Parametric model indeed dominates all the other models for 1 month forecast throughout all 18 years and the Black-Cox model dominates all the other but the Non Parametric models throughout 18 years. The ranking of the other models (Flat Barrier, Longstaff-Schwartz, Geske and Merton models) also generally confirm the result we observe by using the entire sample.

While not significant, the ARs do increase in general over time. The Non Parametric model, for example, increases from 81.52% in 1985 to 96.7% in 2002 for the 1 month forecast and the Merton model increases from 77.85% to 88.96%. Similarly, for the 12 month forecast, the Non Parametric model increases from 77.49% to 90.03% and the Merton model climbs from 69.79% to 83.5%. Other models present a similar upward pattern. This could be due to a number of reasons. It could be the financial data better represent the true credit information of the companies; or the cross market efficiency has improved and hence the equity prices carry more accurately credit information.

Finally, we should note that in the year to year analysis, the number of defaulted firms is divided up by the number of years and hence the number of defaults in each year decreases considerably resulting the z tests to be mostly insignificant. Therefore, the results in Figure 6 are for qualitative examination.

6. Conclusion

In this study, we compare the powers of various structural models in predicting defaults. In particular, we measure the power of each model by its capability to separate default firms from non-default firms. The measure used across all models is the Distant to Default, or commonly known as DD. We adopt the ROC methodology that provides an easy testing statistic on the accuracy ratio (AR) and allows us to compare different models. In general, the rankings are: Non Parametric, Black-Cox, Geske, Longstaff-Schwartz, Flat Barrier, and Merton when the forecast periods are short and Black-Cox, Non Parametric, Geske, Longstaff-Schwartz, Flat Barrier, and Merton when the forecast period is long. However, the Longstaff-Schwartz model is not significantly better than the Flat Barrier model in most cases.

We have a number of interesting results. First, the fact that the Longstaff-Schwartz model is not significantly superior to the Flat Barrier model indicates that random interest rates are not an important factor default prediction. This contrasts the findings in pricing credit derivatives in which interest rate dynamics play an important role. Furthermore, both the Longstaff-Schwartz and the Flat Barrier models do not outperform the simple Merton model in the near term default forecast. This is expected in that continuous default should have a larger impact as the prediction horizon lengthens.

Second, the Non Parametric model, which is a single period default model as Merton, performs better than all other models except for the Black-Cox model in long horizons. This indicates that distribution assumption is extremely important for the short term. The known fact of skewness and fat tails in equity returns for the short horizon plays an exceptionally important role in capturing the credit risk of the company. As the forecast period lengthens, the equity returns gradually approach normality, the advantage of the Non Parametric model disappears and the other factors of the default model, such as multi-period default and exponential barrier become relatively important.

Third, All except for the Merton and the Non Parametric models (both permit only single period default) present a non-monotonic pattern in predicting defaults for different time horizons. This is clear evidence of the importance of multi-period default. In Merton and Non Parametric, for various default prediction horizons, a similar single period computation is applied. Hence, as the period lengthens, the prediction power diminishes monotonically. On the other hand, the multi-period default models consider the interaction between asset price and barrier over time and is better able to capture the default likelihood over time.

Fourth, the fact that the two period Geske model is significantly better than the Longstaff-Schwartz model and the Longstaff-Schwartz model is not significantly better than the Flat Barrier model supports the importance of endogenous recovery, since both the Geske model and the Flat Barrier model set recovery to be the asset value upon default while the Longstaff-Schwartz model has a fixed recovery.

Fifth, we test the models in sub-samples to compare to previous findings. We find that not all models perform better in the non-financial sector than the financial sector, contrasting the previous findings that financial firms are harder to model due to the fact that it is a highly regulated industry and its financial information is industry specific. On the other hand, we confirm that all models perform better in the lower grades than in higher grades, agreeing with the previous literature that low equity price contains more credit information.

Finally, the models we study in this paper are amazingly robust to sample selection, proving the point that structural models are good for default prediction. The relative model performances in full sample as well as various sub-samples are very close.

7. References

Altman, E. I., 1968, "Financial ratios, Discriminant analysis, and the prediction of corporate bankruptcy," Journal of Finance 23, 589-609.

Bélanger, Alain, Steven Shreve, and Dennis Wong, 2002, "A Unified Model for Credit Derivatives," Mathematical Finance, 14 (3), 317-350.

Bharath, Sreedhar T. and Tyler Shumway, 2004 (December), "Forecasting Default with the KMV-Merton Model," working paper, University of Michigan.

Black, Fisher, and John Cox, 1976, "Valuing Corporate Securities: Some Effects of Bond Indenture Provisions," Journal of Finance, 31, 351-367.

Delianedis, Gordon and Robert Geske, 2003, "Credit Risk and Risk Neutral Default Probabilities: Information about Rating Migrations and Defaults," working paper, ULCA.

Eom, Young-Ho, Jean Helwege and Jing-zhi Huang, 2004, "Structural Models of Corporate Bond Pricing: An Empirical Analysis," Review of Financial Studies, 17, 499-544.

Geske, Robert, 1977, "The Valuation of Corporate Liabilities as Compound Options," Journal of Financial and Quantitative Analysis, 541-552.

Hanley, J.A. and B.J. McNeil, 1982, "The meaning and use of the area under the Receiver Operating Characteristic (ROC)," Radiology, 143, 29-36.

Hanley, J.A. and B.J. McNeil, 1983, "The Method of Comparing the Areas under Receiver Operating Characteristic Curves Derived from the Same Cases," Radiology, 148, 839-843.

Hull, John, 2004, Option, Futures, and Other Derivatives, Hoboken, NJ: Prentice Hall.

Longstaff, Francis, and Eduardo Schwartz, 1995, "A Simple Approach to Valuing Risky Fixed and Floating Rate Debt," Journal of Finance, 50, 789-819.

Madan, Dilip B; Haluk Unal, 2000 (March), "A Two-Factor Hazard Rate Model for Pricing Risky Debt and the Term Structure of Credit Spreads," Journal of Financial and Quantitative Analysis, 35 (1), 43-65.

Merton, Robert C., 1974, "The Pricing of Corporate Debt: The Risk Structure of Interest Rates," Journal of Finance, 29, 449–470.

Ohlson, J. S., 1980, "Financial Ratios and the Probabilistic Prediction of Bankruptcy," Journal of Accounting Research 19 109-31.

Shumway, Tyler, 2001, "Forecasting Bankruptcy More Accurately: A Simple Hazard Model," Journal of Business, 101-124.

Stein, Roger M., 2002, "Benchmarking Default Prediction Models: Pitfalls and Remedies in Model Validation," Moody's KMV white paper.

Stein, Roger M., 2005, "The relationship between default prediction and lending profits: Integrating ROC analysis and loan pricing" Journal of Banking and Finance, 29, 1213-1236.

Vassalou, Maria and Yuhang Xing, 2004, "Default Risk in Equity Returns," Journal of Finance, 59 (2), 831.

Wei, David, and Dajiang Guo, 1997, "Pricing Risky Debt: An Empirical Comparison of Longstaff and Schwartz and Merton Models," Journal of Fixed Income, 7, 8-28.

Zhou, Chunsheng 1997, "A Jump-Diffusion Approach to Modeling Credit Risk and Valuing Defaultable Securities," working paper, the Federal Reserve Board.

Table 1: Financial Data Description – Number of Bankrupt Firms that Have Financial Data Prior to Default

Year	# of samples	# of total firms	# of defaulted firms												
			# of months prior to default financial information is available												
			1	2	3	4	5	6	7	8	9	10	11	12	actual
1985	48,383	4,428	18	18	18	18	18	18	18	18	18	18	18	18	18
1986	47,376	4,412	26	26	27	28	29	29	29	30	30	30	30	30	30
1987	49,628	4,601	11	11	12	12	13	13	13	13	13	13	13	13	15
1988	51,303	4,675	27	28	28	29	29	29	29	29	29	29	29	29	29
1989	49,547	4,474	41	42	43	44	46	46	46	46	46	46	46	46	46
1990	48,327	4,392	48	49	52	52	53	54	54	54	55	55	55	55	59
1991	48,179	4,303	49	49	51	53	53	53	53	53	53	54	54	54	64
1992	48,615	4,473	37	37	37	37	37	37	37	37	37	37	37	37	51
1993	51,629	4,701	26	27	27	28	29	29	29	29	29	29	29	29	39
1994	61,457	5,741	18	18	18	18	18	19	19	19	19	19	19	19	25
1995	66,165	5,993	29	30	30	30	30	30	30	30	30	30	30	30	32
1996	68,190	6,333	27	28	28	29	29	29	29	29	29	29	29	29	34
1997	73,166	6,665	32	32	32	32	32	32	32	32	32	32	32	32	32
1998	73,010	6,702	56	57	58	59	59	59	59	59	59	59	59	59	59
1999	69,845	6,391	61	62	63	64	65	66	66	66	66	66	66	66	66
2000	67,969	6,303	87	90	93	94	94	94	94	94	94	94	94	94	94
2001	65,044	5,910	144	148	148	148	148	148	148	148	149	149	149	149	166
2002	59,894	5,319	99	102	104	104	104	104	104	104	104	104	104	104	138
2003	-	-	11	17	26	35	40	46	49	50	54	55	56	56	94
Total	1,047,727		847	871	895	914	926	935	938	940	946	948	949	949	1091

Note: Sample period is January 1985 through December 2002. For example, in 2002, there were 99 bankrupt companies that had financial data a month before bankruptcy and 104 bankrupt companies that had financial data 12 months prior to bankruptcy. Since our data ends in 2002, the 11 defaulted companies in 2003 that should have financial data 1 month before must occur in January 2003. This table also shows the number of firms that went bankrupt in a particular year AND have financial data *n* months prior to default. For example, there are 19 firms that went bankrupt in 1994 and have financial data 6 months prior to default..

Table 2: Market Cap = stock price × outstanding shares (Millions)

N	Mean	Std Dev	Minimum	Maximum
1,047,727	1,223.26	8,289.91	0.0289	602,432.92

Table 2-1: Group by Market Cap

Mean of Size (Millions)		# of bankrupt (1 month)	# of bankrupt (12 month)
0.03 ~	7.08	422	499
7.08 ~	15.41	162	179
15.41 ~	28.39	93	86
28.39 ~	49.30	55	69
49.30 ~	85.47	38	34
85.47 ~	153.86	24	29
153.86 ~	299.52	23	29
299.53 ~	656.03	17	12
656.05 ~	1,921.57	8	7
1,921.65 ~	602,432.92	5	5
		847	949

Table 3: Leverage = total liability/(stock price×outstanding shares+ total liability)

N	Mean	Std Dev	Minimum	Maximum
1,047,727	0.42	0.28	-0.02	1.00

Table 3-1: Group by leverage

Leverage	# of bankrupt (1 month)	# of bankrupt (12 month)
0.0347947	6	42
0.1005334	6	46
0.1756098	8	43
0.2573654	12	46
0.3478355	12	59
0.4454404	35	80
0.5457453	31	93
0.6559855	70	122
0.7934137	124	211
0.9120219	543	207
Total	847	949

Table 4: Group of bankrupt firms by SIC code

Division	0	1	2	3	4	5	6	7	8	9	Total
# of bankrupt firms (1 month)	2	62	86	194	108	146	74	127	44	4	847
# of bankrupt firms (12 month)	3	67	102	218	119	164	78	142	49	7	949

Division 0: Agriculture, Forestry, And Fishing
Division 1: Mining, and Construction
Division 2, 3: Manufacturing
Division 4: Transportation, Communications, Electric, Gas, And Sanitary Services
Division 5: Wholesale Trade, Retail Trade
Division 6: Finance, Insurance, and Real Estate
Division 7,8: Services
Division 9: Public Administration

Table 5: Group by Credit rating

S&P credit ratting	code	Non-default(#)	Default(#)	Default+ Non Default	Default (%)
AAA	2	3,571	0	3,571	0.00%
AA+	4	1,835	0	1,835	0.00%
AA	5	7,880	0	7,880	0.00%
AA-	6	8,139	13	8,152	0.16%
A+	7	13,824	11	13,835	0.08%
A	8	22,462	0	22,462	0.00%
A-	9	16,896	0	16,896	0.00%
BBB+	10	16,637	46	16,683	0.28%
BBB	11	20,088	32	20,120	0.16%
BBB-	12	13,824	24	13,848	0.17%
BB+	13	10,273	35	10,308	0.34%
BB	14	13,505	85	13,590	0.63%
BB-	15	16,178	220	16,398	1.34%
B+	16	18,931	412	19,343	2.13%
B	17	7,780	370	8,150	4.54%
B-	18	4,010	334	4,344	7.69%
CCC+	19	1,576	234	1,810	12.93%
CCC	20	738	150	888	16.89%
CCC-	21	439	50	489	10.22%
CC	23	120	110	230	47.83%
D	27	598	363	961	37.77%
Not Meaningful	28	140	78	218	35.78%
SD	29	40	17	57	29.82%
Total Sample Size (=202,066)		199,482	2,584	202,066	1.28%

Table 6: Distance to Default

Merton (sample size: 1047727)

	Mean	Std	Median
1	9.34	61.13	7.32
2	6.57	43.24	5.16
3	5.34	35.30	4.20
4	4.60	30.57	3.62
5	4.01	89.95	3.23
6	3.72	24.95	2.93
7	3.43	23.10	2.71
8	3.19	21.61	2.52
9	2.99	20.37	2.37
10	2.83	19.33	2.23
11	2.68	18.43	2.12
12	2.55	17.65	2.02

Black-Cox (sample size: 1047727)

	Mean	Std	Median
1	9.27	60.79	7.29
2	6.46	42.75	5.10
3	5.21	34.70	4.12
4	4.45	29.88	3.54
5	3.93	27.72	3.15
6	3.54	24.13	2.87
7	3.23	22.21	2.64
8	2.98	20.69	2.44
9	2.77	19.43	2.26
10	2.59	18.32	2.11
11	2.44	17.40	1.97
12	2.30	16.59	1.84

Longstaff-Schwartz (sample size: 1047727)

	Mean	Std	Median
1	9.31	60.80	7.29
2	6.53	42.76	5.12
3	5.28	34.71	4.15
4	4.54	29.90	3.57
5	4.02	26.59	3.17
6	3.64	24.15	2.87
7	3.34	22.23	2.64
8	3.09	20.70	2.45
9	2.89	19.44	2.29
10	2.72	18.36	2.16
11	2.57	17.42	2.04
12	2.44	16.61	1.94

Flat Barrier (sample size: 1047727)

	Mean	Std	Median
1	9.30	60.79	7.29
2	6.51	42.75	5.12
3	5.27	34.71	4.15
4	4.53	29.88	3.57
5	3.99	40.14	3.17
6	3.64	24.12	2.87
7	3.34	22.85	2.64
8	3.11	20.68	2.45
9	2.89	24.00	2.29
10	2.73	19.36	2.16
11	2.57	22.96	2.04
12	2.46	17.44	1.94

Geske (sample size: 1047727)

	Mean	Std	Median
1	7.44	612.81	7.29
2	5.52	389.64	5.12
3	5.02	184.46	4.15
4	4.31	159.72	3.57
5	4.01	26.81	3.17
6	3.63	24.35	2.87
7	3.33	22.41	2.63
8	3.09	20.87	2.45
9	2.88	19.59	2.29
10	2.71	18.46	2.16
11	2.56	17.51	2.04
12	2.43	17.10	1.94

Non Parametric (sample size: 861917)

	Mean	Std	Median
1	7.79	5.34	6.57
2	5.51	3.78	4.65
3	4.49	3.08	3.80
4	3.89	2.67	3.29
5	3.48	2.39	2.94
6	3.18	2.18	2.68
7	2.94	2.02	2.48
8	2.75	1.89	2.32
9	2.60	1.78	2.19
10	2.46	1.69	2.08
11	2.35	1.61	1.98
12	2.25	1.54	1.90

Table 7: Accuracy Ratios and z-tests (Full Sample)

1_month_forecast_	NP	blkcox	geske	longsch	barrier	merton
	91.08%	85.02%	84.52%	83.58%	83.40%	83.37%
	(NA)	(1.17)	(2.16)	(0.41)	(0.07)	
2_month_forecast_	NP	blkcox	geske	barrier	longsch	merton
	90.21%	86.27%	85.33%	83.86%	83.30%	82.97%
	(NA)	(3.21)	(4.87)	(1.82)	(1.06)	
3_month_forecast_	NP	blkcox	geske	barrier	longsch	merton
	89.52%	87.14%	85.52%	83.65%	83.57%	82.41%
	(NA)	(6.90)	(7.62)	(0.32)	(4.58)	
4_month_forecast_	NP	blkcox	geske	barrier	longsch	merton
	88.92%	87.41%	85.70%	83.97%	83.58%	81.68%
	(NA)	(8.50)	(8.23)	(1.82)	(8.63)	
5_month_forecast_	NP	blkcox	geske	barrier	longsch	merton
	88.23%	87.21%	85.48%	83.73%	83.48%	81.05%
	(NA)	(9.25)	(9.29)	(1.30)	(11.84)	
6_month_forecast_	NP	blkcox	geske	barrier	longsch	merton
	87.54%	87.07%	85.22%	83.47%	83.27%	80.39%
	(NA)	(10.81)	(10.16)	(1.14)	(15.29)	
7_month_forecast_	NP	blkcox	geske	barrier	longsch	merton
	86.96%	86.87%	84.87%	83.11%	83.05%	79.76%
	(NA)	(11.74)	(10.97)	(0.37)	(18.12)	
8_month_forecast_	blkcox	NP	geske	barrier	longsch	merton
	86.57%	86.37%	84.63%	82.90%	82.80%	79.17%
	(4.86)	(NA)	(4.15)	(0.65)	(21.23)	
9_month_forecast_	blkcox	NP	geske	barrier	longsch	merton
	86.10%	85.81%	84.27%	82.51%	82.51%	78.53%
	(12.00)	(NA)	(12.3)	(0.00)	(24.51)	

Table 7: Accuracy Ratios and z-tests (Full Sample) – continued

10_month_forecast_	blkcox	NP	geske	barrier	longsch	merton
	85.71%	85.23%	83.90%	82.24%	82.18%	77.95%
	(12.03)	(NA)	(12.15)	(0.43)	(26.41)	
11_month_forecast_	blkcox	NP	geske	barrier	longsch	merton
	85.12%	84.70%	83.57%	81.83%	81.82%	77.41%
	(11.02)	(NA)	(13.25)	(0.08)	(28.67)	
12_month_forecast_	blkcox	NP	geske	longsch	barrier	merton
	84.53%	84.18%	83.18%	81.42%	81.38%	76.85%
	(10.24)	(NA)	(13.88)	(3.11)	(31.50)	

Note: The z statistic is defined as follows:

$z = \frac{A_1 - A_2}{\sigma_{A_1 - A_2}}$ where A_i represents the area under the ROC curve (accuracy ratio) for model i.
For example, for 1-month forecast, the Black-Cox model that has an accuracy ration of 85.02% is better than then Geske model that has an accuracy ration of 84.52% with a t statistic of 1.17, which is insignificant. On the other hand, the Geske model is better than then Merton model (accuracy ratio of 83.58%) with a t value of 2.16, which is significant. The Non Parametric model uses a different sample size as the other models. Hence, formal statistical test is not possible (NA). The sample size of the Non Parametric model is 861,917. The sample size of the other models is 1,047,727. Numbers in parentheses are t values.

Table 8: Accuracy Ratios and z-tests (Common Sample)

1 month forecast	NP	blkcox	geske	merton	longsch	barrier
	0.9112	0.8505	0.8468	0.8359	0.835	0.8346
	(10.23)	(0.85)	(2.47)	(0.2)	(0.09)	
2 month forecast	NP	blkcox	geske	barrier	longsch	merton
	0.9026	0.8632	0.8544	0.8391	0.8332	0.8297
	(10.34)	(2.95)	(4.97)	(1.86)	(1.1)	
3 month forecast	NP	blkcox	geske	barrier	longsch	merton
	0.8959	0.8723	0.8563	0.8373	0.8361	0.8246
	(7.87)	(6.44)	(7.59)	(0.47)	(4.44)	
4 month forecast	NP	blkcox	geske	barrier	longsch	merton
	0.8898	0.8751	0.8584	0.8409	0.8363	0.8174
	(5.66)	(7.84)	(8.16)	(2.1)	(8.39)	
5 month forecast	NP	blkcox	geske	barrier	longsch	merton
	0.883	0.8737	0.8559	0.8384	0.8354	0.8111
	(3.88)	(9.66)	(9.08)	(1.53)	(11.55)	
6 month forecast	NP	blkcox	geske	barrier	longsch	merton
	0.8761	0.8728	0.8539	0.8363	0.8337	0.8049
	(1.43)	(10.79)	(9.97)	(1.44)	(14.89)	
7 month forecast	blkcox	NP	geske	barrier	longsch	merton
	0.8713	0.8704	0.8511	0.8331	0.8322	0.799
	(0.41)	(10.16)	(10.94)	(0.54)	(18.41)	
8 month forecast	blkcox	NP	geske	barrier	longsch	merton
	0.8688	0.8645	0.849	0.8315	0.8301	0.7936
	(1.99)	(8.63)	(11.31)	(0.89)	(20.74)	
9 month forecast	blkcox	NP	geske	barrier	longsch	merton
	0.8647	0.8589	0.8459	0.828	0.8278	0.7876
	(2.72)	(7.58)	(12.17)	(0.13)	(24.02)	

Table 8 Accuracy Ratios and z-tests (Common Sample) – Continued

10 month forecast	blkcox	NP	geske	barrier	longsch	merton
	0.8612	0.8532	0.8424	0.8256	0.8246	0.7821
	(3.83)	(6.39)	(11.95)	(0.7)	(25.7)	
11 month forecast	blkcox	NP	geske	barrier	longsch	merton
	0.8556	0.8479	0.8395	0.8218	0.8214	0.777
	(3.7)	(5.16)	(13.08)	(0.1)	(11.24)	
12 month forecast	blkcox	NP	geske	longsch	barrier	merton
	0.8503	0.8427	0.8358	0.8176	0.8174	0.7715
	(3.7)	(4.37)	(14.06)	(0.15)	(30.85)	

Note: The common dataset for all the models to be comparable to one another is 965,680.

The z statistic is defined as follows:

$z = \frac{A_1 - A_2}{\sigma_{A_1 - A_2}}$ where A_i represents the area under the ROC curve (accuracy ratio) for model i.

Table 9: Area under ROC

Financials

Area under ROC	NP	merton	barrier	blkcox	geske	longsch
1	94.74%	79.53%	85.72%	91.07%	91.11%	76.35%
2	94.07%	79.62%	87.01%	91.56%	90.35%	75.07%
3	93.83%	80.53%	87.21%	91.34%	91.27%	75.71%
4	93.65%	79.81%	89.15%	90.91%	91.55%	76.14%
5	93.45%	79.69%	87.49%	90.40%	91.13%	76.68%
6	93.12%	79.13%	88.45%	89.78%	90.54%	76.54%
7	92.75%	78.60%	88.46%	88.87%	89.92%	76.82%
8	92.37%	78.25%	88.30%	87.93%	89.27%	77.01%
9	92.00%	77.01%	88.19%	86.57%	88.47%	77.14%
10	91.50%	76.54%	87.14%	85.13%	87.52%	77.10%
11	91.05%	76.14%	86.13%	83.53%	86.58%	76.70%
12	90.53%	75.40%	85.15%	81.94%	85.51%	76.05%

Industrials

Area under ROC	NP	merton	barrier	blkcox	geske	longsch
1	90.49%	83.28%	82.65%	83.82%	83.44%	83.61%
2	89.58%	82.59%	83.02%	85.34%	84.38%	83.53%
3	88.83%	81.93%	82.86%	86.52%	84.57%	83.83%
4	88.14%	81.23%	83.10%	87.03%	84.82%	83.85%
5	87.38%	80.57%	83.04%	87.07%	84.64%	83.71%
6	86.64%	79.98%	82.77%	87.19%	84.54%	83.59%
7	86.03%	79.46%	82.49%	87.27%	84.35%	83.44%
8	85.41%	78.95%	82.38%	87.22%	84.24%	83.26%
9	84.79%	78.48%	82.08%	87.08%	84.05%	83.05%
10	84.19%	78.03%	81.99%	87.02%	83.85%	82.80%
11	83.63%	77.60%	81.75%	86.74%	83.70%	82.56%
12	83.09%	77.17%	81.45%	86.51%	83.49%	82.31%

Note: The sample size for the Financials is 154,402. The sample size of the non-Financials is 811,278. Details of the z test are omitted and can be obtained from the authors upon request.

Table 9: Area under ROC (Continued)

Investment Grades

Area under ROC	NP	merton	barrier	blkcox	geske	longsch
1	86.10%	75.30%	75.30%	75.10%	75.30%	75.30%
2	80.87%	67.41%	67.77%	67.59%	67.77%	67.77%
3	76.86%	64.68%	65.15%	64.80%	65.15%	65.15%
4	72.76%	61.04%	61.71%	61.50%	61.71%	61.63%
5	71.04%	60.10%	60.82%	60.44%	60.85%	60.85%
6	70.51%	59.09%	60.43%	60.43%	60.43%	60.33%
7	69.57%	58.56%	60.43%	60.43%	60.45%	60.45%
8	68.12%	58.25%	60.44%	60.59%	60.46%	60.39%
9	67.25%	58.33%	60.90%	61.30%	60.91%	60.75%
10	66.70%	58.83%	61.47%	62.54%	61.50%	61.37%
11	66.70%	58.83%	61.47%	62.54%	61.50%	61.37%
12	66.87%	60.00%	62.02%	64.12%	63.09%	62.85%

Speculative Grades

Area under ROC	NP	merton	barrier	blkcox	geske	longsch
1	90.70%	82.48%	83.70%	84.01%	84.27%	80.93%
2	90.12%	81.97%	84.04%	86.46%	85.64%	81.14%
3	89.65%	81.68%	83.45%	86.34%	86.32%	82.04%
4	89.42%	80.79%	84.05%	86.36%	86.71%	82.36%
5	88.82%	80.33%	83.54%	86.57%	86.30%	82.40%
6	88.23%	79.80%	83.31%	86.41%	86.05%	82.28%
7	87.82%	79.33%	83.06%	86.18%	85.49%	82.13%
8	87.38%	79.02%	83.06%	85.83%	85.23%	82.00%
9	86.91%	78.53%	82.69%	85.04%	84.77%	81.84%
10	86.41%	78.23%	83.00%	84.64%	84.35%	81.66%
11	86.01%	78.03%	82.63%	84.20%	84.00%	81.41%
12	85.52%	77.74%	82.32%	83.77%	83.61%	81.11%

Note: The sample size for the investment grades is 109,146. The sample size for the speculative grades is 87,384. Details of the z test are omitted and can be obtained from the authors upon request.

Table 10: Regression Test

1 month forecast

Model/all	NP	Merton	Barrier	Blkcox	Geske	Longsch
Intercept	8.18	8.39	8.43	8.46	12.36	8.39
	76.17	6.92	6.99	7.02	0.97	6.96
Size	0.03	0.03	0.03	0.03	0.03	0.03
	53.45	4.09	4.11	4.11	0.44	4.10
Bankrupt	-5.98	-5.45	-4.87	-5.76	0.67	-4.89
	-31.53	-2.55	-2.29	-2.71	0.03	-2.30
Leverage	0.53	1.61	1.43	1.28	-9.31	1.53
	23.42	6.29	5.65	5.05	-3.48	6.01
S1	-1.26	-1.06	-1.06	-1.05	-0.52	-1.06
	-11.46	-0.86	-0.86	-0.85	-0.04	-0.86
S2	-0.05	0.26	0.26	0.27	-0.15	0.26
	-0.46	0.21	0.21	0.22	-0.01	0.21
S3	-1.22	-0.88	-0.88	-0.87	-1.10	-0.88
	-11.32	-0.72	-0.73	-0.72	-0.09	-0.73
S4	2.83	3.46	3.46	3.50	5.08	3.45
	26.05	2.82	2.84	2.87	0.40	2.83
S5	-0.95	-0.13	-0.13	-0.11	0.73	-0.14
	-8.72	-0.11	-0.10	-0.09	0.06	-0.11
S6	2.11	2.63	2.57	2.58	-4.64	2.66
	19.50	2.15	2.12	2.12	-0.36	2.19
S7	-2.44	-1.16	-1.16	-1.16	-1.61	-1.15
	-22.54	-0.95	-0.95	-0.95	-0.13	-0.95
S8	-2.28	-2.29	-2.25	-2.28	-8.55	-2.23
	-20.61	-1.83	-1.81	-1.84	-0.65	-1.80
S9	-1.63	6.89	6.86	6.89	7.46	6.86
	-12.59	4.72	4.72	4.74	0.49	4.72
Adj R-Sq	0.0985	0.0010	0.0010	0.0010	0.0000	0.0010

Note: The dependent variable is DD of each model. Size is market cap, Bankrupt is a dummy (1 for default), Leverage is book value debt over market cap plus book value debt, S1 ~ S9 are SIC dummy variables.

Table 10 Continued

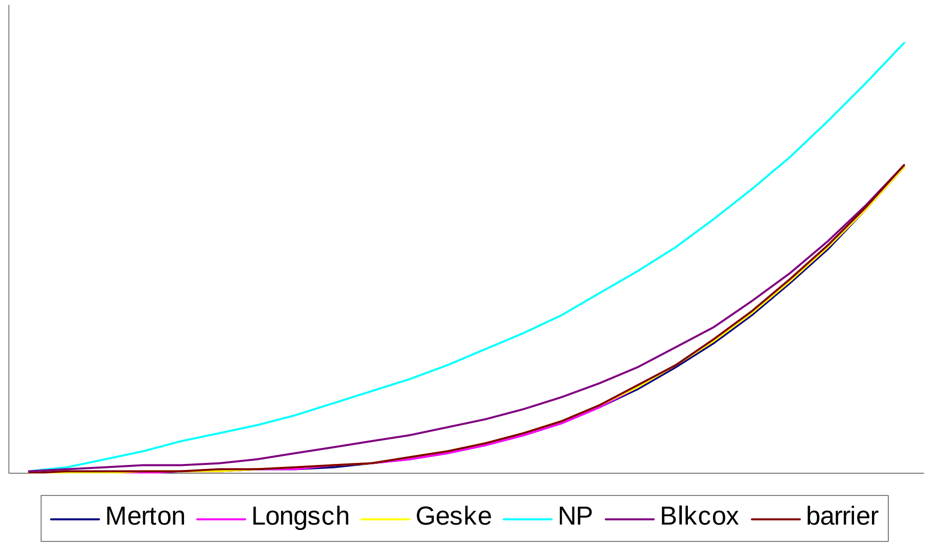
12 month forecast

Model/all	NP	Merton	Barrier	Blkcox	Geske	Longsch
Intercept	2.35	2.57	2.64	2.73	2.66	2.65
	76.22	7.35	7.64	8.35	7.85	8.08
Size	0.01	0.01	0.01	0.01	0.01	0.01
	53.32	4.46	4.25	4.80	4.45	4.51
Bankrupt	-1.38	-1.32	-1.25	-1.45	-1.45	-1.40
	-87.22	-7.39	-7.06	-8.66	-8.39	-8.34
Leverage	0.23	0.10	-0.30	-0.93	-0.42	-0.36
	35.24	1.37	-4.08	-13.44	-5.84	-5.23
S1	-0.37	-0.29	-0.29	-0.30	-0.29	-0.30
	-11.70	-0.81	-0.83	-0.88	-0.85	-0.88
S2	-0.02	0.05	0.04	0.04	0.04	0.04
	-0.64	0.15	0.11	0.13	0.12	0.12
S3	-0.36	-0.27	-0.31	-0.29	-0.28	-0.28
	-11.56	-0.76	-0.90	-0.89	-0.83	-0.84
S4	0.81	1.05	1.02	1.11	0.99	1.03
	25.80	2.97	2.90	3.35	2.89	3.11
S5	-0.27	0.00	-0.02	-0.05	-0.02	-0.02
	-8.81	-0.01	-0.07	-0.15	-0.07	-0.05
S6	0.57	0.02	0.20	0.29	0.22	0.12
	18.31	0.06	0.57	0.88	0.64	0.35
S7	-0.70	-0.35	-0.38	-0.39	-0.38	-0.37
	-22.57	-0.99	-1.09	-1.20	-1.12	-1.12
S8	-0.66	-0.66	-0.65	-0.68	-0.66	-0.64
	-20.70	-1.84	-1.84	-2.01	-1.90	-1.90
S9	-0.47	2.00	1.90	1.88	1.91	1.90
	-12.73	4.74	4.57	4.77	4.69	4.82
Adj R-Sq	0.1049	0.0007	0.0006	0.0009	0.0007	0.0007

Note: The dependent variable is DD of each model. Size is market cap, Bankrupt is a dummy (1 for default), Leverage is book value debt over market cap plus book value debt, S1 ~ S9 are SIC dummy variables.

Figure 1: Left Tail Probability 5% Plots of the DD Distributions

One month forecast



Twelve month forecast

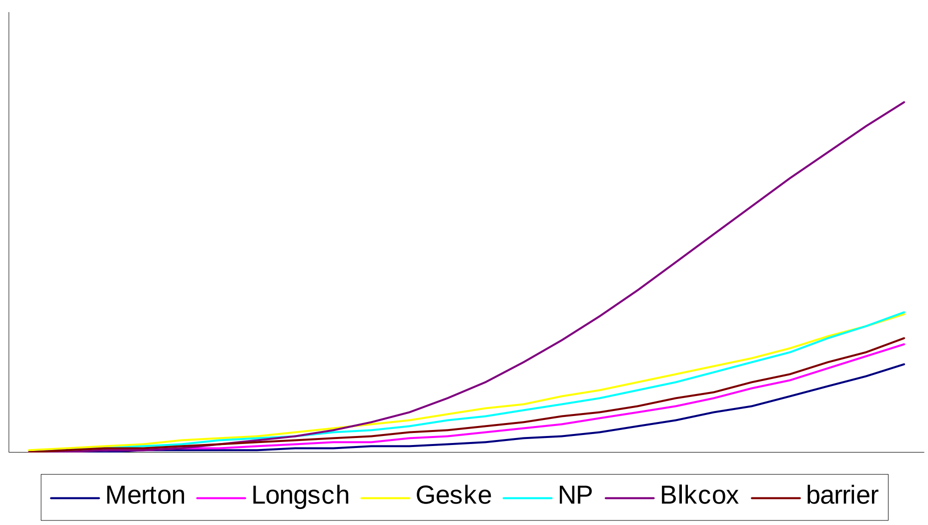
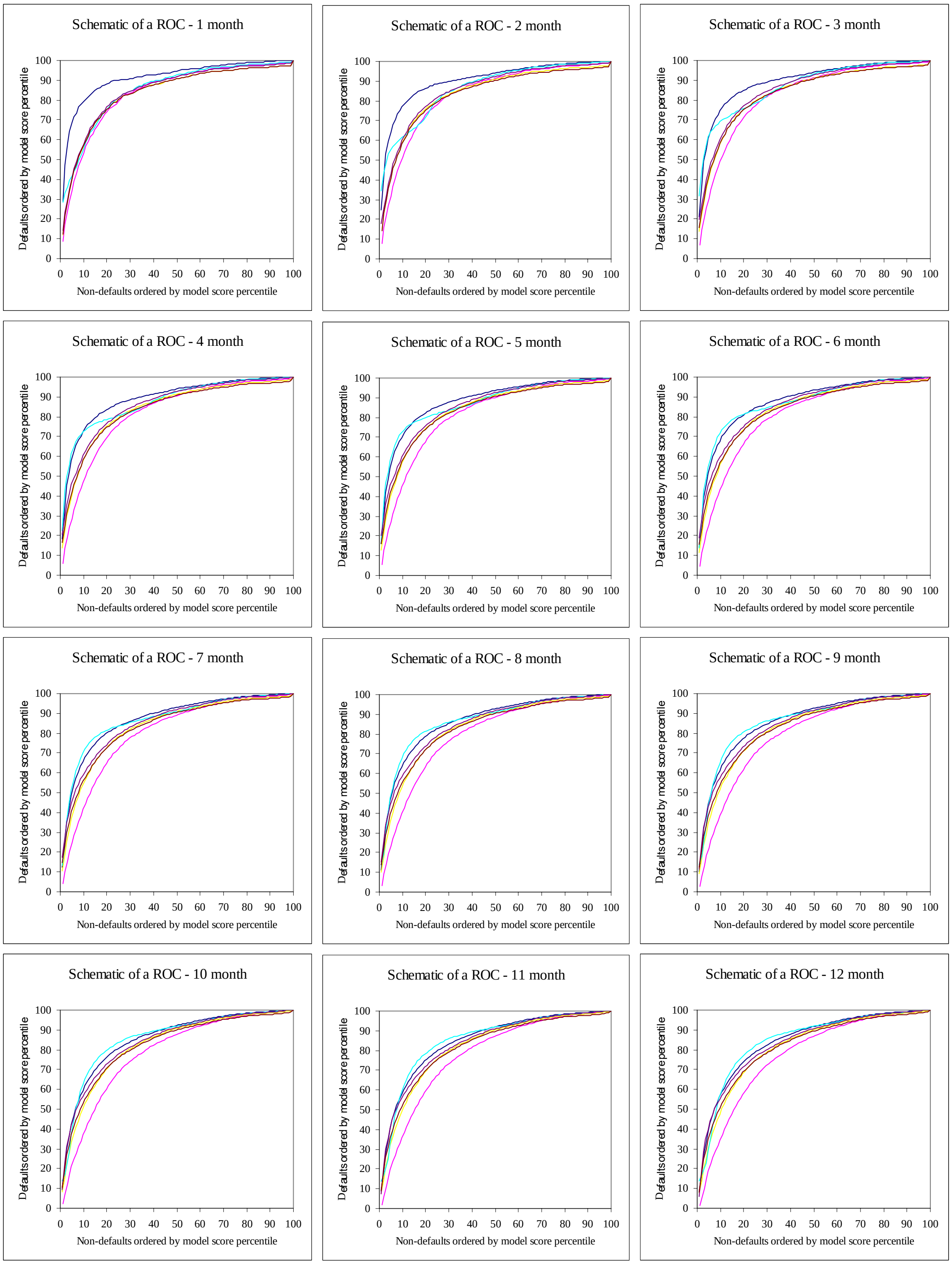


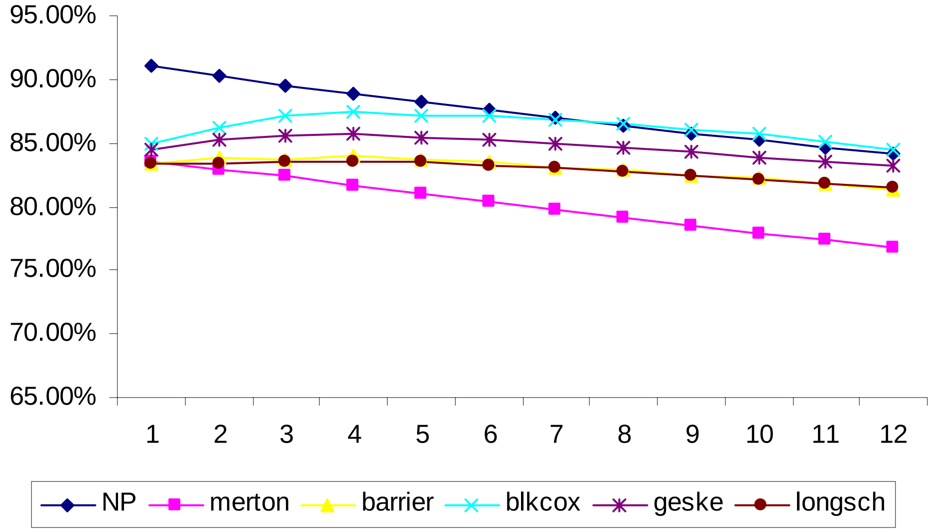
Figure 2: Receiver Operating Characteristic Curves: All Models



Note: The ARs for various models are represented in Table 7.

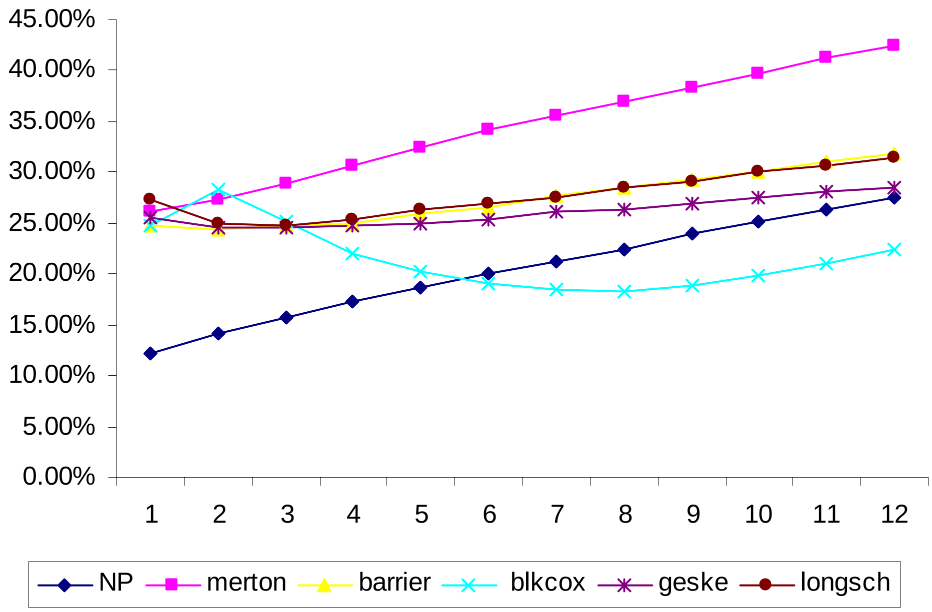
NP Merton Barrier BlkCox Geske LongSch

Figure 3: Accuracy Ratios of Various Models (full sample)



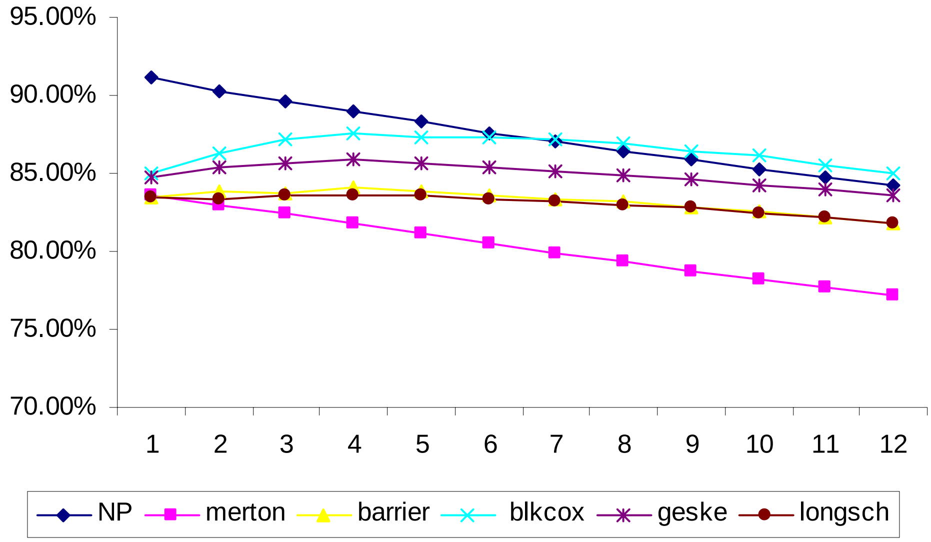
Note: The dataset used in the models is the full sample. The dataset for the Non Parametric model is different from the other models.

Figure 4: Type-I Errors (α) of Various Models (full sample)



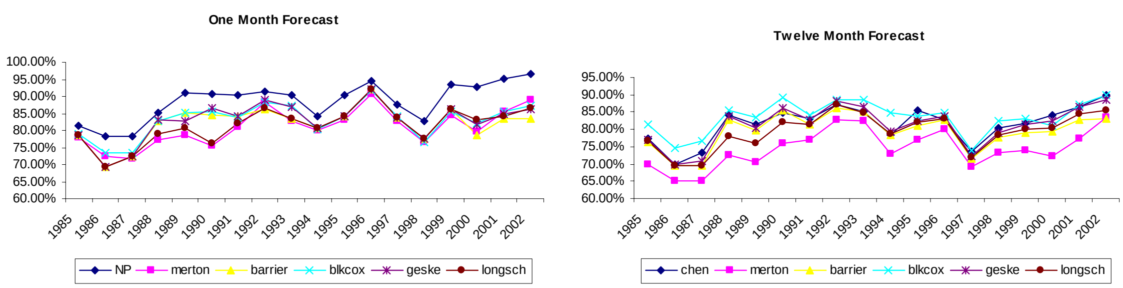
Note: The Tyle-II error (β) is controlled at 20%.

Figure 5: Accuracy Ratios of Various Models (common sample)



Note: The dataset used in the models is the smaller sample over which all models are tested upon.

Figure 6: Accuracy Ratios of Various Models by Year (common sample)



8. Appendix

In all of the following models, we solve the simultaneous equations for the asset value and asset volatility:

$$\begin{cases} E = E^{*} \\ \sigma \dfrac{A}{E}\dfrac{\partial E}{\partial A} = \sigma_E^{*} \end{cases}$$

A. Black-Scholes-Merton model

$$E = VN(d_+) - e^{-rT}KN(d_-)$$

where

$$d_{\pm} = \frac{\ln V / K + (r \pm \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}$$

B. Black-Cox model

$$E(V,t) = V - \left\{Pe^{-r(T-t)}N(z_1) - Pe^{-r(T-t)}y^{2\vartheta-2}N(z_2) + Ve^{-a(T-t)}N(z_3) + Ve^{-a(T-t)}y^{2\vartheta}N(z_4) + Vy^{\vartheta+\zeta}N(z_5) + Vy^{\vartheta-\zeta}N(z_6) - Ve^{-a(T-t)}y^{\vartheta+\eta}N(z_7) - Ve^{-a(T-t)}y^{\vartheta-\eta}N(z_8)\right\}$$

where, for the sake of completeness,

$$y = Ce^{-\gamma(T-t)} / V$$

$$\vartheta = (r - a - \gamma + \frac{1}{2}\sigma^2) / \sigma^2$$

$$\delta = (r - a - \gamma - \frac{1}{2}\sigma^2) / \sigma^2 + 2\sigma^2(r - \gamma)$$

$$\zeta = \sqrt{\delta} / \sigma^2$$

$$\eta = \sqrt{\delta - 2\sigma^2 a} / \sigma^2$$

$$z_1 = \left[\ln V - \ln P + (r - a - \frac{1}{2}\sigma^2)(T - t)\right] / \sqrt{\sigma^2(T - t)}$$

$$z_2 = \left[\ln V - \ln P + 2\ln y + (r - a - \frac{1}{2}\sigma^2)(T - t)\right] / \sqrt{\sigma^2(T - t)}$$

$$z_3 = \left[\ln P - \ln V - (r - a + \frac{1}{2}\sigma^2)(T - t)\right] / \sqrt{\sigma^2(T - t)}$$

$$z_4 = \left[\ln V - \ln P + 2\ln y + (r - a + \frac{1}{2}\sigma^2)(T - t)\right] / \sqrt{\sigma^2(T - t)}$$

$$z_5 = \left[\ln y + \zeta\sigma^2(T-t)\right] / \sqrt{\sigma^2(T-t)}$$
$$z_6 = \left[\ln y - \zeta\sigma^2(T-t)\right] / \sqrt{\sigma^2(T-t)}$$
$$z_7 = \left[\ln y + \eta\sigma^2(T-t)\right] / \sqrt{\sigma^2(T-t)}$$
$$z_8 = \left[\ln y - \eta\sigma^2(T-t)\right] / \sqrt{\sigma^2(T-t)}$$

The mistake is the underscored term. It should be $Ve^{-a(T-t)}y^{\vartheta+\eta}N(z_7)$ but the original article has a misprint of $\vartheta - \eta$ for the power of y.

C. Geske Model

$$E_0 = V_0 M(h_{1+}, h_{2+}; \sqrt{T_1 / T_2}) - e^{-r(T_2-t)} F_2 M(h_{1-}, h_{2-}; \sqrt{T_1 / T_2}) - e^{-r(T_1-t)} F_1 N(h_{1-})$$

where

$$h_{j\pm} = \frac{\ln V_0 - \ln \bar{V}_j + (r \pm \sigma^2 / 2)(T_j - t)}{\sigma\sqrt{T_j - t}}$$

and $\bar{V}$ is the solution to $E = K$.

D. Flat Barrier Model

$$D(0,T) = A(0)\left[1 - \left\{N(d^+) - \left(\frac{H}{A(0)}\right)^{2q} N(h^+)\right\}\right] + e^{-rT}K\left[N(d^-) - \left(\frac{H}{A(0)}\right)^{2q-2} N(h^-)\right]$$

for $H > K$ where

$$d^\pm = \frac{\ln A(0) - \ln H}{\sigma\sqrt{T}} + q^\pm\sigma\sqrt{T}$$
$$h^\pm = \frac{\ln H - \ln A(0)}{\sigma\sqrt{T}} + q^\pm\sigma\sqrt{T}$$
$$q^\pm = \frac{r}{\sigma^2} \pm \tfrac{1}{2}\sigma$$

and

$$D(0,T) = A(0)\left[N(-x^+) + \left(\frac{H}{A(0)}\right)^{2q} N(y^+)\right] + e^{-rT}K\left[N(x^-) - \left(\frac{H}{A(0)}\right)^{2q-2} N(y^-)\right]$$

for $H < K$

where

$$x^\pm = \frac{\ln A(0) - \ln K + rT}{\sigma\sqrt{T}} \pm \tfrac{1}{2}\sigma\sqrt{T}$$
$$y^\pm = \frac{2\ln H - \ln A(0) - \ln K}{\sigma\sqrt{T}} + q^\pm\sigma\sqrt{T}$$

Fixed Recovery – this is to compare to LS

$D = e^{-rT}K(\Pr[A(t) > H] + R\Pr[A(t) < H])$ which is

$$e^{-rT}K\left[N(x^-) - \left(\frac{H}{A(0)}\right)^{2q-2} N(y^-) + R\left\{1 - N(x^-) + \left(\frac{H}{A(0)}\right)^{2q-2} N(y^-)\right\}\right] \text{ for } H < K \text{ and}$$

$$e^{-rT}K\left[N(d^-) - \left(\frac{H}{A(0)}\right)^{2q} N(h^-) + R\left\{1 - N(d^-) + \left(\frac{H}{A(0)}\right)^{2q} N(h^-)\right\}\right] \text{ for } H > K$$

E. Longstaff-Schwarts Model

$$D = P(r,T)[1 - wU(X,r,T)]$$

where

$$U(X,r,T) = \sum_{i=1}^{n} u_i$$
$$u_1 = N(a_1)$$
$$u_i = N(a_i) - \sum_{j=1}^{i-1} u_j N(b_{ij})$$
$$a_i = \frac{-\ln X - M(\frac{iT}{n},T)}{\sqrt{S(\frac{iT}{n})}}$$
$$b_{ij} = \frac{M(\frac{jT}{n},T) - M(\frac{iT}{n},T)}{\sqrt{S(\frac{iT}{n}) - S(\frac{jT}{n})}}$$
$$M = \left(\frac{\alpha - \rho\sigma\eta}{\beta} - \frac{\eta^2}{\beta^2} - \frac{\sigma^2}{2}\right)t + \left(\frac{\rho\sigma\eta}{\beta^2} + \frac{\eta^2}{2\beta^3}\right)e^{-\beta T}\left(e^{\beta t} - 1\right)$$
$$+\left(\frac{r}{\beta} - \frac{\alpha}{\beta^2} + \frac{\eta^2}{\beta^3}\right)\left(1 - e^{-\beta t}\right) - \left(\frac{\eta^2}{2\beta^2}\right)e^{-\beta T}\left(1 - e^{-\beta t}\right)$$
$$S = \left(\frac{\rho\sigma\eta}{\beta} + \frac{\eta^2}{\beta^2} + \sigma^2\right)t - \left(\frac{\rho\sigma\eta}{\beta^2} + \frac{2\eta^2}{\beta^3}\right)\left(1 - e^{-\beta t}\right)$$
$$+\left(\frac{\eta^2}{2\beta^3}\right)\left(1 - e^{-2\beta t}\right)$$

F. Estimation of the Vasicek Model

The Vasicek model is a one-factor mean reverting Gaussian model with the following short rate dynamics:

$$dr = \alpha(\mu - r)dt + \upsilon dW_2$$

under the risk-neutral measure. This is an AR(1) process in discrete time:

$$r_t = \left(1 - e^{-\alpha h}\right)\mu + e^{-\alpha h} r_{t-h} + e_t$$

where h is time interval and e_t has mean 0 and variance $\frac{1-e^{-2\alpha h}}{2\alpha}\upsilon^2$.

The closed-form solution from equation is derived by Vasicek as follows:

$$P_{t,T} = e^{-r_t F(T-t) - G(T-t)}$$

where

$$F(\tau) = \frac{1 - e^{-\alpha\tau}}{\alpha}$$

$$G(\tau) = \left(\mu - \tfrac{\upsilon^2}{2\alpha^2}\right)(\tau - F(\tau)) + \frac{\upsilon^2 F(\tau)^2}{4\alpha}$$

and $t + \tau = T$ represents the maturity of the bond. Note that the 3-month CMT rates have a constant rolling maturity of 3 months. Hence $\tau = 1/4$ and remains fixed. Inverting the pricing formula gives the instantaneous rate:

$$r_t = -\frac{\ln P_{t,T} + G(T-t)}{F(T-t)}$$

Approximating the CMT rates as continuously compounded money market rates and substituting this into the regression equation, we obtain the following revised regression equation where we can use the CMT rates:

$$\frac{-\ln P_{t,T}}{\tau} = \left(\frac{1 - e^{-\alpha h}}{\tau}\right)(\mu F_\tau + G_\tau) + e^{-\alpha h}\frac{-\ln P_{t+h,T+h}}{\tau} + \frac{F_\tau}{\tau} e_t$$

where

$$\text{Slope} = e^{-\alpha h}$$

$$\text{Intercept} = \left(\frac{1 - e^{-\alpha h}}{\tau}\right)(\mu F_\tau + G_\tau)$$

$$\text{SSE} = \tfrac{F_\tau^2}{\tau^2}\tfrac{\upsilon^2(1 - e^{-2\alpha h})}{2\alpha}$$

where $h = \frac{1}{252}$ for daily data. To obtain parameter estimates, we solve the slope for α, then SSE for υ, and finally Intercept for μ.