DBJ Discussion Paper Series, No.1214

Corporate Bond Premia

Yoshio Nozawa
(Development Bank of Japan)

March 2013

Discussion Papers are a series of preliminary materials in their draft form. No quotations, reproductions or circulations should be made without the written consent of the authors in order to protect the tentative characters of these papers. Any opinions, findings, conclusions or recommendations expressed in these papers are those of the authors and do not reflect the views of the Institute.

Corporate Bond Premia*

Yoshio Nozawa†

March 13, 2013

Abstract

I identify level and slope factors in corporate bond returns. I show that these two factors can explain 93% of the variation in average excess returns on corporate bonds. To show this result, I describe expected excess returns and risks as functions of characteristics of corporate bonds such as bond spreads and use a parametric characteristic-based asset pricing test. This approach allows one to test the model's ability to explain the variation in average excess returns associated with multiple characteristics. The two factor model does well for all characteristics except equity momentum.

*I thank my advisors John Cochrane, Lars Hansen, Ralph Koijen and Pietro Veronesi for their continuous guidance. I am grateful for comments and suggestions from Eugene Fama, Zhiguo He, Bryan Kelly, Tobias Moskowitz, Diogo Palhares, Shri Santosh, Brian Weller and participants in the Finance Workshop and Economics Dynamics working group at the University of Chicago and Development Bank of Japan. Research support from the Fama-Miller Center for Research in Finance is gratefully acknowledged.

†Development Bank of Japan, University of Chicago Booth School of Business and Department of Economics. Email: ynozawa@chicagobooth.edu

1 Introduction

I study empirically the behavior of the corporate bond premia and corporate bond spreads using the historical data of corporate bond prices since 1973 in the US. The corporate bond premium is defined as the expected return on corporate bonds in excess of synthetic treasury bonds with an identical cash flow schedule. The corporate bond spread is defined as the relative price of a corporate bond to its corresponding treasury bond, scaled by the time to maturity. I construct a factor pricing model based on the bond spread to understand the variation of the corporate bond premia observed in the data.

My asset pricing model consists of the first two principal components of the excess returns on corporate bond portfolios sorted on bond spreads. The first principal component is a level factor, on which all corporate bonds have the same loading. The second principal component is a slope factor, which affects the bonds with high bond spread positively and the bonds with low bond spread negatively. I show that this simple two factor model can account for the majority of the variation in bond premia when bonds are sorted into portfolios based on 14 characteristics such as bond spread, coupon yields and momentum.

The two factor model has a problem in pricing the corporate bond portfolios sorted on the issuers' past equity returns. That is, the equity momentum in the bond market identified by Gebhardt, Hvidkjaer and Swaminathan (2005b) causes a problem for the two factor model.

I also show that these two factors explain more than 70% of the variation in realized excess returns on portfolios sorted on other characteristics. This finding suggests that there is a strong factor structure in corporate bond excess returns and shows that the empirical result of this article can be interpreted as an application of the Arbitrage Pricing Theory of Ross (1976).

Next, I show a novel approach to conduct an asset pricing test which complements the classic test result using portfolio sorts.

As pointed out by Cochrane (2011), the asset pricing test based on portfolio sorts amounts to matching the expected excess return function (as a function of securities' characteristics) to the risk function produced by the model. The classic test based on portfolio sorts implicitly assumes that both expected excess returns and risk are functions of characteristics. I parametrically characterize the expected excess return and risk functions of characteristics, using individual security data. This approach yields a number of benefits for the test.

With the parametric characteristic-based asset pricing test, I can test the model using multiple characteristics at once. With the classic test based on portfolio sorts, using multiple characteristics is challenging due to the limited data amount. The parametric characteristic-based test is particularly useful in analyzing my sample, where the sample size varies significantly over time. With a few number of observations in the early 1970s, it is impossible to conduct multidimensional sorts using my corporate bond data. With the parametric characteristic-based test, using four or five characteristics is fairly easy and can be done without losing precision in the estimated bond premia.

To implement the test, I use a three-step procedure based on pooled OLS regressions using individual bond-level observations. In the first and second steps, I estimate the expected excess return function and the risk function parametrically using a large panel of individual security returns. That is, I forecast returns and comovements with the characteristics of the bonds, such as bond spread and time to maturity. I deliberately remove the 'name' of the security from the regressor of the forecasting regression. By conditioning on characteristics and not on the name of the security, I can estimate the expected excess returns and expected comovements precisely. This result shows that the corporate bond premia and the risks associated with corporate bonds are more stable functions of the bond characteristics than the security name.

In the third step, I project the estimated expected excess returns onto the estimated risk by minimizing the sum of squared pricing errors. These pricing errors are small if the expected excess return function is close to a linear combination of the risk functions. That is, if the estimated bond premia is matched well by a linear combination of the estimated risks produced by the model, the model is successful in pricing these test assets.

Using the parametric characteristic-based asset pricing test, I show that the two factor model explains 93% of the variation in estimated bond premia even when the four characteristics that best predict excess returns are used as instruments.

From a methodological perspective, the parametric characteristic-based test is attractive for a number of reasons. In addition to the capability of handling multiple characteristics in the test, using individual securities gives rise to a large variation in estimated bond premia and risks. A large variation in estimated risks leads to more efficient estimates of the model parameters than classic portfolio sorts. Furthermore, the parametric characteristic-based test offers the flexibility of testing both time-varying bond premia as well as cross-sectional variation in bond premia.

Corporate bond premia can arise due to an exposure to systematic risk, transaction costs or mispricing in bonds with some characteristics. The transaction cost to trade corporate bonds can be large. Since corporate bonds are traded in the Over-the-counter (OTC) market, an investor may have to search for a trading counterparty. As she anticipates costly search when selling the bond in the future, the price today may reflect a discount[1]. My empirical result does not identify which one of these three sources is the main reason for the bond premia. The fact that 93% of the variation in bond premia can be explained by the exposure to the systematic risk only shows that risk-based explanation is a plausible explanation of corporate bond premia.

The rest of the article proceeds as follows: In section 2, I discuss the related literature. In

[1] Acharya and Pedersen (2005) show that the existence of liquidity can lead to liquidity risk, which may then show up as a systematic risk. My factor model may be a proxy for liquidity risk and I do not separate liquidity risk from credit risk. Rather, with the two factor model I separate the risk premia (the sum of credit and liquidity risk premia) and the liquidity (transaction cost) itself.

section 3, I explain the data to be used for the empirical analysis. I conduct the asset pricing test of the two factor model based on classic portfolio sorts in Section 4. I introduce the parametric characteristic-based approach in Section 5 and show the test result. The last section provides concluding comments.

2 Literature Review

There is a strand of literature which identifies and explains the risk premia priced in the corporate bond market. Gebhardt, Hvidkajaer and Swaminathan (2005), the paper closest to mine, use a two factor model of corporate bonds consisting of default (the long-term aggregate corporate bond returns minus the long-term treasury bond returns) and term (the long-term treasury bond returns minus T-bill rates) factors. They find that the two-factor model prices the cross-sectional variation in expected excess returns of corporate bonds associated with duration and credit ratings. However, the model fails to price the variation associated with yield to maturity.

The key difference in this article is that I adopt two factors consisting of the first two principal components of the bond excess returns. I find that the second principal component, a slope factor, is particularly important in explaining the variation in expected excess returns on corporate bonds. With my two factor model, I can price the variation in corporate bond premia associated not only with the three characteristics tested in Gebhardt, Hvidkajaer and Swaminathan (2005), but also many others.

Another strand of literature identifies the risk premia priced in corporate bonds based on models. By constructing either a reduced form model or a structural model which is consistent with the observed credit spread and default probability, one can study the source and the magnitude of risk premia. Papers based on a reduced form model includes Driessen (2005), while papers that use a structural model include Leland (1994), Chen, Collin-Dufresne and Goldstein (2009) and Bharmra, Kuehn and Strebulaev (2010), among others. The focus of these papers are limited to explaining the variation in risk premia associated with time to maturity and credit ratings. In this article, I explore a wider range of characteristics.

This article proposes a novel approach to conduct asset pricing tests based on characteristics. The idea of explicitly expressing moments of the returns on securities using characteristics is pioneered by Rosenberg (1974), who models the beta of stocks that depend on the stocks' characteristics. The use of characteristics of securities in an asset pricing test is proposed by Cochrane (2011) and partially implemented by Gao (2009). Cochrane (2011) points out that the traditional asset pricing test based on portfolio sorts can be thought of as a characteristics-based test by

"An implicit assumption underlies everything we do: Expected returns, variances, and covariances are stable functions of characteristics such as size and book-to-market ratio, and not security names. This assumption is why we use portfolios in the first place."

I formalize Cochrane (2011)'s idea and show that the parametric characteristic-based test can be conducted using an instrumental variable regression. Gao (2009) estimates non-parametrically the expected excess returns and covariance as a function of stocks' characteristics. However, Gao (2009) applies a classic asset pricing test of Fama and MacBeth (1973), and tests if his measure of covariance is priced in the cross-section of equities. In this article, I present a coherent framework of both the measurement of moments, and the asset pricing test based on characteristics.

More recently, Connor, Hagmann and Linton (2012) propose an efficient non-parametric asset pricing test for their characteristic-based factor models. My approach is parametric and allows risks to depend on multiple characteristics at once. Therefore my work is complimentary to Conner, Hagmann and Linton (2012).

The parametric characteristics-based test presented in this paper allows both the expected excess returns and risk measures to vary over time. The existing research approaches the time-varying expected excess returns differently. Ferson and Harvey (1991) attempt to explain the time-variation of expected excess returns for 10 size-sorted portfolios and industry portfolios using standard 5-year rolling betas. The characteristics-based test I present in this article allows one to measure the time-varying risk more precisely without limiting the focus to portfolios sorted in a particular way. Other articles regarding asset pricing tests with time-varying expected excess returns include Nagel and Singleton (2011), and Adrian, Crump and Moench (2011).

The parametric characteristic-based test uses individual securities for the test, not the portfolios. Ang, Liu and Schwartz (2010) show that using individual stocks instead of portfolios helps reduce the standard errors for the estimated factor weights of the stochastic discount factor. Gagliardini, Ossola and Scaillet (2011) use individual panel data to test the three factor model of Fama and French (1993). Avramov and Chordia (2006) also propose an asset pricing test using individual securities, extending Fama and MacBeth (1973). All of these studies use the noisy estimates of betas based on time-series regressions for each security. In contrast, I condition the measure of risk on characteristics and reduce noises in the estimates significantly.

3 Data

I obtain the monthly price observation of senior unsecured corporate bonds from the following four data sources. First, from 1973 to 1997, I use the Lehman Brothers Fixed Income Database which provides month-end bid prices. Since Lehman Brothers used these prices to construct the Lehman Brothers bond index while simultaneously trading it, the traders at Lehman Brothers had an incentive to provide correct quotes. Thus although the prices in the Lehman Brothers Fixed Income Database are quote-based, they are considered to be reliable.

In the Lehman Brothers Fixed Income Database, some observations are dealers' quotes while others are matrix prices. Matrix prices are set using algorithms based on the quoted prices of other bonds with similar characteristics. Since matrix prices are less reliable than actual dealer quotes (Warga and Welch (1993)), I remove them from the sample and treat the data points as missing from the sample. The exception is December 1984, where all the prices are recorded as matrix priced, and thus I retain these matrix prices only for this month.

Second, from 1994 to 2011, I use the Mergent FISD/NAIC Database. This database consists of actual transaction prices reported by insurance companies.

Third, from 2002 to 2011, I use TRACE data which provides actual transaction prices. TRACE covers more than 99 percent of the OTC activities in the US corporate bond markets after 2005.

The data from Mergent FISD/NAIC and TRACE are transaction-based data and therefore the observations are not exactly at the end of months. Thus, I use only the observations that are in the last five days of each month. If there are multiple observations in the last five days, I use the latest one and treat it as a month-end observation.

Lastly, I use the DataStream database which provides month-end price quotes from 1990 to 2011.

Since there are some overlaps among the four databases, I prioritize in the following order: the Lehman Brothers Fixed Income Database, TRACE, Mergent FISD/NAIC and DataStream. As Jostova, Nikolova, Philipov and Stahel (2012) find, the number of the overlaps are not large relative to the total size of the dataset, with the largest overlaps between TRACE and Mergent FISD being 3.3% of the non-overlapping observations. To check the data consistency, I compare the difference across databases in Appendix A using the overlapping observations.

The Lehman Brother's Fixed Income Database and Mergent FISD[2] provide the other characteristics specific to the issue of bonds, such as the maturity dates, credit ratings, coupon rates and optionalities of the bonds. I remove bonds with floating rates and with any option features other than callable bonds. Until the late 1980s, there are very few bonds that are non-callable. Thus, removing callable bonds would reduce the length of the sample period significantly and it is for this reason that I include callable bonds in my sample. As the callable bond price reflects the discount due to the call option value, the yield on these bonds are not exactly comparable to the yield on non-callable bonds. Crabbe (1991) estimates that call options contribute nine basis points to the bond spread on average for investment grade bonds. Moreover, Crabbe and Helwege (1994) show that speculative grade bonds typically have lower call option values than investment grade bonds. Therefore, the effect of call options does not seem large enough to affect my results significantly[3].

I apply several filters to remove the observations that are likely to be subject to erroneous recording. First, I remove the pricing observations that are higher than matching treasury bond

[2] Mergent FISD provides relatively limited price information but provides bond characteristic information for most of the bonds since 1994.

[3] As a recent example, Gilchrist and Zakrajsek (2010) also use the Lehman Brothers Fixed Income Database including callable bonds in their analysis.

prices (170,784 observations). Some corporate bond price observations seem to be unreasonably high with a number even exceeding 1,000,000 per 100. By comparing the bond's price with the price of the synthetic treasury with the same repayment schedule (explained below), I determine whether the corporate bond price is unrealistically high or not. On the other hand, several bonds have extremely low price observations, including zeros. Zero prices are unrealistic as the bond is a security with limited liability. Furthermore, unless the firm is in default, a price very close to zero may not be reasonable, either. Thus, I drop the price observations below one cent per dollar, except for bonds that are in default (4,794 observations)[4]. Some bonds have price observations after the maturity date or before the issue date. I drop these observations as well (3,202 observations).

In addition, I remove the return (defined below) observations that show a large bounce back. Specifically, I compute the product of the adjacent return observations and remove both observations if the product is less than -0.04. That is, if the same bond jumps up more than 20% in one month and comes down more than 20% in the following month, I assume that the price observation in the middle is recorded with errors. This filter removes 2,394 observations. Finally, I see some bonds whose prices do not change for extended period (observed frequently in DataStream). As these constant price observations are likely to be subject to lack of liquidity or data entry errors[5], I remove returns that are exactly zero for more than three consecutive months. This filter removes 162,827 observations.

For the information regarding defaults, I use Moody's Default Risk Service which provides the historical record of bond defaults from 1970. I compute the returns only up to the month in which the default occurs. The same source also provides the secondary market value of the defaulted bond one month after the incidence. If the price observation in the month when a bond defaults is missing, I add Moody's secondary market price to my dataset. I also add a price observation of 100 if the price at maturity is missing and the bond does not default nor is called.

After the filtering, I have an unbalanced panel of 1,454,363 bond-month price observations with 39,120 bonds over 468 months.

With the filtered price, I compute the return on corporate bond i by

$$R_{i,t+1} = \frac{(P_{i,t+1} + AI_{i,t+1} + Coupon_{i,t+1}) - (P_{i,t} + AI_{i,t})}{P_{i,t} + AI_{i,t}}$$

where $P_{i,t}$ is the price for corporate bond i at time t, $AI_{i,t}$ is the accrued interest for bond i at time t and $Coupon_{i,t}$ is the coupon payment for bond i at time t.

To compute some of the characteristics of bonds, I use accounting information from Compustat. The short-term debt ratio is computed by dividing the amount of short-term debt by the amount

[4] One might be concerned by the fact that the bond price falls before default and removing low prices then might remove low returns before default, which is informative. The concern does not invalidate this filter as only 0.09% of the price observations upon default is less than one cent.

[5] Chen, Lesmond and Wei (2007) find that zero return observations are highly related to illiquidity.

of total debt outstanding. The tangibility ratio is computed by dividing the value of property, plant and equipment by the value of total assets. I use CRSP for monthly stock price observations and return volatilities.

To compute the corporate bond premia and the corporate bond spread, I need to construct prices of the synthetic treasury bonds that match corporate bonds. To this end, I use the Federal Reserve's constant maturity yields data. First, I interpolate the treasury yield curve using cubic splines and construct zero coupon curves for treasuries by bootstrapping. At each month and for each corporate bond in the dataset, I construct the future cash flow schedule for the coupon and principal payments. Then I multiply each cash flow with the zero coupon treasury bond price with the corresponding time to maturity[6]. I add all the cash flows to obtain the synthetic treasury bond price which matches the corporate bond. I do this process for all corporate bonds at each month to obtain the panel data of matching treasury bond prices. I compute the returns on the synthetic treasury $R^f_{i,t+1}$ using the same definition as the corporate bonds.

To compare my result with the previous literature, I form a term factor $term_t$ based on the difference in returns between long-term treasury bonds and one month t-bills. I also form a default factor def_t which is the difference in returns between long-term corporate bonds and long-term treasury bonds. I obtain the data for these variables from Ibbotson's Stocks, Bonds, Bills and Inflation Yearbook.

Finally, I merge all four databases using the CUSIP identifiers at the firm and at the issue level.

In what follows I define the key variables that I use for the analysis. The bond premium is defined by the expected returns on the corporate bond in excess of the matching treasury bond.

$$E\left[R^e_{i,t+1}\right] = E\left[R_{i,t+1} - R^f_{i,t+1}\right]$$

Throughout this article, I use the word 'excess returns' on a corporate bond to mean $R_{i,t+1} - R^f_{i,t+1}$. By focusing on the returns in excess of the matching treasury bond returns, I eliminate the mechanical effect of shocks to the treasury yield curve on the corporate bond returns. Thus, this definition of excess returns allows me to study the risk-return trade off that is unique in the corporate bond market.

The bond spread is defined by the relative price scaled by time to maturity $\tau_{i,t}$

$$s_{i,t} = \frac{1}{\tau_{i,t}} \log \frac{P^f_{i,t}}{P_{i,t}}$$

[6] I match the maturity of the zero coupon treasury prices to the cash flow exactly by linearly interpolating continuous compounding forward rates.

Lastly, the default loss is defined by

$$loss_{i,t^*} = \log \frac{P_{i,t^*}}{P^f_{i,t^*}}$$

where t^* is the time of default.

4 Cross-section of Corporate Bond Premia

4.1 Portfolios Sorted by Bond Spread

To understand the information contained in the variation of the individual bond spreads, I start with a classic portfolio analysis of the cross-section of corporate bonds sorted on bond spreads. At the end of July every year, I form ten portfolios based on the average bond spreads between January and June. I record the monthly value-weighted average returns for each portfolio from August to the following July. The one month gap between the time to form portfolios and the time of the information used to sort bonds is necessary to remove high frequency market microstructure noise which may induce spurious return forecastability. I also take six-month averages of the bond spread to average out the potential measurement errors in bond prices.

A preliminary analysis shows that there is a distinct jump in the average excess returns between the ninth and tenth portfolios. To see the information in the last decile better, I split the tenth portfolio into three sub-portfolios based on bond spreads when rebalancing portfolios.

The top panel of Table 1 shows the average monthly portfolio returns in excess of matching treasuries. The average excess returns on corporate bonds rise monotonically from -0.01% per month[7] to 0.64% per month as they move from the lowest decile to the last decile. The difference in average excess returns between the highest and the lowest decile is 0.64% with t-statistics 3.2.

The excess returns on these portfolios are serially correlated, as shown in the lag k autocorrelation coefficients AR_k. The signs of the AR_1 coefficients suggest that bonds with a high price (low bond spread) tend to be followed by a bad return while bonds with a low price (high bond spread) also tend to be followed by a bad return. Except for the seventh and eighth deciles and the last subportfolio (10c), the AR_1 coefficients are statistically significant at the 5% level. On the other hand, the AR_2 and AR_3 coefficients (not reported) are all statistically insignificant.

There are two potential reasons why the autocorrelations are present in the data. First, it is possible that high frequency measurement errors cause spurious mean reversion in prices. Second, individual bond returns should be negatively autocorrelated if the variation in the bond spread is associated with the news about expected excess returns, as opposed to the news about expected

[7] The negative average excess returns might be due to i) the interpolation in treasury yield curve, ii) Omission of callablility of corporate bonds and iii) recording errors in the data. Since the negative average excess returns is economically and statistically small, I will not impose further filters just to remove it.

Table 1: Portfolios Sorted On Bond Spreads (Percentage Per Month)

Sample Statistics:

	1(low)	2	3	4	5	6	7	8	9	10	10a	10b	10c	10-1
$E(R^e)$	-0.01	0.04	0.04	0.05	0.04	0.05	0.17	0.18	0.21	0.64	0.42	0.68	0.91	0.64
$se[E(R^e)]$	(0.08)	(0.08)	(0.07)	(0.07)	(0.08)	(0.08)	(0.09)	(0.10)	(0.13)	(0.23)	(0.19)	(0.25)	(0.31)	(0.20)
$AR_1(R^e)$	-0.12*	-0.14*	-0.17*	-0.14*	-0.19*	-0.14*	-0.06	0.05	0.14*	0.26*	0.19*	0.29*	0.07	
$AR_2(R^e)$	-0.04	-0.06	-0.05	-0.05	-0.05	-0.04	-0.06	-0.05	-0.03	0.01	-0.03	0.04	-0.03	
$E[s]$	0.03	0.04	0.05	0.06	0.07	0.09	0.11	0.14	0.20	0.30	0.28	0.33	0.37	
$E[loss]$	0.00	0.00	0.00	0.00	0.03	0.02	0.01	0.01	0.03	0.20	0.09	0.19	0.41	
$E[\tau]$	14.94	14.44	13.48	12.67	12.01	11.60	11.23	10.20	8.94	9.58	8.31	7.63	11.86	
$E[MV]$	7.91	7.05	6.70	6.53	6.23	5.79	5.36	5.28	4.82	4.36	1.51	1.43	1.42	

Issuer Information:

	1(low)	2	3	4	5	6	7	8	9	10	10a	10b	10c	10-1
MV_{eq}	22.71	14.06	10.70	7.63	6.32	4.56	3.80	2.56	1.22	1.16	0.93	0.66	1.74	-20.97
B/M_{eq}	0.29	0.36	0.40	0.38	0.42	0.51	0.52	0.56	0.60	0.52	0.57	0.59	0.46	0.17
mom_{eq}	0.32	0.40	0.42	0.38	0.39	0.41	0.44	0.51	0.42	0.43	0.46	0.47	0.34	0.02

Time-series Regression: $R^e_{i,t} = \alpha_i + \beta_{1,i} Level_t + \beta_{2,i} Slope_t + \varepsilon_{i,t}$

	1(low)	2	3	4	5	6	7	8	9	10	10a	10b	10c	10-1
α	-0.03	0.01	0.00	0.01	0.01	-0.01	0.03	0.01	-0.08	0.03	-0.08	0.05	0.17	0.05
$se(a)$	(0.04)	(0.02)	(0.02)	(0.03)	(0.02)	(0.04)	(0.02)	(0.03)	(0.06)	(0.03)	(0.06)	(0.10)	(0.13)	(0.03)
β_1	0.27	0.30	0.30	0.29	0.31	0.33	0.33	0.31	0.31	0.39	0.34	0.41	0.45	0.12
$se(\beta_1)$	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.04)	(0.01)
β_2	-0.24	-0.25	-0.22	-0.19	-0.23	-0.17	-0.05	0.04	0.26	0.80	0.62	0.79	0.99	1.04
$se(\beta_2)$	(0.03)	(0.02)	(0.01)	(0.01)	(0.03)	(0.02)	(0.02)	(0.02)	(0.05)	(0.02)	(0.09)	(0.07)	(0.12)	(0.03)
R^2	0.79	0.89	0.94	0.91	0.89	0.84	0.88	0.85	0.77	0.96	0.67	0.80	0.63	

GMM χ^2 test:	10.82	[0.545]

At the end of July, the bonds are sorted into portfolios according to their average bond spreads between January and June. The returns are recorded from August to next July, when portfolio rebalancing occurs. The time period of the data is from 1973 to 2011. *Level* and *Slope* correspond the first two principal components of excess returns. $AR_k(\cdot)$ is the univariate autocorrelation coefficient with lag k and $*$ shows that the coefficient is statistically significant at 5% level. s denotes the value-weighted average of the bond spreads on the portfolios. $loss$ is the value-weighted average default loss rate. τ is time to maturity and MV is the total value of the portfolio in billion dollars. MV_{eq} is the (bond) value-weighted market value of the issuers' equities in million dollars. B/M_{eq} is the value-weighted average of the issuers' book-to-market ratio. mom_{eq} is the value-weighted average of the issuers' average equity returns from $t-12$ to $t-2$. Standard errors (in parenthesis) are adjusted for the serial correlations up to 12 lags following Newey and West (1987). The GMM χ^2 test shows the test of the hypothesis that the intercepts of the regressions are jointly zero. The p-value is shown in the bracket.

default. If the bond price falls due to the news about discount rates, the return on the bond going forward must be higher on average.

The second source of serial correlation in returns is analyzed using a term structure model in Nozawa (2012). The lesson from this exercise is that the individual bond returns show negative autocorrelations when the bond is held until maturity. Even so, the serial correlation is not too persistent and accounting for 12 to 24 lags is enough to conduct statistical inference. As I analyze portfolios of bonds where each individual bond can leave or join the portfolio in the middle of its life, the serial correlation of portfolio returns must be attenuated relative to the model-implied serial correlations in Nozawa (2012). Thus, in the following analysis, I account for 12 lags using the weights of Newey and West (1987) when computing standard errors.

The average ex-post[8] bond spread (per month) $E[s]$ and the default loss rate $E[loss]$ for the portfolios rise almost monotonically from the lowest decile to the highest, with a sharp rise from the ninth decile to the tenth decile. The bonds in the first four deciles never default in this sample. The loss rate in the ninth decide is only 3 basis points per month while it rises to 20 basis points for the highest decile. The fact that the first four deciles have positive spreads and yet have no defaults at all does not imply that there is an arbitrage opportunity. An investor who buys a bond in the first four deciles may still suffer from the news about the credit event, as such news can send the bond into lower deciles with lower prices. A default may occur after the bond migrates to the lower deciles, but the lower returns due to the news is recorded in the original low spread portfolio.

The bonds in the lowest spread decile have longer time to maturity on average (high $E[\tau]$) and larger total market value (high $E[MV]$). Also, an analysis on the equities on the issuers of these bonds shows an interesting pattern. The issuers of the high spread bonds have small equity market values. On the other hand, there is no significant variation in the issuers' book-to-market ratio[9] B/M_{eq} and the past average equity returns mom_{eq}. Though B/M_{eq} rises from the first decile to the tenth decile, the variation is not economically large. This result seems contrary to the idea that the book-to-market ratio is the proxy for financial distress.

4.2 Two Factor Model

To examine if the variation in average excess returns across the bond spread sorted portfolios can be explained by a risk exposure, I form a simple factor model consisting of the first two principal components of excess returns on ten portfolios sorted on bond spread.

The third panel of Table 1 shows the loadings of the two principal components on the ten portfolios. The first principal component is a level factor ($Level_t$) which has nearly the same loadings on all ten portfolios, as shown in β_1. $Level_t$ explains 73% of the common variation in

[8] The ex-post bond spread is the spread observed after the portfolio formation, which can be different from the bond spread used to form the portfolios.

[9] Though the high spread portfolios tend to have slightly higher book-to-market ratio, the difference is economically small. The value firms in Fama and French (1996) have the book-to-market ratio greater than one.

Table 2: Summary Statistics of Two Principal Components (Percentage Per Month)

	Descriptive Statistics of Principal Components:					Factor Correlations:					
	$E[\cdot]$	$\sigma[\cdot]$	*skew*	*kurt*	$E[\cdot]/\sigma[\cdot]$	$RMRF_t$	SMB_t	HML_t	$term_t$	def_t	$dVIX_t$
$Level_t$	0.49	5.71	0.04	1.24	0.09	0.26	0.11	-0.03	0.03	0.33	-0.43
(se)	(0.27)		(0.11)	(0.23)							
$Slope_t$	0.52	2.61	0.52	3.87	0.20	0.29	0.28	0.02	-0.08	0.35	-0.23
(se)	(0.12)		(0.11)	(0.23)							

Monthly returns from 1973 to 2011. $Level_t$ and $Slope_t$ are the first and the second principal components of the excess returns on ten bond spread sorted portfolios. *skew* is skewness and *kurt* is excess kurtosis. $RMRF_t$, SMB_t and HML_t are the three factors of equity returns of Fama and French (1993). $term_t$ is a return on long-term treasury bonds minus t-bills. def_t is the return on long-term corporate bond index minus long-term treasury bonds. $dVIX_t$ is the change in VIX index. The correlation between $dVIX_t$ is computed using the data from 1990 to 2011.

realized excess returns. The pattern in β_2 shows that the second principal component is a slope factor ($Slope_t$) whose loadings monotonically go up from the lowest decile to the highest decile. $Slope_t$ captures another 15% of the common variation. The fact that $Level_t$ and $Slope_t$ explain nearly 90% of the common variation in excess returns shows that there is a strong factor structure in excess returns on the ten portfolios.

The summary statistics of $Level_t$ and $Slope_t$ are shown in the left panel of Table 2. $Level_t$ has statistically insignificant average excess returns. The average excess returns scaled by the standard deviation is low at 0.09. $Slope_t$, on the other hand, has the high ratio of average excess returns to standard deviation (0.20). The high risk price of $Slope_t$ raises the possibility that $Slope_t$ may proxy for a shock to an important state variable that is of concern to the average investor.

To study the nature of the risk in the two factors, the top panel of Figure 1 shows the cumulative log returns on $Level_t$ and $Slope_t$. The gray area in the background shows the recessions identified by NBER. $Slope_t$ is highly cyclical and tends to fall in the first half of recessions before subsequently recovering. For the last three recessions, the fall in $Slope_t$ actually precedes each recession, which suggests that $Slope_t$ somewhat forecasts these recessions. Furthermore, Table 2 shows that $Slope_t$ is fat-tailed with excess kurtosis of 3.87. The cyclicality and fat-tailness of $Slope_t$ show that the high ratio of average excess returns to the standard deviation for $Slope_t$ is not necessarily a proof of market inefficiency where a good investment opportunity is left on the table. In fact, $Slope_t$ seems like a risky investment strategy.

The right panel of Table 2 shows the correlation of $Level_t$ and $Slope_t$ with the five bond and equity pricing factors of Fama and French (1993). The correlations with the three equity factors are in general low. The highest correlation is only 0.29 between $Slope_t$ and the excess returns on equity market portfolio ($RMRF_t$). Though issuers of high spread bonds have small equities, the correlation between $Slope_t$ and SMB_t is only 0.28. This low correlation between $Slope_t$ and SMB_t suggests that the equity and the bond markets are not fully integrated at monthly frequency.

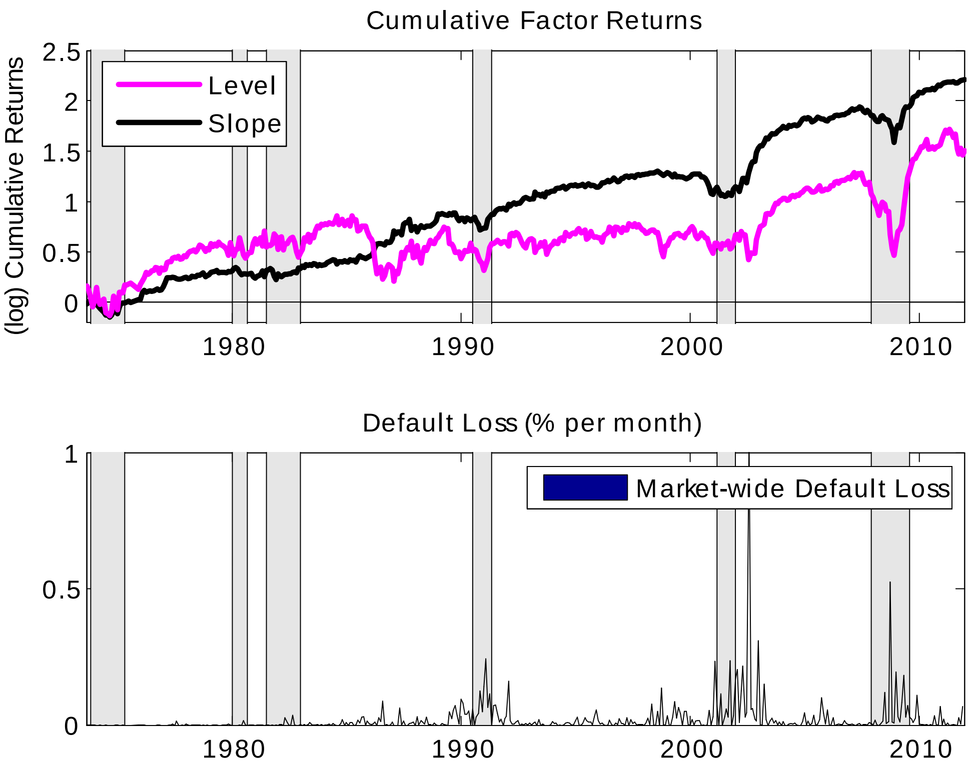

Figure 1: Cumulative Factor Returns and Aggregate Default Loss Rate (value-weight average of default loss of individual bonds)

The correlations between the two factors and the change in the VIX index $dVIX_t$ are negative. This is expected as an increase in volatility tends to increase an option value. Since a long position in defaultable bonds are equivalent as a short position in put options on firm's asset value and a long position in a treasury bond, an increase in the VIX should affect the defaultable bond returns negatively. However, the correlation between $dVIX_t$ and $Level_t$ is only -0.43 and I cannot replace one of my factors with $dVIX_t$ to keep the pricing capability of the two factor model.

I test if the two factor model consisting of $Level_t$ and $Slope_t$ can explain the cross-sectional variation in average excess returns. The third panel in Table 1 shows the estimated coefficients for the following time-series regression

$$R^e_{i,t} = \alpha_i + \beta_{1,i} Level_t + \beta_{2,i} Slope_t + \varepsilon_{i,t} \tag{1}$$

and the corresponding standard errors. If the variation in bond premia is explained by the two-factor model, then

$$E\left[R^e_{i,t}\right] = \beta_{1,i} E\left[Level_t\right] + \beta_{2,i} E\left[Slope_t\right]$$

must hold.

I estimate (1) for all deciles (which I use to form my factors) and the three subportfolios of the last decile.

In Table 1, β_1 is around 0.3 for all portfolios but the extreme deciles, and the loading on the spread factor β_2 rises monotonically from -0.24 (the lowest decile) to 0.99 (10c of the highest decile). The high R-squared in each regression shows that the two factors explain most of the time-series variation in excess returns, which we expect from the fact that my factors are principal components. The pricing errors, α, are small in general, ranging from -0.06% to 0.17% per month and statistically insignificant. The GMM χ^2 test which tests if the intercepts are jointly zero fails to reject the two-factor model at the 10% level. (This test would be the same as the joint test of Gibbons, Ross and Shanken (1989) if there is no lags in computing standard errors. Since the returns show autocorrelation, I account for it for robust statistical inference.)

The fact that the betas with respect to $Slope_t$ match the cross-section of bond premia suggests that corporate bond premia are risk premia. That is, investors may demand the premia on corporate bonds due to their comovements with some systematic risks.

The corporate bonds with a high bond spread earn high average excess returns for three potential reasons. First, the average investor may dislike the bond that is likely to default and therefore demand a premium. Second, the high spread bonds are illiquid and costly to trade. Thus, the bond premia for high spread bonds reflect the expected transaction cost which incurs at the time of sale in the future. Third, the high spread bonds comove positively with the average investor's marginal utility of consumption and are exposed to the systematic risk proxied by $Slope_t$.

The fact that the $Slope_t$ betas match the pattern of average excess returns of corporate bonds can be consistent with any three of these explanations above. A challenge for the irrational preference or transaction-based explanations is that they also need to be consistent with the factor structure of excess returns. A bond with high transaction costs should earn a high bond premium, but why does it have to comove with other bonds with high transaction costs? Either way, in this article, I do not offer any direct evidence in support of one explanation over the other.

4.3 Other Characteristics

To further examine the performance of the two-factor model consisting of the level and slope factors, I sort the bonds into deciles according to the fourteen characteristics listed in Table 3. I use previous literature in finding the characteristics associated with bond premia. I use four price-related variables: current yield cy, market value of a bond $size$, the past one year average excess returns (excluding the last month) mom and the yield to maturity on the synthetic treasury that corresponds to the corporate bond i, try. The bond specific characteristics used in the analysis are time to maturity τ, duration dur and the time after the issuance of the bond age. The issuer's fundamental variables include the (risk neutral) distance-to-default DD, the ratio of short-term debt to the total debt $sdtd$, the ratio of the fixed assets over the total assets $tangible$ and issuer's credit ratings $rating$. Since credit ratings are expressed with symbols, I transform them into numerical variables such that AAA is 1 and C is 21. Finally, I use equity anomaly variables such as equity size MV_{eq}, equity book-to-market ratio B/M_{eq} and equity momentum mom_{eq}.

The (risk neutral) distance-to-default for firm i at time t is defined by

$$DD_{i,t} = \frac{\ln \frac{A_{i,t}}{D_{i,t}} + \left(r_t - \delta_{i,t} - \frac{1}{2}\sigma^2_{i,t,A}\right)}{\sigma_{i,t,A}}$$

where $A_{i,t}$ is the market value of the firm, $D_{i,t}$ is the default boundary, r_t is the risk-free rate, $\delta_{i,t}$ is the payout rate and $\sigma_{i,t,A}$ is the conditional volatility of the returns on the firm asset. The computational detail is shown in Appendix B. Roughly speaking, the market value of the firm is identified such that the call option on the firms' value calculated using the Black-Scholes formula is equal to the observed stock price in the data. If Merton's (1973) model is correct, the cumulative density of standard normal distribution evaluated at $-DD_{i,t}$ should be equal to the risk-neutral probability of default.

The motivation for using the characteristics listed in Table 3 is clear for some variables such as equity anomaly variables but not for others. I use $sdtd$ because the maturity structure of the firm is associated with the rollover risk suggested by He and Xiong (2012). I test if $tangible$ is associated with expected excess returns since it is used as a proxy for the tangibility of assets which serve as collateral for borrowing under the incomplete contract theory. Though the corporate

Table 3: List of Characteristics for Corporate Bonds

	Name	Definition
Price Related:		
s	Bond spread	$1/\tau \log P^f_{i,t}/P_{i,t}$
cy	Current yield	Coupon rate (per face value) divided by $P_{i,t}$
$size$	Size	Amount outstanding multiplied by $P_{i,t}$
mom	Bond momentum	Average bond return between $t-12$ and $t-2$
try	Treasury yield	Yield to maturity of the matching treasury bond at time t
Issue-specific:		
τ	Time to maturity	Time to maturity of the bond in years
dur	Duration	Macaulay duration of the bond
age	Age	Time passed since the issuance of the bond in years
Issuer's fundamentals:		
DD	Distance to default	Risk-neutral distance-to-default of Merton's model
$Sdtd$	Sdtd	Short-term debt (book value) / Total debt (book value)
$tangible$	Tangibility	Plant, Property and Equipment (book value) / Total asset (book value)
$rating$	Ratings	Credit ratings by Moody's or S&P
Equity Anomalies:		
MV_{eq}	Equity size	Market value of equity
B/M_{eq}	Equity book-to-market ratio	Book value of equity divided by market value of equity
mom_{eq}	Equity momentum	Average equity return between $t-12$ and $t-2$

bonds I analyze are uncollateralized bonds, higher debt capacity of the issuer might still affect its bond premia.

The use of age as a forecaster of the returns is motivated by Bao, Pan and Wang (2010) who show that the bond's age is related to their measure of liquidity[10]. The use of try is motivated by the findings of Gebhardt, Hvidkjaer and Swaminathan (2005) that the yield to maturity of corporate bonds is a strong predictor of returns.

The expected excess returns on the ten portfolios sorted on characteristics as well as the regression estimates of equation (1) for each portfolio are shown in Table 4. For brevity, I report the test result only for the first and last deciles.

Table 4 shows that the two factor model explains most of the variation in average excess returns for portfolios sorted on all characteristics but mom_{eq}. Among the characteristics I test, cy, $size$, DD, $sdtd$, MV_{eq} and B/M_{eq} give rise to statistically significant variations in estimated bond premia. The root mean squared alpha $RMSE_\alpha$ and the mean squared alpha MAE_α are economically small for these characteristics, as the pricing errors are mostly less than 5 basis points per month. Also, these alphas are small relative to the variation in estimated bond premia (as shown in the difference between *high* and *low* portfolios), which shows that the model performs well in pricing these portfolios.

[10] Driessen (2005) also uses age as a proxy for liquidity.

Table 4: Asset Pricing Test with Variety of Characteristics (Percentage Per Month)

	1 (*low*)		10 (*high*)								
Characteristics	$E[R^e]$	α	$E[R^e]$	α	$10-1$	(*se*)	$RMSE_\alpha$	MAE_α	χ^2	$pv[\chi^2]$	avg R^2
Price-related											
cy	0.07	0.02	0.41	-0.01	0.34	(0.16)	0.03	0.02	10.6	[0.39]	0.81
size	0.33	0.17	0.02	-0.08	-0.31	(0.11)	0.07	0.05	13.7	[0.19]	0.76
mom	0.23	-0.04	0.11	-0.03	-0.12	(0.10)	0.02	0.02	5.7	[0.84]	0.75
try	0.14	0.06	0.04	-0.05	-0.10	(0.09)	0.05	0.05	12.8	[0.23]	0.71
Issue-specific											
τ	0.08	0.01	0.01	-0.07	-0.07	(0.09)	0.04	0.03	24.3	[0.01]	0.78
dur	0.09	0.02	-0.04	-0.08	-0.13	(0.08)	0.04	0.03	20.3	[0.03]	0.77
age	0.07	-0.01	0.08	0.00	0.00	(0.05)	0.03	0.03	16.1	[0.10]	0.78
Issuer's fundamentals											
DD	0.43	0.03	0.03	0.01	-0.40	(0.13)	0.03	0.02	5.7	[0.84]	0.75
sdtd	0.11	0.00	0.01	-0.06	-0.10	(0.05)	0.02	0.02	4.4	[0.93]	0.73
tangible	0.05	-0.08	0.06	-0.03	0.01	(0.06)	0.05	0.04	14.1	[0.17]	0.72
rating	-0.01	-0.01	0.33	-0.06	0.34	(0.17)	0.03	0.02	5.0	[0.89]	0.76
Equity Anomalies											
MV_{eq}	0.31	0.01	-0.03	-0.07	-0.34	(0.12)	0.04	0.03	22.2	[0.01]	0.76
B/M_{eq}	0.05	-0.03	0.24	0.03	0.19	(0.08)	0.04	0.03	10.3	[0.42]	0.76
mom_{eq}	0.03	-0.26	0.21	0.09	0.18	(0.12)	0.09	0.06	44.6	[0.00]	0.77
All Bond/Issuer Characteristics (High and Low only)							0.05	0.04	34.21	[0.06]	
All Characteristics (High and Low only)							0.07	0.05	69.81	[0.00]	

Asset pricing test using time series regressions: $R^e_{i,t} = \alpha_i + \beta_{1,i} Level_t + \beta_{2,i} Slope_t + \varepsilon_{i,t}$. Monthly returns from 1973 to 2011. The bonds are sorted into 10 portfolios every year according to characteristics except for *mom* and mom_{eq}. For *mom* and mom_{eq}, the portfolios are sorted every month. Low is the first decide and High is the tenth decile and H-L shows the difference between the tenth and first deciles. $RMSE_\alpha \equiv \sqrt{\frac{1}{10}\sum_{i=1}^{10}\alpha_i^2}$ and $MAE_\alpha \equiv \frac{1}{10}\sum_{i=1}^{10}|\alpha_i|$. χ^2 is the GMM test statistics of the null that the intercepts are jointly zero for all ten deciles and χ^2 [*pv*] is the corresponding p-value. avg R^2 is the R-squared of the time-series regression averaged over 10 portfolios. The characteristics used are defined in Table 3. The standard errors are in parenthesis and computed using Newey and West (1987) 12 lags.

The GMM χ^2 test which tests if the intercepts α_i are jointly zero for all of the ten portfolios mostly fails to reject the model except for τ, dur, MV_{eq} and mom_{eq}. For τ, dur and MV_{eq}, the pricing errors are economically small as $RMSE_\alpha$ is 4 basis points and MAE_α is 3 basis points. Therefore, the statistical rejection of the model seems to come from the fact that these alphas are very well measured. In such cases, the two factor model can still be used as a reasonable description of the bond premia.

On the other hand, the two factor model fails to explain the variation in estimated bond premia associated with mom_{eq}. In this case, the model works in the 'wrong' direction in the sense that the magnitude of alphas is greater than the magnitude of estimated bond premia. For the first decile (*low* portfolio), the estimated bond premium is only 3 basis points while the magnitude of the alpha is as large as -26 basis points. This large alpha leads to relatively large $RMSE_\alpha$ (9 basis points) and the statistical rejection of the model.

The last two rows of Table 4 show the joint test if the intercepts are zero for multiple characteristics. The second last row shows the test if intercepts are jointly zero for the first and the last deciles of the first eleven characteristics (from cy to $rating$). In other words, I am testing if the 22 portfolios in the high end and the low end of the 11 *univariate* sorts have jointly significant alphas. $RMSE_\alpha$ is small at 5 basis point, which is good given that I am using only extreme deciles and thus the variation in estimated bond premia is large. The model cannot be rejected at the conventional 5% level. This result suggests that the two factor model can price the portfolios associated with multiple characteristics at once, when portfolios are formed using univariate sorts.

On the other hand, when I test if the intercepts are jointly zero for the first and the last deciles of all of the 14 characteristics (i.e. 28 portfolios in total), the model is rejected at the 5% level. Including the equity momentum sorted portfolios causes a problem for the two factor model again.

Another interesting point in Table 4 is that the factor structure of corporate bond excess returns. avg R^2 in Table 4 shows the R^2 of time-series regressions averaged across the ten portfolios for each characteristic. avg R^2 is in general high and ranges from 0.7 to 0.8. The high avg R^2 implies that the first two principal components constructed from the portfolios sorted on bond spread also explain the common variation in excess returns on other portfolios. The common factor structure in corporate bond returns seems to exist no matter what characteristics we use to sort bonds.

The high avg R^2 in realized return is comforting as I can motivate my two factor model by the Arbitrage Pricing Theory of Ross (1976). On the other hand, the high avg R^2 raises a question why I use the two principal components of the bond spread-sorted portfolios rather than other principal components. Maybe the first two principal components of the portfolios sorted on current yields will do an equally good job. Instead of showing the empirical result using 14 other potential sets of factors against hundreds of potential test assets, I support my choice by claiming that the bond spread is the cleanest price variable associated with bond premia.

For an individual bond, the bond spread must forecast its excess return or its default loss. For a portfolio, the bond spread must forecast the portfolio excess returns or the credit shocks to the portfolio which cause individual bonds in the portfolio to migrate to another portfolio. Bond spreads are the variables that should have information about the distribution of future excess returns on corporate bonds. Therefore, it is the most natural variable to sort corporate bonds and form factors. By sorting securities using bond spreads, I can alleviate the concern that my two factors earn premia and thus work as a pricing kernel only in my sample period. Unless one firmly believes that the greater bond spread always forecasts negative credit events (such as defaults), it must forecast excess returns and therefore $Slope_t$ should earn the premia.

4.4 Comparison with Gebhardt, Hvidkjaer and Swaminathan (2005)

To clarify the contribution of this paper to the literature, I also test the two factor model of Gebhardt, Hvidkjaer and Swaminathan (2005) which is based on $term_t$ and def_t of Fama and French (1993), using the same test assets as in Table 1 and Table 4. The detailed result is available upon request and I summarize my findings here. When I test if the two factor model consisting of $term_t$ and def_t can price the ten portfolios sorted on the bond spread, I find that the model has no power to explain the variation in estimated bond premia.

The reason for the poor performance of $term_t$ and def_t is simple: these two factors have nearly zero average returns in my sample and the resulting alphas are basically the same as the estimated bond premia. For example, def_t, which takes a long position in long-term corporate bonds and takes a short position in a long-term (about 20 years) treasury bond, has the average excess returns of -1 basis point per month in my sample. This value is lower than 2 basis points reported in Fama and French (1993) who use the sample period from 1963 to 1990. Gebhardt, Hvidkajaer and Swaminathan (2005) construct def_t in a slightly different manner but following the same principle (long-term corporate bond returns based on their sample minus long-term treasury bond returns) and find that def_t earns the average excess returns of 4 basis points over the sample period from 1973 to 1996. Since my sample includes the financial crisis of 2008, it makes sense that def_t has even lower average excess returns in my test. Due to the near zero average excess returns, the two factor model consisting of $term_t$ and def_t have nearly no explanatory power of the test assets analyzed in this article.

5 Parametric Characteristic-based Asset Pricing Test

5.1 Idea

In the previous section, I test the two factor model using traditional portfolio sorts based on one characteristic in each test. This is not satisfactory as investors can take advantage of multiple characteristics at once in forming their investment strategy. A simple way to address this concern is multidimensional sorts. Just as Fama and French (1993) sort stocks based both on size and book-to-market ratio, I can do two-way sorts of corporate bonds. However, a higher dimensional sort is impractical particularly when there is a lot of information in the tail of the distribution of characteristics. In Table 1, splitting the last decile into three subportfolios uncovers a large variation in bond premia. If an economist does the same for multiple characteristics, she will quickly run out of data points to form portfolios. This problem is even worse for corporate bonds since the size of my sample is small in the 1970s. For example, I only have 19 bonds in January 1973. The small sample size of the corporate bond data makes it necessary for a researcher to be more resourceful in using the data.

In this section, I characterize bond premia as a smooth function of characteristics. I also characterize risk as a function of characteristics, and test if these two functions match. I call this approach the parametric characteristic-based approach.

To get an idea of the difference between portfolio sorts and the approach adopted here, I make two plots which describe these tests in Figures 2 and 3. Figure 2 is a graphical reproduction of the result of Table 1. In the classic test, we express the bond premia and betas as a function of portfolio rankings which reflect the variation in bond spreads. As in the figure, the estimated betas for $Slope_t$ go up as we move from left to right, so do the average excess returns. Since the increasing pattern in the average excess returns is matched by the increasing patterns in the betas, the model prices these portfolios successfully.

The top panel of Figure 3 shows the corresponding result of the parametric characteristic-based test. Before going into the detail, I will explain the intuition of the test. In this case, I express the bond premia and a measure of the risk, expected comovements (that will be defined clearly below), as a smooth function of bond spreads. In this example, I express the bond premia and expected comovements as functions of a bond spread, a squared bond spread and square-root of a bond spread.

Figure 3 also shows that the model seems successful in pricing the variation in bond premia associated with bond spreads. The expected excess return function seems to be matched well by the model-implied counterpart, which is a linear combination of the two expected comovement functions $E[V]$. In particular, the expected comovements with the slope factor $Slope\ E[V]$ seems parallel to the expected excess return function in the data, which suggests that the slope factor is the key to the success of the model.

Using a smooth function reduces the number of parameters to be estimated, which makes it easy to accommodate multiple characteristics in an asset pricing test. Of course, the benefit comes with some costs. If the assumed functional form is incorrect, then the test becomes misspecified. However, in this case, there seems to be little nonlinearity in the expected excess returns. To see this, I re-express the point estimates of bond premia from portfolio sorts as a function of bond spreads (not a function of ranking) in the top panel of Figure 3. The dotted line with circles

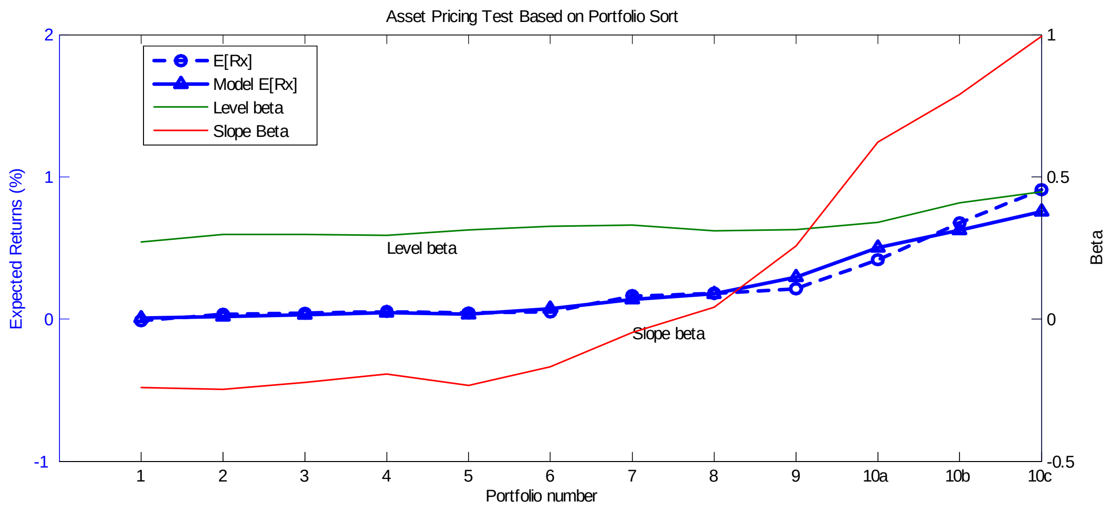

FIGURE 2: ASSET PRICING TEST BASED ON PORTFOLIO SORTS

shows the estimated bond premia based on portfolio sorts[11]. This non-parametric estimates of bond premia do not produce a complex functional form of bond premia. Rather, the estimated bond premia seem to go up smoothly from left to right.

When we look at the pattern in estimated bond premia in Table 1, we might conclude that the bond premia jump up in the last decile and it is hard to replicate such a jump with a smooth function. When we express the bond premia as a function of underlying characteristics, the jump goes away and it seems feasible to fit a smooth function.

The bottom panel of Figure 3 shows the histogram of the bond spread. As we can see, the distribution is highly skewed and the most of the observations are below 0.2% per month. The white and gray colors in the panel correspond to the cutoff values (averaged over time) for the ten portfolios analyzed in Table 1. For example, the white area at the left end of the figure shows these bonds will be mostly in portfolio one, while the gray area at the center to the right shows that these bonds will be mostly in the last decile. This means that when sorting securities into bins, we are throwing a lot of variation in characteristics away by averaging all the extreme observations in the last decile.

Of course, the point of fitting a functional form is to use multiple characteristics at once, not just bond spreads as shown in Figures 2 and 3. Based on the intuition developed by comparing Figures 2 and 3, I formally define the parametric characteristic-based approach in the next section.

[11] Strictly speaking, the data points in Figure 3 differs from the portfolio sorts in Table 1 as the figure shows the result of the regression (3) using dummy variables. Since the weights assigned to each security in regression (3) is different from portfolio sorts, there is a slight gap in the estimated bond premia between Table 1 and Figure 3.

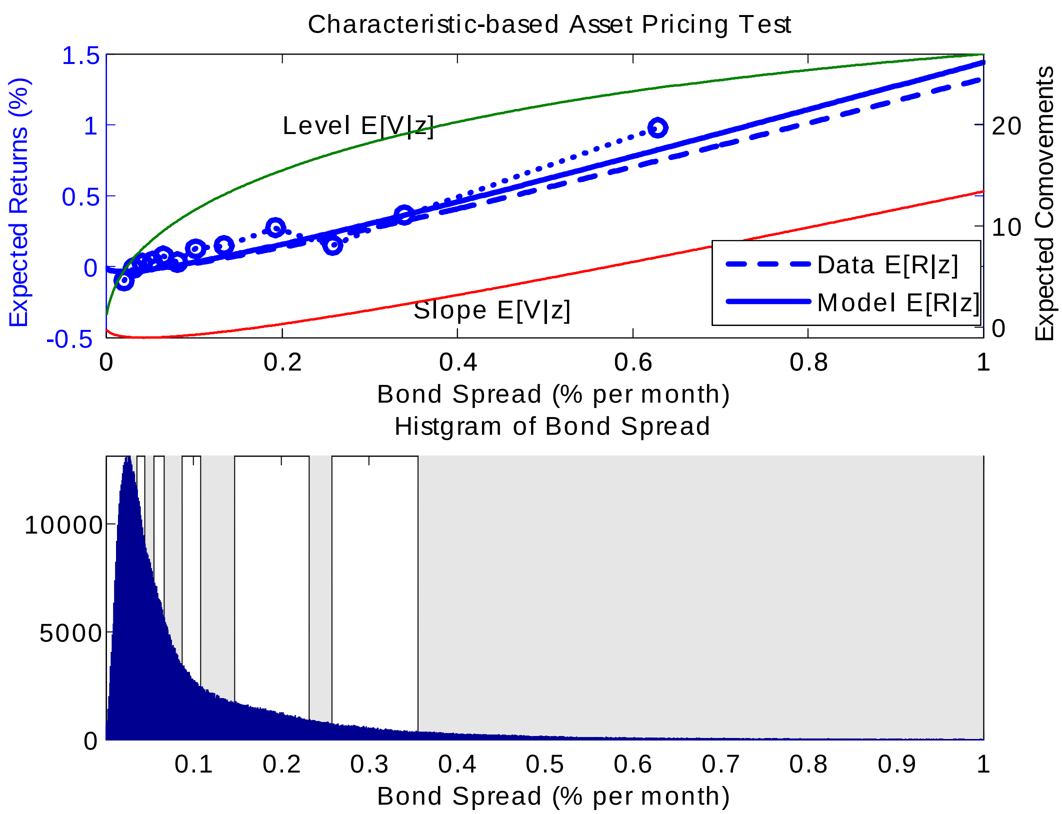

Figure 3: Parametric Characteristic-based Test using s, s^2 and $s^{0.5}$ as instruments (Top) and the Distribution of Bond Spread (Bottom): E[R|z] is the expected excess return function and E[V|z] is the expected comovement function. The dotted line with circles are expected excess returns using 12 dummy variables corresponding to portfolios.

5.2 Three-step Procedure

I describe the parametric characteristic-based asset pricing test using a simple three-step estimation procedure. I test the model which specifies K factors $f_t = (f_{1,t}, \ldots, f_{K,t})$ driving the stochastic discount factor. Let $z_{i,t}$ be the L dimensional vector of instruments for bond i observed at time t. Traditionally, these instruments are macro variables such as lagged market returns or consumption growth rates (for example, see Ferson and Harvey (1991)). In the characteristic-based test, $z_{i,t}$ includes smooth functions of characteristics $c_{i,t}$ of each asset i at time t. The characteristics drive the individual securities' bond premia and expected comovements with factors.

Let $V_{i,t+1} \equiv R^e_{i,t+1} f_{t+1}$. The asset pricing model states that

$$E\left[R^e_{i,t+1} | c_{i,t}\right] = E\left[V_{i,t+1} | c_{i,t}\right] \lambda \tag{2}$$

where $c_{i,t}$ is a vector of characteristics and λ is a K dimensional vector of factor risk weights. That is, if the model is correct the expected excess return function is equal to a linear combination of the expected comovement functions $E\left[V_{i,t+1} | c_{i,t}\right]$.

I test (2) using a simple three-step estimation procedure. First, I pick an instrument vector $z_{i,t}$ which is a function of characteristics $c_{i,t}$. I estimate the expected excess return function by the pooled OLS regression

$$R^e_{i,t+1} = z_{i,t} b_r + \varepsilon^r_{i,t+1} \qquad\qquad i = 1, \ldots, N,\ t = 1, \ldots, T \tag{3}$$

where N is the number of bonds and T is the number of months. Recall that i denotes *individual* bond i as opposed to a portfolio. I forecast individual security excess returns $R^e_{i,t+1}$ with the function of its characteristics $z_{i,t}$.

Second, I estimate the expected comovement function by the pooled OLS regression

$$V_{i,t+1} = z_{i,t} b_v + \varepsilon^v_{i,t+1} \qquad\qquad i = 1, \ldots, N,\ t = 1, \ldots, T \tag{4}$$

(3) and (4) yield the estimated functions $\hat{E}\left[R^e_{i,t+1} | c_{i,t}\right] = z_{i,t} \hat{b}_r$ and $\hat{E}\left[V_{i,t+1} | c_{i,t}\right] = z_{i,t} \hat{b}_v$, where $\hat{b}_r$ and $\hat{b}_v$ are the OLS estimates of b_r and b_v respectively.

Third, I estimate the factor risk weight λ by regressing estimated expected excess returns on estimated expected comovements. That is, I run the pooled OLS regression

$$z_{i,t} \hat{b}_r = z_{i,t} \hat{b}_v \lambda + \alpha_{i,t} \qquad\qquad i = 1, \ldots, N,\ t = 1, \ldots, T \tag{5}$$

Let $\hat{\lambda}$ be the estimate of λ. The sample counterpart of equation (2) implies that

$$\hat{b}_r = \hat{b}_v \hat{\lambda} \tag{6}$$

That is, the parameters that characterize the expected excess return function must be completely matched by a linear combination of the parameters of the expected comovement functions. Since $z_{i,t}$ is L dimensional, we need to match the L dimensional vector b_r with a combination of K vectors that are the columns of b_v (b_v is a $L \times K$ matrix).

This procedure is closely related to the classic asset pricing test based on portfolio sorts. In the characteristic-based test, the expected excess returns are estimated using the average excess returns on the dynamic investment strategy based on $z_{i,t}$. Each entry of $z_{i,t}$ gives us the weight on security i to execute the strategy.

In the classic test, the portfolios are formed by sorting securities into bins. Suppose that $z_{i,t}$ are dummy variables which take a value of one if the observation is for a particular portfolio. Then running regression (3) is equivalent to taking sample average of realized portfolio returns and running (4) is essentially the same as taking average of realized comovements at the portfolio level. The third regression (5) is a cross-sectional regression using portfolio average excess returns and portfolio average comovements. Therefore, the asset pricing test based on classic portfolio sorts is a special case of the characteristic-based approach. The key difference between classic portfolio sorts and the parametric approach lies in the choice of instruments. The instruments I use are smooth functions of characteristics, not dummy variables associated with characteristics.

The average excess returns on the portfolios formed by sorting securities into bins are the non-parametric estimates of the expected excess return function. In the parametric characteristic-based approach, I essentially form portfolios by running regressions and estimate the expected excess return function parametrically.

The dimension of $z_{i,t}$ corresponds to the number of test assets in the classic test. Thus for the characteristic-based test to have power, one needs to have $L > K$. Fama and French (1996) test the three factor model using 25 size and book-to-market sorted portfolios and thus have 25 assets to price. In the characteristic-based approach, if expected excess returns are linear functions of size and book-to-market ratio, $z_{i,t}$ will be a two-dimensional vector, and we have two assets to price.

To see if we statistically reject the null hypothesis (6), we can apply the standard χ^2 test result of the GMM estimator. That is, compute

$$E_T\left[z'_{i,t}h_{i,t+1}\right]' cov\left(E_T\left[z'_{i,t}h_{i,t+1}\right]\right)^{-1} E_T\left[z'_{i,t}h_{i,t+1}\right] \sim \chi^2_{L-K} \tag{7}$$

where $E_T[\cdot]$ is a sample average operator and $h_{i,t+1} \equiv R^e_{i,t+1} - V_{i,t+1}\lambda$. This test asks if the alphas for L strategies formed based on $z_{i,t}$ are jointly statistically significant or not.

5.3 Benefits of Parametric Characteristic-based Test

There are a number of benefits of using the parametric characteristic-based asset pricing test instead of the classic test based on portfolio sorts. Rather than going into detail, I list the benefits here and provide supporting evidence in the empirical analysis that follows.

First, with the parametric characteristic-based approach, one can conduct an asset pricing test with multiple characteristics. By sorting securities into portfolios along some characteristics, variations in other characteristics will be lost. Practically speaking, we can only sort securities using up to two or three characteristics even for equities that have a larger sample size. Thus, sorting securities into portfolios makes it impossible to test the performance of the proposed asset pricing model along more than three characteristics at once. Due to this limitation researchers often test the model along one or two characteristics at a time, neglecting the interaction between multiple characteristics. This is the main reason why I adopt this approach in analyzing corporate bonds as there are a few securities in the earlier part of my sample.

Second, for the asset pricing test to be powerful, greater variation in expected comovements is desirable. By sorting securities into bins along particular characteristics, some of the variation in expected comovements is lost, which results in (statistically) less efficient estimates of the factor weight parameter λ_0. This is why Ang, Liu and Schwartz (2010) propose to use individual securities to conduct an asset pricing test.

Third, when sorting securities into portfolios, we essentially impose time fixed effects in estimating expected excess return functions and expected comovement function. That is, a researcher focuses on cross-sectional variation in expected excess returns in evaluating the asset pricing model.

In a classic asset pricing test based on sorts, the test assets are the portfolios sorted on the past characteristics. These portfolios are created based on the strategies of buying the bonds with *relatively* high characteristics at each point in time. If the characteristics of all bonds increase at some time, this trading strategy does not do anything in capturing the variation in expected excess returns observed then relative to the past. Thus in the standard asset pricing test using portfolios, economists ask if the asset pricing model can price the cross-sectional variation in expected excess returns, but not the time-series variation.

With the characteristic-based test with no time fixed effects, the model can be asked to price variation of bond premia both over time and in a cross section. Therefore, using all the individual bonds in a characteristics-based test helps us test an implication of the model other than cross-section of expected excess returns. Since the goal of this paper is to explain the cross-sectional variation in bond premia, I will not take advantage of this flexibility in this article and include time fixed effects in the empirical analysis that follows.

In the classic test based on sorts, a key motivation to form portfolios is to reduce the noise in the estimated betas. In the parametric characteristic-based asset pricing test, the expected excess return function (3) and the expected comovement function (4) can be estimated with all the panel observations of individual bonds. The use of the entire panel of individual bonds makes the estimated expected excess returns and comovements precise given the specification of the functional form. As a result, the concern about noise in the estimates does not invalidate the use of individual bonds as test assets. This point will be quantitatively supported in the next section.

5.3.1 Standard Errors of Expected Comovements

One reason why researchers started to use portfolio sorts was to reduce the noise in estimated betas. The parametric characteristic-based approach addresses the issue by projecting the realized comovements onto characteristics. To show the reliability of the estimated expected comovements, I use a bond spread, a squared bond spread and square-root of a bond spread for $z_{i,t}$ as an illustrating example. I run a regression (4) and I compute the corresponding standard errors of the estimated expected comovements $\sigma\left[\hat{E}\left[V_{i,t+1}^{[k]}|z_{i,t}\right]\right]=\sqrt{z_{i,t}cov\left(\hat{b}_v^{[k]}\right)z_{i,t}'}$ where $V_{i,t+1}^{[k]}$ is the k-th element in $V_{i,t+1}$ and $\hat{b}_v^{[k]}$ is the k-th column of the matrix $\hat{b}_v$. Since the standard error depends on $z_{i,t}$, I pick the percentiles of the distribution of bond spreads.

Table 5 shows the estimates of the standard errors and the estimated expected comovements for the percentiles of the bond spread. As we can see, the standard errors of the expected comovements are small relative to the variation in expected comovements. For $Slope_{t+1}$, the expected comovements vary from -0.7 to 11.9 as the spread increases from the 1st percentile to the 99th percentile, while their standard errors are 0.37 and 3.09, respectively. The difference between the 5th and 95th percentiles are highly statistically significant. Since the precision of the estimates depends on the noise in $\hat{b}_v$, this conclusion does not depend on the particular choice of the percentiles. Overall, these figures show that the estimated expected comovements are well measured despite the fact that I use individual securities.

Sorting securities into portfolios is an effective way to reduce the standard errors of the estimated expected comovements. However, portfolio sorts are not the only way to improve precision of the estimates. What is truly needed is a projection of comovements onto a set of characteristics that forecast comovements well. Portfolio sorts can be thought of as a non-parametric regression, while in contrast the parametric characteristic-based approach uses a parametric regression. The key is the choice of instruments. The fact that the betas of individual securities are typically poorly measured suggests that the dummy variables for the name of securities are not good instruments to project on. The bond spread seems to be a good instrument in this respect and we can do projections either parametrically or non-parametrically.

5.4 Estimating Expected Excess Return Function

To describe the variation in expected excess returns on individual bonds, I run regression (3) using instruments $z_{i,t}$ that are linear functions of characteristics associated with expected excess returns. To see whether the assumption of linearity is satisfactory, in Appendix C, I show a series of analysis about the functional form using non-parametric regression with Gaussian kernel for each characteristic I use. Based on the non-parametric regression result, I transform size and time to

Table 5: Standard Errors of Estimated Expected Comovements

	$Level_{t+1}$		$Slope_{t+1}$	
$s_{i,t}$	$\hat{E}[V_{i,t+1}\|z_{i,t}]$	$\sigma[\hat{E}[V_{i,t+1}\|z_{i,t}]]$	$\hat{E}[V_{i,t+1}\|z_{i,t}]$	$\sigma[\hat{E}[V_{i,t+1}\|z_{i,t}]]$
1%	3.3	(0.74)	-0.7	(0.37)
5%	4.6	(1.01)	-0.8	(0.51)
50%	9.1	(1.79)	-1.0	(0.94)
95%	20.2	(3.54)	3.2	(1.94)
99%	26.3	(6.68)	11.9	(3.09)
$95\% - 5\%$	15.6	(3.01)	4.0	(1.56)

$\hat{E}\left[V_{i,t+1}\right] = z_{i,t}\hat{b}_v$, where $\hat{b}_v$ is the OLS estimate. The values in parenthesis are standard errors of $\hat{E}\left[V_{i,t+1}\right]$ computed by $\sqrt{z_{i,t}cov\left(\hat{b}_v^{[k]}\right)z_{i,t}'}$ where $z_{i,t}$ are functions of the percentile values of the bond spread. The F-statistics for $\hat{b}_v$ for $Level_{t+1}$ is 16.7 with the p-value of 0.00. The F-statistics for $\hat{b}_v$ for $Slope_{t+1}$ is 11.0 with the p-value of 0.00.

maturity with log to make the relationship with returns closer to be linear. I also demeaned market value-related variables such as $size$ by the cross-sectional average in each month to maintain the stationarity. After such transformations, when I analyze each characteristic separately, a linear function seems to fit the data reasonably well. I also scale each characteristic so that it has a unit standard deviation.

Table 6 shows the estimation result. I run both univariate and multivariate regressions using the characteristics which give rise to statistically significant variations in estimated bond premia when I use them to sort corporate bonds into deciles. Specifically, I use s, cy, $\log size$, DD, $sdtd$, $\log MV_{eq}$ and B/M_{eq}. Furthermore, I include mom_{eq} in this test as my two factor model has a trouble pricing the portfolios sorted on mom_{eq}.

The top panel of Table 6 shows the result of the univariate regressions. Since I am using only the characteristics that are associated with estimated bond premia based on portfolio sorts, the slope coefficient b_r should be statistically significant. The result in the top panel of Table 6 shows that cy, $\log size$ and B/M_{eq} do not have statistically significant slope coefficients possibly due to the linearity assumption.

The bottom panel of Table 6 shows the result of the multivariate regression. When all the characteristics are used at the same time, only five characteristics have economically significant coefficients: s, cy, $\log size$, $\log MV_{eq}$ and mom_{eq}. Thus, in my main empirical result using the parametric characteristic-based approach, I use these five characteristics as instruments.

Table 6: Multivariate Return Forecasting Regression

Characteristics	s	cy	$\log size$	DD	$sdtd$	$\log MV_{eq}$	B/M_{eq}	mom_{eq}
Univariate Regression:								
b_r	0.51	0.02	-0.04	-0.06	-0.03	-0.06	0.01	0.06
$t(b_r)$	(4.18)	(0.61)	(-1.41)	(-3.73)	(-2.25)	(-2.91)	(0.79)	(2.69)
$\sqrt{\frac{1}{NT}\sum_{i,t}(z_{i,t}b_r)^2}$	0.26	0.01	0.03	0.05	0.03	0.06	0.01	0.05
Multivariate Regression:								
b_r	0.52	0.06	-0.14	-0.02	-0.01	0.05	-0.01	0.07
$t(b_r)$	(4.32)	(0.92)	(-2.57)	(-1.07)	(-0.57)	(1.85)	(-0.89)	(3.51)
$\sqrt{\frac{1}{NT}\sum_{i,t}(z_{i,t}b_r)^2}$	0.35					F-statistics:	10.0	[0.000]

Monthly observations from 1973 to 2011. I estimate b_r by regressions $R^e_{i,t+1} = z_{i,t}b_r + \varepsilon^r_{i,t+1}$. $s_{i,t}$ is the bond spread for bond i at time t. To put time fixed effects, I demeaned the vector $z_{i,t}$ by subtracting the time-specific mean $\bar{z}_t \equiv \frac{1}{N_t}\sum_{i=1}^{N_t} z_{i,t}$. Each observation is weighted by the square root of relative value of the bond at time t. The numbers in parenthesis are t-statistics, where standard errors are computed taking into account 12 lags of serial correlations (with Newey and West (1987)'s weighing).

5.5 Practical Issues in Parametric Characteristic-based Test

5.5.1 Attenuating Extreme Observations

In the characteristic-based test using individual bonds, the estimated coefficients b_r and b_v may be sensitive to extreme observations. This is true especially when I include higher-order terms of characteristics (such as squared bond spread) in $z_{i,t}$. Small difference in b_r and $b_v\lambda$ can translate into a large pricing error for a large value of the characteristic.

In the classic asset pricing test using portfolio sorts, a security with an extreme value of characteristics is averaged out with other securities in the extreme portfolio. This is especially the case for the size and book-to-market sorted 25 portfolios of Fama and French (1993) where the threshold of characteristics are set using NYSE stocks only. As a result, the stocks with extreme characteristics (that are typically small and more likely to be NASDAQ stocks) will end up in one or two portfolios that have many securities to average across. This practice is justified for noise-reduction reasons.

To downweight the extreme observations in the characteristic-based asset pricing test, I value-weight the observations. Since the characteristic-based test involves panel data and the market value of bonds grows over time, a naive use of the market value at time t as a weight will downweight old observations too much. Thus I multiply $\left(R^e_{i,t+1}, V_{i,t+1}, z_{i,t}\right)$ with weight $w_{i,t}$ defined by

$$w_{i,t} \equiv \sqrt{\frac{MV_{i,t}}{\sum_{j\in J_t} MV_{j,t}}}$$

where J_t is the set of the bonds that are in sample at time t.

5.5.2 Computing Standard Errors

The three step estimation of λ described in (5) boils down to a one step minimum distance estimation of the moment conditions $E\left[z'_{i,t}\left(R^e_{i,t+1} - V_{i,t+1}\lambda\right)\right] = 0$, with a weighting matrix $E\left[z'_{i,t}z_{i,t}\right]^{-1}$. Thus it is straightforward to compute standard errors using the GMM framework. The panel data I use has a large number of securities that is highly correlated with each other. To account for the correlation across securities, I first compute the standard errors clustered by time and then add several time lags. The exact number of lags added is specified in each test result.

As the cross section is large relative to the time-series, one may be concerned about the reliability of the estimated standard errors. To address this concern I also compute the standard errors imposing some structural assumptions on shock vectors. I apply the structural assumptions suggested by Ang, Liu and Schwartz (2010) and compare the performance of the computed standard errors in Appendix D. Appendix D shows that these alternative methodologies of computing standard errors seem to produce similar results. Thus in the following empirical analysis, I show the standard errors clustered by time (with several lags added).

5.6 Empirical Result

Below I implement the parametric characteristics-based test of the two-factor asset pricing model using various set of characteristics.

5.6.1 All Characteristics but Equity Momentum

First, I use all the characteristics except for mom_{eq} in the set of instruments. That is, I use s, cy, $\log size$ and $\log MV_{eq}$ as my test characteristics. I apply the three step procedure and report the result in Table 7.

The first step regression (in the top panel) shows estimated bond premia as a linear function of the characteristics that I use. The corporate bond with a high spread, a high current yield, a low market value and a high issuer's market value of equity earns higher excess returns going forward. The resulting variation in estimated bond premia $\sqrt{\frac{1}{TN}\sum_{i,t}\left(z_{i,t}\hat{b}_r\right)^2}$ is 38 basis points per month. In comparison, the same statistic for the variation in $E\left(R^e\right)$ of the ten portfolios in Table 1 is 23 basis points. Thus, in the parametric characteristic-based approach, we have economically sizable variations in estimated bond premia which the model needs to match.

The second step regression (in the middle panel) shows the estimates of the expected comovement function. As I have two factors, there are two regressions. The point estimates for b_v show the way in which the estimated risk varies along with the characteristics. The comovements with

$Level_t$ has a large positive loading on cy. Thus, $Level_t$ helps explain the variation in estimated bond premia associated with cy. The comovements with $Slope_t$, on the other hand, has a large positive loading on s and thus helps explain the variation in bond premia along s. Both factors have the loadings on $\log size$ and $\log MV_{eq}$ that have the same sign as the estimated b_r in the first step. That is, both factors help explain the variation in bond premia associated with $\log size$ and $\log MV_{eq}$.

The third step regression is reported in the third panel of Table 7. The estimated λ is 0.01 and 0.10 for $Level_t$ and $Slope_t$, respectively. The estimated λ for $Level_t$ is insignificant while it is highly significant for $Slope_t$. This is consistent with the fact that $Level_t$ does not earn statistically significant average excess returns in my sample while $Slope_t$ earns significant average excess returns (as shown in Table 2).

If the model performs well, (6) should hold. To see this, I report $\hat{b}_v\hat{\lambda}$ in the third panel. As we can see, $\hat{b}_v\hat{\lambda}$ are very close to $\hat{b}_r$, which implies that the model does well in pricing these test assets. In other words, the model can produce the expected comovement functions that closely match the expected excess return function.

The third panel of Table 7 shows another diagnosis of the model. $RMSE_\alpha$ and MAE_α are 0.11% and 0.09% per month, respectively. They are larger than the case of univariate sorts and it is reasonable given that I use all the four characteristics jointly here. The R-squared is as high as 0.93, meaning 93% of the variation in estimated bond premia is matched by the two factor model. The pricing errors are jointly statistically insignificant, as the GMM χ^2 test fails to reject the model at the 10% level.

Multiplying λ with $E[f'f]$ yields the estimated bond premium for each factor. The bond premia for $Level_t$ and $Slope_t$ are 0.37% and 0.71% per month, respectively. $Level_t$ is assigned a smaller premium than its average excess returns (0.49%) and $Slope_t$ is assigned larger premium than its average excess returns (0.52%). The gap between the average excess returns and the estimated bond premium using the parametric characteristic-based approach raises the concern that I may be mispricing the factors to price the test assets.

To alleviate this concern, I also test if the two factors are priced correctly or not, using the point estimates $\hat{\lambda}$. This is a calibration-verification exercise of Hansen (2008) which accounts for the fact that $\hat{\lambda}$ are estimated using the test assets. The test (not in the table) fails to reject the null that alphas for the two factors are jointly zero and thus supports the model's performance.

In the parametric characteristic-based approach, it is important to have large values of F-statistics for the second step regression. As the three step procedure is an application of a standard instrumental variable regression, the usual caution about weak instruments applies. Staiger and Stock (1997) point out that including instruments that weakly forecast endogenous variables ($V_{i,t}$ in my case) causes a problem in the small sample behavior of the estimates. To detect weak instruments, one can compute the statistics of Cragg and Donald (1993), a relative of F-statistics which accounts for the interactions between multiple endogenous variables. In my specific setting, the interaction between $V_{i,t}$ for $Level_t$ and $Slope_t$ does not matter as these two factors are uncorrelated by construction. Thus, the statistics of Cragg and Donald (1993) reduce to the minimum of the two F-statistics computed in the second step.

In Table 7, I have the minimum F-statistics of 15.2. This value is large relative to the cutoff values of a weak instrument test reported in Stock and Yogo (2002). Thus, the set of instruments s, cy, $\log size$ and $\log MV_{eq}$ jointly forecasts comovements well enough.

In conclusion, the two factor model matches well the variation in estimated bond premia associated with s, cy, $\log size$ and $\log MV_{eq}$ jointly.

5.6.2 All Characteristics but Equity Momentum with Interaction

The advantage of the parametric characteristic-based approach over the joint test using univariate sorts (as in Table 4) is that one can allow for interactions among characteristics. If one uses univariate sorts of equities based on size and book-to-market ratio and tests if the intercepts are jointly zero for all portfolios, the 'small growth stocks' effect (Fama and French (1993)) will be missed.

With the parametric characteristic-based approach, one can easily account for interactions by including the interaction terms in the set of instruments and conducting the test in the same manner.

As an example, I include the four characteristics tested in the previous section (s, cy, $\log size$ and $\log MV_{eq}$) and the interaction terms with log time to maturity $\tilde{\tau}$. Though log time to maturity itself does not lead to a significant variation in bond premia, it is interesting to ask if the effect of the characteristics becomes greater for long-term bonds than for short-term bonds.

Table 8 shows the test result. Now I have eight instruments as I add s, cy, $\log size$ and $\log MV_{eq}$ multiplied by $\tilde{\tau}$ to the previous set of instruments. $s \cdot \tilde{\tau}$ and $\log size \cdot \tilde{\tau}$ give rise to a statistically significant variation in estimated bond premia. In the first step, the effect of the bond spread is more pronounced for long-term bonds (b_r for $s \cdot \tilde{\tau}$ is positive) and the effect of size is attenuated for long-term bonds (b_r for $\log size \cdot \tilde{\tau}$ is positive).

The eight characteristics jointly significantly forecast comovements as well, as we can see from the F-statistics in the second step.

The result for the third step regression shows that despite the greater number of instruments, the two factor model has no problem in pricing these test assets. $RMSE_\alpha$ and MAE_α are small at 0.06% and 0.04% per month, respectively. The model cannot be rejected at the 10% level.

Table 7: Asset Pricing Test with Four Characteristics

1st step: $R^e_{i,t+1} = z_{i,t}b_r + \varepsilon^r_{i,t}$

	s	cy	$\log size$	$\log MV_{eq}$	$\sqrt{\frac{1}{NT}\sum(z_{i,t}\hat{b}_r)^2}$	F-stats
$\hat{b}_r$	0.53	0.06	-0.16	0.04	0.38	6.7
$t(\hat{b}_r)$	(4.32)	(0.95)	(-2.38)	(1.63)		[0.000]

2nd step: $V_{i,t+1} = z_{i,t}b_v + \varepsilon^v_{i,t}$

	s	cy	$\log size$	$\log MV_{eq}$	$\sqrt{\frac{1}{NT}\sum(z_{i,t}\hat{b}_v)^2}$	F-stats
$\hat{b}_{v,Level}$	1.05	4.56	-1.05	0.29	2.99	15.8
$t(\hat{b}_{v,Level})$	(0.54)	(4.98)	(-1.25)	(0.92)		[0.000]
$\hat{b}_{v,Slope}$	5.41	-0.39	-0.62	0.17	2.80	15.2
$t(\hat{b}_{v,Slope})$	(6.58)	(-0.74)	(-1.45)	(1.35)		[0.000]

3rd step: $z_{i,t}\hat{b}_r = z_{i,t}\hat{b}_v\lambda + \alpha_{i,t}$

	s	cy	$\log size$	$\log MV_{eq}$
$\hat{b}_v\hat{\lambda}$	0.55	0.01	-0.07	0.02

	Level	*Slope*	$RMSE_\alpha$	MAE_α	R^2_α	χ^2
$\hat{\lambda}$	0.01	0.10	0.11	0.09	0.93	2.04
$t(\hat{\lambda})$	1.10	4.70				[0.360]
Premium	0.37	0.71				

Monthly observations from 1973 to 2011. I estimate b_r, b_v and λ by three regressions $R^e_{i,t+1} = z_{i,t}b_r + \varepsilon^r_{i,t+1}$, $V_{i,t+1} = z_{i,t}b_v + \varepsilon^v_{i,t+1}$ and $z_{i,t}\hat{b}_r = z_{i,t}\hat{b}_v\lambda + \alpha_{i,t}$. To put time fixed effects, I demeaned the vector $z_{i,t}$ by subtracting the time-specific mean $\bar{z}_t \equiv \frac{1}{N_t}\sum_{i=1}^{N_t} z_{i,t}$. Each observation is multiplied by the square root of relative value of the bond at time t. The numbers in parenthesis are t-statistics, where standard errors are computed taking into account 1,12 and 12 lags of serial correlations (with Newey and West (1987)'s weighing) for λ, b_r and b_v, respectively. Premium is computed by $E\left[f'_{t+1}f_{t+1}\right]\lambda$ and expressed in percentage per month. χ^2 is chi-squared statistics of Hansen (1982). The numbers in bracket are p-values. $RMSE_\alpha \equiv \sqrt{\frac{1}{NT}\sum_{i,t}\alpha^2_{i,t}}$ and $MAE_\alpha \equiv \frac{1}{NT}\sum_{i,t}|\alpha_{i,t}|$. R^2_α is R-squared defined by $1 - \frac{\sum\alpha^2_{it}}{\sum(z_{it}b_r)^2}$.

Table 8: Asset Pricing Test with Four Characteristics and Interactions

1st step: $R^e_{i,t+1} = z_{i,t} b_r + \varepsilon^r_{i,t}$

	s	cy	$\log size$	$\log MV_{eq}$	$s \cdot \tilde{\tau}$	$cy \cdot \tilde{\tau}$	$\log size \cdot \tilde{\tau}$	$\log MV_{eq} \cdot \tilde{\tau}$	$\sqrt{\frac{1}{NT}\sum(z_{i,t}\hat{b}_r)^2}$	F-stats
$\hat{b}_r$	0.28	0.02	-0.33	0.18	0.25	-0.01	0.07	-0.04	0.37	4.6
$t(\hat{b}_r)$	(2.39)	(0.30)	(-3.38)	(2.59)	(3.28)	(-0.65)	(2.40)	(-1.78)		[0.000]

2nd step: $V_{i,t+1} = z_{i,t} b_v + \varepsilon^v_{i,t}$

	s	cy	$\log size$	$\log MV_{eq}$	$s \cdot \tilde{\tau}$	$cy \cdot \tilde{\tau}$	$\log size \cdot \tilde{\tau}$	$\log MV_{eq} \cdot \tilde{\tau}$	$\sqrt{\frac{1}{NT}\sum(z_{i,t}\hat{b}_v)^2}$	F-stats
$\hat{b}_{v,Level}$	3.42	-2.41	-1.00	1.38	2.06	1.80	0.45	-0.22	5.16	11.6
$t(\hat{b}_{v,Level})$	(2.93)	(-3.22)	(-0.79)	(1.95)	(2.28)	(7.59)	(1.25)	(-0.78)		[0.000]
$\hat{b}_{v,Slope}$	2.11	0.59	-1.19	0.59	2.24	-0.47	0.23	-0.11	2.80	9.7
$t(\hat{b}_{v,Slope})$	(2.77)	(1.70)	(-1.98)	(1.59)	(6.57)	(-4.00)	(1.30)	(-0.91)		[0.000]

3rd step: $z_{i,t}\hat{b}_r = z_{i,t}\hat{b}_v \lambda + \alpha_{i,t}$

	s	cy	$\log size$	$\log MV_{eq}$	$s \cdot \tilde{\tau}$	$cy \cdot \tilde{\tau}$	$\log size \cdot \tilde{\tau}$	$\log MV_{eq} \cdot \tilde{\tau}$
$\hat{b}_v\hat{\lambda}$	0.25	0.02	-0.13	0.08	0.25	-0.02	0.03	-0.01

	$Level$	$Slope$		$RMSE_\alpha$	MAE_α	R^2_α	χ^2
$\hat{\lambda}$	0.02	0.09		0.06	0.04	0.97	10.52
$t(\hat{\lambda})$	0.42	4.64					[0.105]
Premium	0.56	0.67					

Monthly observations from 1973 to 2011. I estimate b_r, b_v and λ by three regressions $R^e_{i,t+1} = z_{i,t} b_r + \varepsilon^r_{i,t+1}$, $V_{i,t+1} = z_{i,t} b_v + \varepsilon^v_{i,t+1}$ and $z_{i,t}\hat{b}_r = z_{i,t}\hat{b}_v\lambda + \alpha_{i,t}$. $\tilde{\tau}$ is log time to maturity. To put time fixed effects, I demeaned the vector $z_{i,t}$ by subtracting the time-specific mean $\bar{z}_t \equiv \frac{1}{N_t}\sum_{i=1}^{N_t} z_{i,t}$. Each observation is multiplied by the square root of relative value of the bond at time t. The numbers in parenthesis are t-statistics, where standard errors are computed taking into account 1,12 and 12 lags of serial correlations (with Newey and West (1987)'s weighing) for λ, b_r and b_v, respectively. Premium is computed by $E\left[f'_{t+1} f_{t+1}\right]\lambda$ and expressed in percentage per month. χ^2 is chi-squared statistics of Hansen (1982). The numbers in bracket are p-values. $RMSE_\alpha \equiv \sqrt{\frac{1}{NT}\sum_{i,t}\alpha^2_{i,t}}$ and $MAE_\alpha \equiv \frac{1}{NT}\sum_{i,t}|\alpha_{i,t}|$. R^2_α is R-squared defined by $1 - \frac{\sum \alpha^2_{it}}{\sum (z_{it} b_r)^2}$.

5.6.3 Including Equity Momentum

Finally, I include mom_{eq} in the set of instruments and test if the model can match the variation in estimated bond premia associated with the five characteristics. The test result is shown in Table 9.

The first step regression shows that $\hat{b}_r$ for mom_{eq} is positive and statistically significant in presence of other characteristics. That is, the bonds issued by past equity winners tend to earn high bond returns going forward.

The second step regression shows that the model works in the opposite direction. Both $Level_t$ and $Slope_t$ have negative loadings $\hat{b}_v$ on mom_{eq}. That is, from the model's perspective, the bonds issued by past winners should be less risky. As a result, the model implied expected excess return function has a wrong sign for mom_{eq}. Table 9 shows $\hat{b}_v\hat{\lambda}$ for mom_{eq} is -0.03 while it is 0.07 for $\hat{b}_r$.

As I include mom_{eq} in the set of instruments, $RMSE_\alpha$ and MAE_α become large at 0.18% and 0.15% per month, respectively. The alphas are jointly statistically significant and the GMM χ^2 test rejects the model at the 1% level. In conclusion, the two factor model cannot price these test assets as the pricing errors are both economically and statistically significant.

6 Conclusion

I show that the simple two factor model consisting of the first two principal components of the 10 portfolios sorted on bond spreads prices much of the variation in corporate bond premia. When the asset pricing tests are done with univariate sorts, the model performs well regardless of the choices of sorting variables except equity momentum.

I introduce the parametric characteristic-based asset pricing test, and show its flexibility of incorporating multiple characteristics into the test. It can describe expected excess returns and covariances of returns with factors as a function of characteristics and check if these functions are equal. By taking a stand on the functional form of the expected comovement function, I can estimate risk for individual securities precisely without sorting securities into bins.

The two factor model consisting of $Level_t$ and $Slope_t$ cannot be rejected using the four characteristics that best forecast excess returns as instruments, providing a further evidence for the model. The model can price the variation in bond premia associated with the characteristics other than the bond spread used to form the factors. However, the model is rejected when it is tested using the five instruments including equity momentum.

Table 9: Asset Pricing Test Including Equity Momentum

1st step: $R^e_{i,t+1} = z_{i,t} b_r + \varepsilon^r_{i,t}$

	s	cy	$\log size$	$\log MV_{eq}$	mom_{eq}	$\sqrt{\frac{1}{NT}\sum(z_{i,t}\hat{b}_r)^2}$	F-stats
$\hat{b}_r$	0.54	0.06	-0.17	0.05	0.07	0.40	12.6
$t(\hat{b}_r)$	(4.53)	(0.91)	(-2.57)	(1.88)	(3.47)		[0.000]

2nd step: $V_{i,t+1} = z_{i,t} b_v + \varepsilon^v_{i,t}$

	s	cy	$\log size$	$\log MV_{eq}$	mom_{eq}	$\sqrt{\frac{1}{NT}\sum(z_{i,t}\hat{b}_v)^2}$	F-stats
$\hat{b}_{v,Level}$	0.96	4.55	-0.97	0.27	-0.38	2.99	12.8
$t(\hat{b}_{v,Level})$	(0.50)	(4.94)	(-1.16)	(0.83)	(-1.38)		[0.000]
$\hat{b}_{v,Slope}$	5.37	-0.40	-0.57	0.15	-0.23	2.79	12.2
$t(\hat{b}_{v,Slope})$	(6.62)	(-0.76)	(-1.33)	(1.29)	(-1.78)		[0.000]

3rd step: $z_{i,t}\hat{b}_r = z_{i,t}\hat{b}_v \lambda + \alpha_{i,t}$

	s	cy	$\log size$	$\log MV_{eq}$	mom_{eq}
$\hat{b}_v\hat{\lambda}$	0.52	0.00	-0.06	0.02	-0.03

	Level	*Slope*	$RMSE_\alpha$	MAE_α	R^2_α	χ^2
$\hat{\lambda}$	0.01	0.10	0.18	0.15	0.80	19.54
$t(\hat{\lambda})$	0.93	4.45				[0.000]
Premium	0.32	0.68				

Monthly observations from 1973 to 2011. I estimate b_r, b_v and λ by three regressions $R^e_{i,t+1} = z_{i,t}b_r + \varepsilon^r_{i,t+1}$, $V_{i,t+1} = z_{i,t}b_v + \varepsilon^v_{i,t+1}$ and $z_{i,t}\hat{b}_r = z_{i,t}\hat{b}_v\lambda + \alpha_{i,t}$. To put time fixed effects, I demeaned the vector $z_{i,t}$ by subtracting the time-specific mean $\bar{z}_t \equiv \frac{1}{N_t}\sum_{i=1}^{N_t} z_{i,t}$. Each observation is multiplied by the square root of relative value of the bond at time t. The numbers in parenthesis are t-statistics, where standard errors are computed taking into account 1,12 and 12 lags of serial correlations (with Newey and West (1987)'s weighing) for λ, b_r and b_v, respectively. Premium is computed by $E\left[f'_{t+1}f_{t+1}\right]\lambda$ and expressed in percentage per month. χ^2 is chi-squared statistics of Hansen (1982). The numbers in bracket are p-values. $RMSE_\alpha \equiv \sqrt{\frac{1}{NT}\sum_{i,t}\alpha^2_{i,t}}$ and $MAE_\alpha \equiv \frac{1}{NT}\sum_{i,t}|\alpha_{i,t}|$. R^2_α is R-squared defined by $1 - \frac{\sum\alpha^2_{it}}{\sum(z_{it}b_r)^2}$.

References

[1] Adrian, Tobias, Richard K. Crump and Emanuel Moench, 2011, Efficient, Regression-Based Estimation of Dynamic Asset Pricing Models, *Working Paper*

[2] Ang, Andrew, Jun Liu and Krista Schwarz, 2010, Using Stocks or Portfolios in Tests of Factor Models, *Working Paper*

[3] Avramov, Doron and Tarun Chordia, 2006, Asset Pricing Models and Financial Market Anomalies, *Review of Financial Studies* 19, 3, 1001-1040

[4] Bao, Jack, Jun Pan and Jiang Wang, 2011, The Illiquidity of Corporate Bonds, *Journal of Finance* 66, 911-946

[5] Bhamra, Harjoat S., Lars-Alexander Kuehn and Ilya A. Strebulaev 2010, The Levered Equity Risk Premium and Credit Spreads: A Unified Framework, *Review of Financial Studies* 23, 2, 645-703

[6] Chen, Long, David A Lesmond and Jason Wei, 2007, Corporate Yield Spread and Bond Liquidity, *Journal of Finance* 62, 1, 119-149

[7] Chen, Long, Pierre Collin-Dufresne and Robert S. Goldstein, 2009, On the Relation Between the Credit Spread and the Equity Premium Puzzle, *Review of Financial Studies* 22 (9), 3367-3409

[8] Cochrane, John H., 1996, A Cross-Sectional Test of an Investment-Based Asset Pricing Model, *Journal of Political Economy* 104, 572-621

[9] Cochrane, John H., 2011, Discount Rates, *Journal of Finance* 66, 1047-1108

[10] Collin-Dufresne, Pierre, Robert S. Goldstein and J. Spencer Martin, 2001, The Determinants of Credit Spread Changes, *Journal of Finance* 56, 6, 2177-2207

[11] Connor, Gregory, Matthias Hagmann and Oliver Linton, 2012, Efficient Semiparametric Estimation of The Fama–French Model and Extensions, *Econometrica* 80, 2, 713-754

[12] Crabbe, Leland E., 1991, Callable corporate bonds: A vanishing breed, *FEDS working paper #155*, Board of Governors of the Federal Reserve System

[13] Crabbe, Leland E. and Jean Helwege, 1994, Alternative Tests of Agency Theories of Callable Corporate Bonds, *Financial Management* 23, 4, 3-20

[14] Cragg, John G. and Stephen G. Donald, 1993, Testing Identifiability and Specification in Instrumental Variable Models, *Econometric Theory* 9, 2, 222-240

[15] Driessen, Joost, 2005, Is Default Event Risk Priced in Corporate Bonds?, *Review of Financial Studies* 18 (1), 165-195

[16] Fama, Eugene F. and Kenneth R. French, 1993, Common Risk Factors in the Returns on Stocks and Bonds, *Journal of Financial Economics* 33, 23-49

[17] Fama, Eugene F. and James D. MacBeth, 1973, Risk, Return and Equilibrium: Empirical Tests, *Journal of Political Economy* 81, 3, 607-636

[18] Ferson, Wayne E. and Campbell R. Harvey, 1991, The Variation of Economic Risk Premiums, *Journal of Political Economy* 99, 2, 385-415

[19] Gagliardini, Patrick, Elisa Ossola and Olivier Scaillet, 2011, Time-varying risk premium in large cross-sectional equity datasets, *Working Paper*

[20] Gao, Pengqin, 2010, Firm Characteristics, Covariances and Cross-Sectional Expected Returns, *Working Paper*

[21] Gebhardt, William R., Soeren Hvidkjaer and Bhaskaran Swaminathan, 2005, The cross-section of expected corporate bond returns: Betas or characteristics?, *Journal of Financial Economics* 75, 85-114

[22] Gebhardt, William R., Soeren Hvidkjaer and Bhaskaran Swaminathan, 2005b, Stock and bond market interaction: Does momentum spill over?, *Journal of Financial Economics* 75, 651-690

[23] Gilchrist, Simon, and Egon Zakrajsek, 2012, Credit Spreads and Business Cycle Fluctuations, *American Economic Review* 102, 4, 1692-1720

[24] Hansen, Lars P., 2008, Generalized Method of Moments Estimation, *New Palgrave Dictionary of Economics*, 2nd edition

[25] Rosenberg, Barr, 1974, Extra-Market Components of Covariance Among Security Prices, *Journal of Financial and Quantitative Analysis* 9, 263-274

[26] He, Zhiguo and Wei Xiong, 2012, Rollover Risk and Credit Risk, *Journal of Finance* 67, 391-429

[27] Jostova, Gergana, Stanislava Nikolova, Alexander Philipov and Christof W. Stahel, 2010, Momentum in Corporate Bond Returns, *Working Paper*

[28] Merton, Robert C., 1973, Theory of Rational Option Pricing, *Bell Journal of Economics and Management Science* 4, 1, 141-183

[29] Nagel, Stefan and Kenneth J. Singleton, 2011, Estimation and Evaluation of Conditional Asset Pricing Models, *Journal of Finance*, Forthcoming

[30] Newey, Whitney K. and West, Kenneth D. (1987), A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica 55*, 703-708

[31] Nozawa, Yoshio (2012), Term Structure Model of Portfolios of Corporate Bonds, *Working Paper*

[32] Petersen, Mitchell A., 2008, Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches, *Review of Financial Studies* 22, 435-480

[33] Ross, Stephan, 1976, The arbitrage theory of capital asset pricing, *Journal of Economic Theory* 13, 3, 341–360

[34] Staiger, Douglas and James H. Stock, 1997, Instrumental Variables Regression with Weak Instruments, *Econometrica* 65, 3, 557-586

[35] Stock, James H. and Motohiro Yogo, 2002, Testing For Weak Instruments In Linear IV Regression, *Working Paper*

[36] Warga, Arthur and Ivo Welch, 1993, Bondholder Losses in Leveraged Buyouts, *Review of Financial Studies* 6, 959-982

A Comparing Overlapping Data Sources

Table 10 compares the mean and percentile differences between the observations used in my analysis and the overlapping observations in alternative databases. The distribution of the difference in prices is skewed so that the mean is significantly greater than the median. Though the median difference is small, some extreme observations exist (shown in 90 percentile values). This is because the database in comparison has a unreasonable price probably due to recording errors. Using the data in comparison instead of the one I use in the main analysis will not change my main result, as when I apply the filters described in Section 3, these unreasonable prices will be completely eliminated.

The findings of Jostova, Nikolova, Phikipov and Stahel (2010) also support my conclusion. They use similar data as mine and prioritize the data source in the same way as I did. They show that whether using only the data with the highest priority and using the average across overlapping observations do not affect the resulting returns.

B Estimation of Distance to Default

To create a characteristic which measures the probability of default, I will use Merton's model. The value of the assets of a firm A_t follows a geometric Brownian motion.

Table 10: Comparing the Overlapping Observations Across Databases

RefDat	Compare	Mean	10ptle	50ptle	90ptle	n
Lehman		1.807	0.125	0.935	4.470	12,929
Lehman	Mergent	1.674	0.113	0.853	4.191	5,504
Lehman	DataStream	1.895	0.125	1.000	4.583	7,425
Trace		1.322	0.010	0.521	3.304	107,770
Trace	Mergent	1.356	0.000	0.477	3.375	55,861
Trace	DataStream	1.293	0.062	0.613	3.250	51,909
Mergent	DataStream	1.646	0.047	0.552	3.678	23,815

The mean and percentiles of price difference (per 100 dollars) are reported. RefDat is the reference data used in the analysis and Compare is the data presented here for comparison. 10ptle, 50ptle and 90ptle are 10, 50 and 90 percentile values respectively. n is the number of observations.

$$\frac{dA_t}{A_t} = \mu dt + \sigma_A dW_t \tag{8}$$

Let D_t be the book value of the debt of the firm at time t. If the value of the firm's asset is less than the book value of debt at the maturity date, then it cannot repay the debt and defaults. When in default, the bond holders take over the firm immediately and the equity holders receive zero. If the assets exceed the debt, then the equity holder receives the difference between A_t and D_t. This way, the market value of equity S_t can be considered as the price of a call option. The equity value is given by Black-Scholes formula for a call option

$$S_t = A_t\Phi(d_1) - D_t\Phi(d_2) \tag{9}$$

where

$$d_1 = \frac{\log(A_t/D_t) + \left(r + \frac{1}{2}\sigma_A^2\right)}{\sigma_A}$$
$$d_2 = \frac{\log(A_t/D_t) + \left(r - \frac{1}{2}\sigma_A^2\right)}{\sigma_A}$$

and r is a risk-free rate and Φ is a cumulative density function of a standard normal distribution.

We cannot directly observe the market value of the asset A_t and its volatility σ_A. Instead, we can observe the market value of equity S_t and its volatility σ_S. By applying Ito's lemma to the equity and imposing no-arbitrage condition, we have the risk neutral dynamics of equity S_t:

$$dS_t = rS_t dt + \frac{\partial S_t}{\partial A_t} A_t \sigma_A dW_t$$

By matching coefficient, we obtain

$$\begin{aligned}\sigma_S =& \frac{\partial S_t}{\partial A_t}\frac{A_t}{S_t}\sigma_A \\ =& \Phi(d_1)\frac{A_t}{S_t}\sigma_A\end{aligned} \tag{10}$$

Equations (9) and (10) give a system of two equations with two unknowns: A_t and σ_A. Since they are non-linear equations, I solve them numerically using KNITRO solver.

C Non-parametric Estimates of the expected excess return function

To investigate an appropriate functional form, I provide graphical investigations based on non-parametric kernel regression of the form

$$R^e_{i,t+1} = f\left(z_{i,t}\right) + \varepsilon_{i,t+1}$$

where the function $f\left(z_{i,t}\right)$ is estimated non-parametrically using Gaussian kernel. Specifically, the kernel regression estimator at z_0 is given by

$$f_t\left(z_0\right) = \frac{\frac{1}{Nh}\sum_{i=1}^{N} K\left(\frac{z_i - z_0}{h}\right) R^e_{i,t+1}}{\frac{1}{Nh}\sum_{i=1}^{N} K\left(\frac{z_i - z_0}{h}\right)} \equiv \sum_{i=1}^{N} w_{i0,h} R^e_{i,t+1}$$

where $K\left(\cdot\right)$ is the normal probability density function and h is the bandwidth of the kernel chosen to minimize the cross-validation $CV\left(h\right)$,

$$\begin{aligned} CV\left(h\right) &\equiv \sum_{i=1}^{N}\left(R^e_{i,t+1} - \hat{f}_{-i}\left(z_i\right)\right)^2 \pi\left(z_i\right) \\ \hat{f}_{-i}\left(z_i\right) &= \frac{\sum_{j\neq i}^{N} w_{ji,h} R^e_{j,t+1}}{\sum_{j\neq i}^{N} w_{ji,h}} \end{aligned}$$

and $\pi\left(\cdot\right)$ is an indicator function which is one if z_i is between 5 percentile and 95 percentile of the distribution and zero otherwise. The idea of cross-validation procedure is that one has to strike a balance between minimizing errors and improving efficiency of an estimate. Putting a large weight on $R^e_{i,t+1}$ to estimate $f\left(z_i\right)$ will reduce an error but will not use as much information from the adjacent observations $R^e_{j,t+1}$, which results in inefficiency. I penalize inefficiency by using $\hat{f}_{-i}\left(z_i\right)$, or a leave one-out estimates in computing $CV\left(h\right)$. Minimizing $CV\left(h\right)$ amounts to choose an optimal balance between error minimization and efficiency of the estimates. I estimate the function

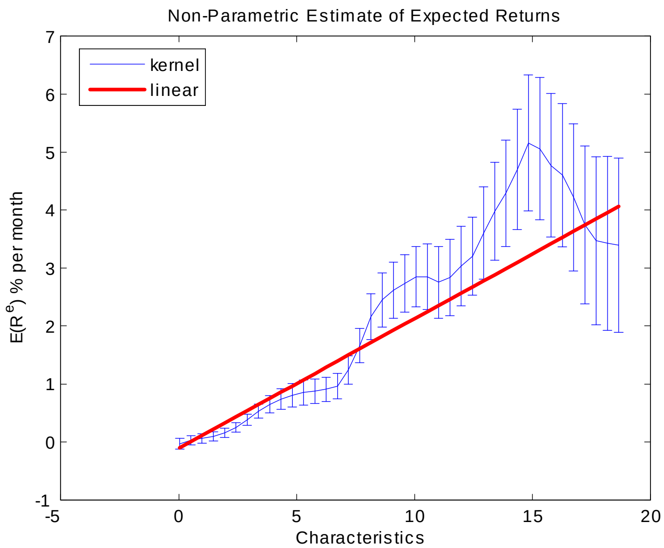

FIGURE 4: BOND SPREAD

$f_t(z)$ for each month and take average of them. The standard errors of the estimated expected excess returns are computed using the time variation of $f_t(z)$ accounting for 12 lags. Figures 4 to 17 show the estimated expected excess return function with error bars for each characteristic. I also plot the fitted line of linear regressions on z_t. The result is robust to the choice of the kernel function and Epanechnikov (quadratic) kernel yields a similar result.

From the figures, we can see that for most of the characteristics, a linear approximation provides a decent fit. Though statistical tests for the fit of parametric estimates to the kernel regression estimates are available, I do not pursue them here as I run a multivariate regression (3) instead of a univariate regression in the asset pricing test. Thus, the estimated kernel regression provides a limited guidance about how to choose the functional form, requiring a researcher to try multiple functional form for robustness of the test result.

D Comparing the performance of the standard errors

To compute standard errors, I need to compute

$$S \equiv E\left[z'_{i,t}\varepsilon_{i,t+1}\varepsilon'_{i,t+1}z_{i,t}\right]$$

where $\varepsilon_{i,t} \in \left\{\varepsilon^r_{i,t}, \varepsilon^v_{i,t}, \varepsilon^r_{i,t} - \varepsilon^v_{i,t}\lambda_0\right\}$. If I cluster the standard errors by time, the estimates of S is given by

$$\hat{S} = \frac{1}{NT}\sum_{t=1}^{T} Z'_t\Omega_t Z_t$$

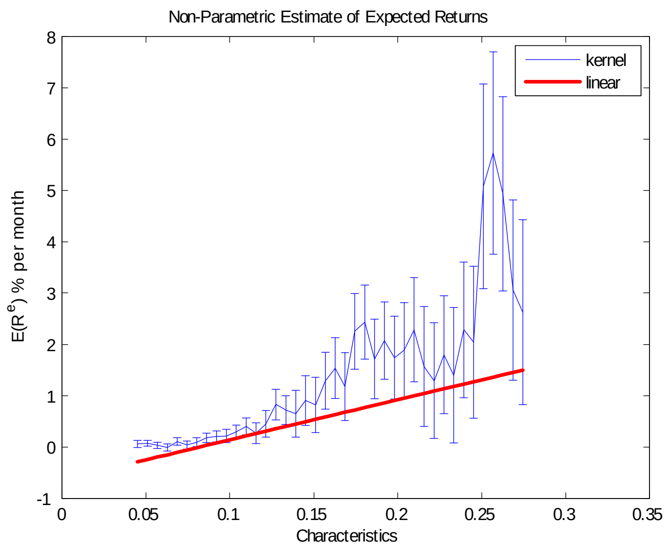

Figure 5: Current Yield

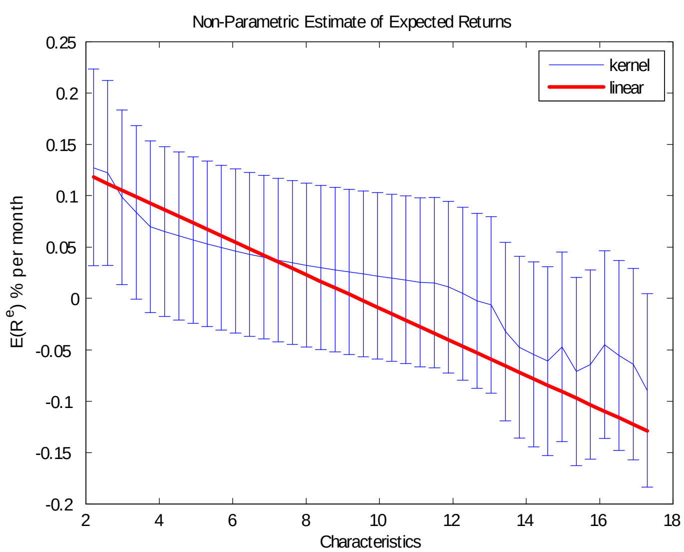

Figure 6: Distance to Default

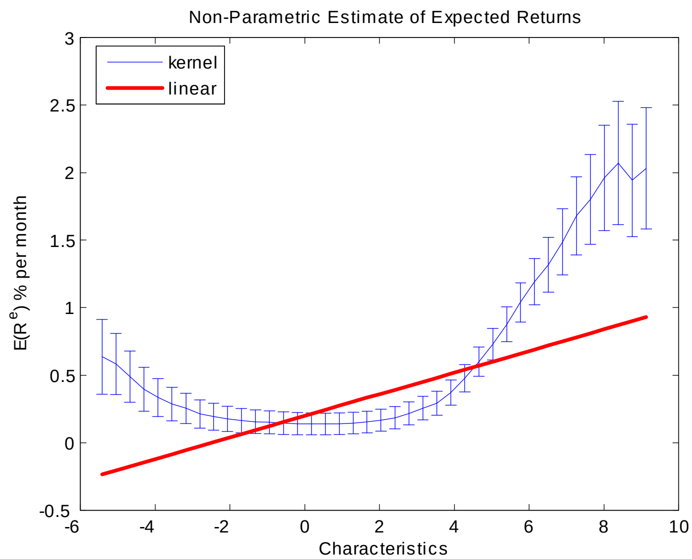

Figure 7: Momentum (-12 to -2 month returns)

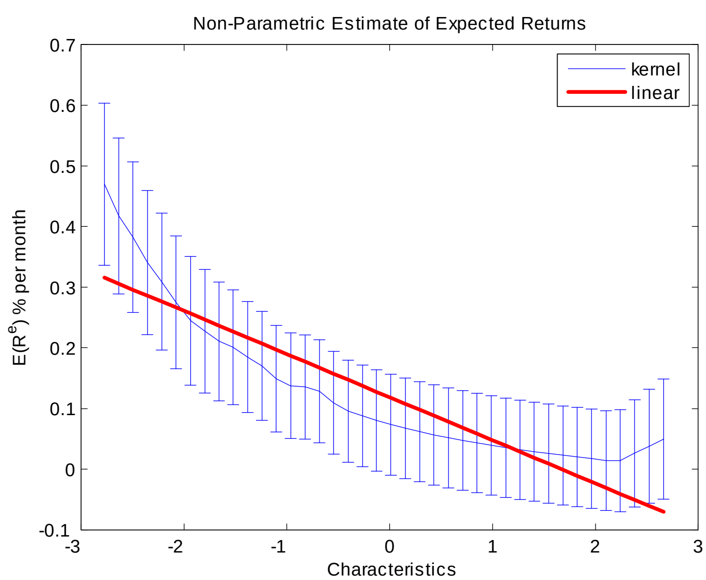

Figure 8: Log Size

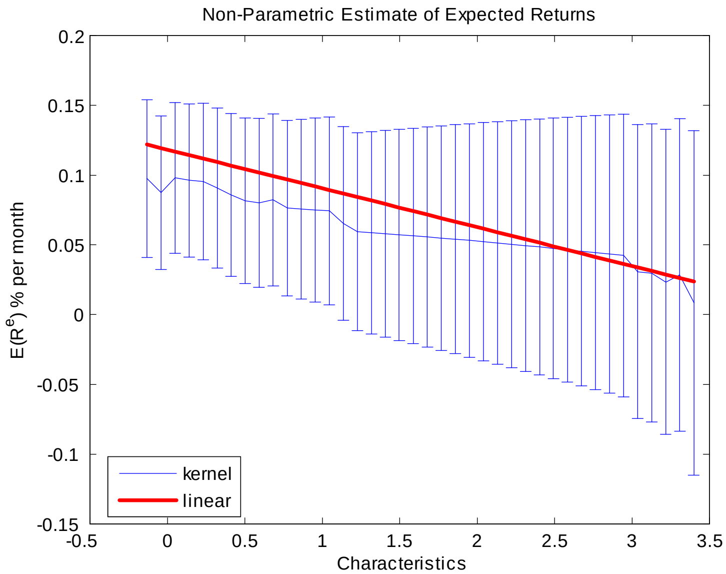

Figure 9: Log Time to Maturity

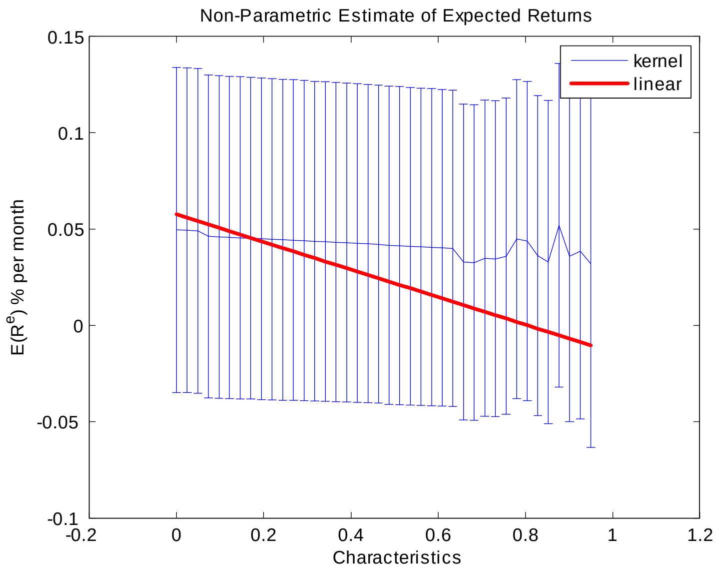

Figure 10: Short-term Debt/Total Debt Ratio

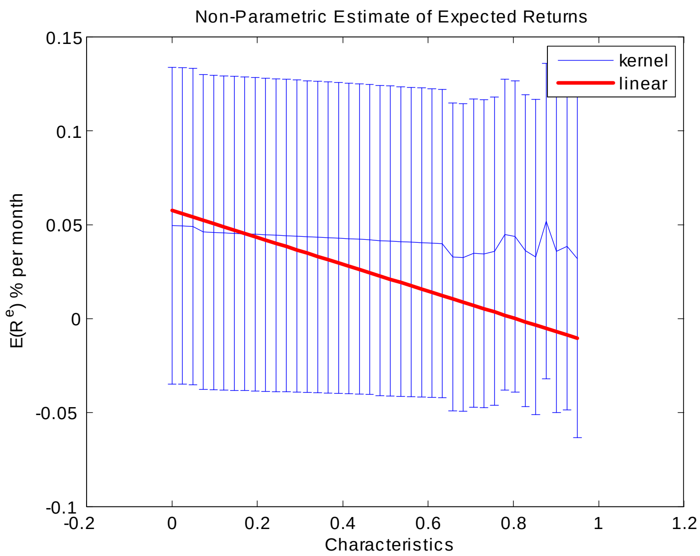

Figure 11: Tangibility (Property, Plant and Equipment/Total Asset)

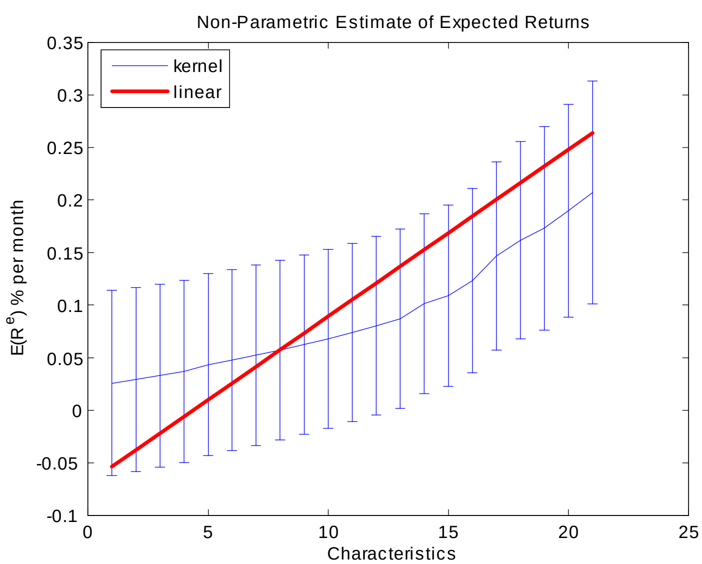

Figure 12: Credit Ratings (AAA as 1, C as 21)

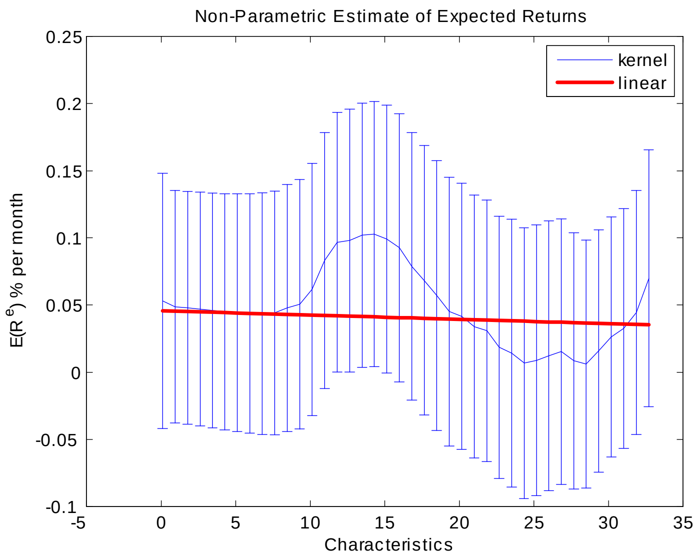

FIGURE 13: AGE

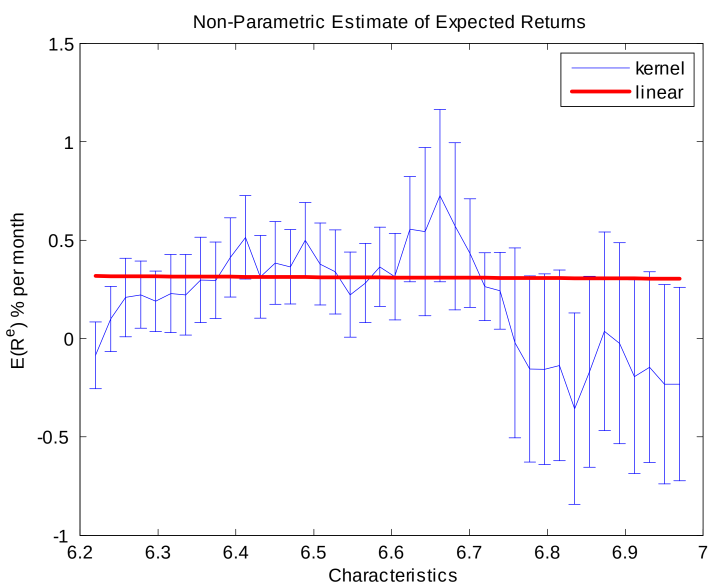

FIGURE 14: YIELD TO MATURITY ON MATCHING TREASURY BONDS

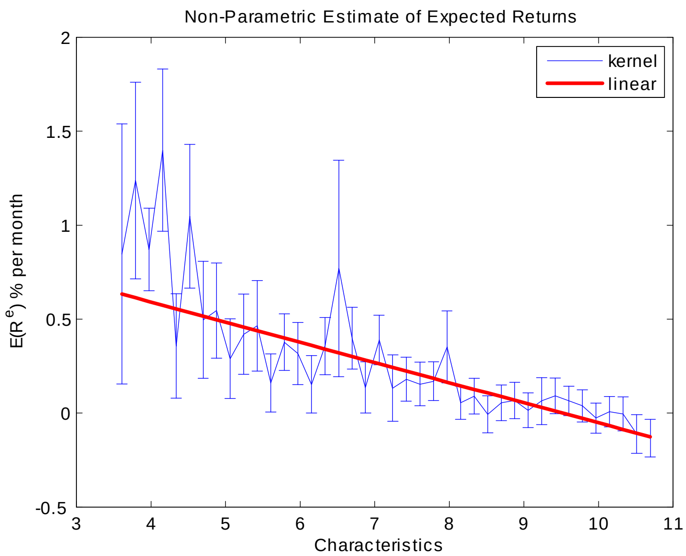

FIGURE 15: EQUITY SIZE

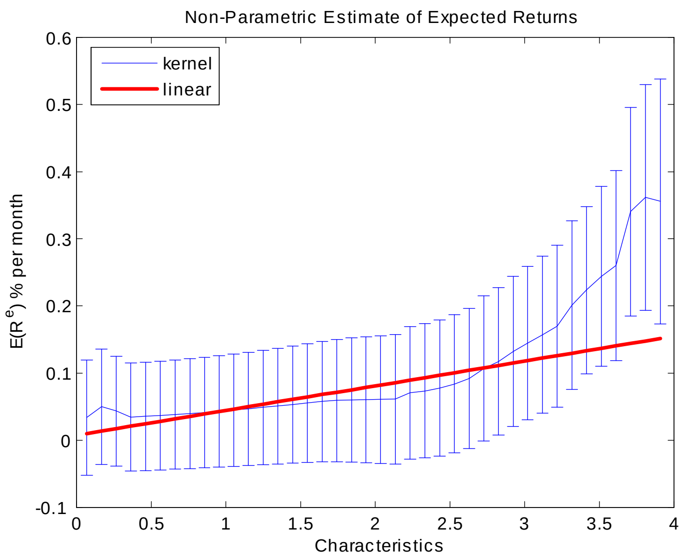

FIGURE 16: EQUITY BOOK-TO-MARKET RATIO

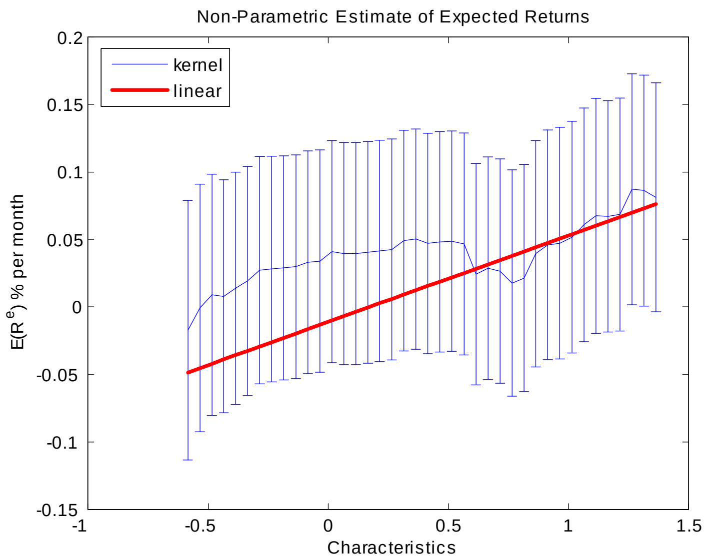

FIGURE 17: EQUITY MOMENTUM (-12 TO -2 EQUITY RETURNS)

where Z_t is $N_t - by - L$ data matrix of instruments and Ω_t is the sample covariance matrix of error term $\varepsilon_{i,t+1}$. Since the number of bonds at time t, N_t, is large relative to length of time ($N_t >> T$), Ω_t may not be reliable estimates of true conditional covariance matrix $E\left[\varepsilon_{i,t+1}\varepsilon'_{i,t+1}|z_{i,t}\right]$. To overcome this issue, I impose a structure on the data-generating process of the errors to obtain reliable estimates. Specifically, as suggested in Ang, Liu and Schwartz (2010), I compute standard errors using two methods. First, I impose one-factor structure in error terms. Second, I assume the deciles defined by distance-to-default drive comovements among error terms[12].

In the first methodology of computing standard errors, I assume that the error terms have the one-factor structure

$$\varepsilon_{i,t+1} = \xi_i u_{t+1} + v_{i,t+1}$$

where u_{t+1} and $v_{i,t+1}$ is iid. Let σ_u^2 denote the variance of u_{t+1} and Σ_v denote a diagonal matrix whose diagonal element is the variance of $v_{i,t+1}$. Then

$$\Omega_t = \xi_t \xi'_t \sigma_u^2 + \Sigma_v$$

where ξ_t is a $N_t - by - 1$ vector which stacks ξ_i for all i that exist at time t.

Empirically, the shock u_{t+1} can be obtained as average errors

$$\tilde{u}_{t+1} = \frac{1}{N_t}\sum_{i=1}^{N_t} \varepsilon_{i,t+1}$$

[12] Instead of using the characteristic-based deciles, Ang, Liu and Schwartz (2010) use ten industries to categolize securities.

and the sensitivity parameter ξ is obtained by regressing $\varepsilon_{i,t+1}$ onto u_{t+1}. That is, regress

$$\varepsilon_{i,t+1} = \gamma + \xi_i \tilde{u}_{t+1} + v_{i,t+1}$$

and obtain the OLS estimates $\hat{\xi}_i$ and $\hat{v}_{i,t+1}$.

In the second methodology, I assume that the bond i in decile D has the shock following

$$\varepsilon_{i,t+1} = \zeta_i u_{D,t+1} + \nu_{i,t+1}$$

where ζ_i is $1-by-10$ indicator vector whose D-th entry is 1 and 0 otherwise. $u_{D,t+1}$ is $10-by-1$ vector of decile-wide shocks with covariance matrix Σ_u and $\nu_{i,t+1}$ is an idiosyncratic shock that is independent across securities and time. Then we have

$$\Omega_t = \zeta_t \Sigma_u \zeta_t' + \Sigma_\nu$$

where ζ_t is $N_t - by - 10$ matrix constructed by stacking ζ_i and Σ_ν is a diagonal matrix whose diagonal entry is the variance of $\nu_{i,t+1}$. In the article, I use distance-to-default as a characteristics that defines the deciles. At each time t, I classify bonds that exist at time t into deciles depending on their value of distance-to-defaults. The decile-wide shock $u_{D,t}$ is constructed by the sample average

$$\tilde{u}_{D,t} = \frac{1}{N_{t,D}} \sum_{i \in D} \varepsilon_{i,t}$$

where $N_{i,D}$ is the number of securities in decile D at time t. Σ_u is estimated using the sample covariance matrix of $\tilde{u}_{D,t}$. The idiosyncratic shock $\nu_{i,t}$ is constructed by

$$\tilde{\nu}_{i,t} = \varepsilon_{i,t} - \zeta_i \tilde{u}_{D,t}$$

and the sample variance of $\tilde{\nu}_{i,t}$ is used to estimate the diagonal entries of Σ_ν.

Collin-Dufresne, Goldstein and Martin (2001) show that the residuals of the bond price change after accounting for fundamentals have a one factor structure. Thus, the one-factor error model has some empirical support. On the other hand, the distance-to-default decile model is logically more consistent with the objective of this article in the sense that all the moments of the model should be estimated as a function of characteristics, not the name of the individual bond.

I show the result of the characteristic-based asset pricing test of the two factor model using three different measures of standard errors. Here I use the set of instruments that best forecast returns, as in Table 7. Since it is time consuming to account for time lags in the methodologies of Ang, Liu and Schwartz (2010), I compute standard errors with no lags for all the three methodologies.

Table 11 shows the estimated standard errors and χ^2 statistics using the three alternative assumptions. I show the result with the same instruments as Table 7 as an example but other

Table 11: Comparing Alternative Models of Standard Errors

	Point Estimates				Cluster by time t-statistics			One factor error model t-statistics			DD-decile error model t-statistics		
	b_r	$b_{v,Level}$	$b_{v,Slope}$	$b_v\lambda$	b_r	$b_{v,Level}$	$b_{v,Slope}$	b_r	$b_{v,Level}$	$b_{v,Slope}$	b_r	$b_{v,Level}$	$b_{v,Slope}$
s	0.53	1.05	5.41	0.55	(3.89)	(0.73)	(7.17)	(4.78)	(0.81)	(6.61)	(3.38)	(0.84)	(8.64)
cy	0.06	4.56	-0.39	0.01	(0.80)	(5.98)	(-1.04)	(0.87)	(8.08)	(-1.04)	(1.55)	(15.44)	(-2.38)
$\log size$	-0.16	-1.05	-0.62	-0.07	(-1.94)	(-1.35)	(-1.60)	(-2.41)	(-1.78)	(-1.87)	(-3.14)	(-2.78)	(-3.08)
$\log MV_{eq}$	0.04	0.29	0.17	0.02	(1.35)	(1.10)	(1.32)	(1.81)	(1.42)	(1.34)	(1.54)	(1.42)	(1.58)
		$Level_t$	$Slope_t$			$Level_t$	$Slope_t$		$Level_t$	$Slope_t$		$Level_t$	$Slope_t$
λ		0.01	0.10			(1.07)	(4.65)		(1.02)	(6.32)		(2.27)	(3.55)
$\chi^2[pv]$						1.81	[0.404]		2.12	[0.346]		4.02	[0.134]

Monthly observations from 1973 to 2011. Each observation is multiplied by the square root of relative value of the bond at time t. The numbers in parenthesis are t-statistics, where standard errors are computed taking into account 0, 0 and 0 lags of serial correlations for λ, b_r and b_v, respectively. χ^2 is chi-squared statistics for the test of the overidentifying restrictions and the numbers in bracket are corresponding p-values.

choices of instruments yield qualitatively the same result. Comparing χ^2 statistic and t-statistics of λ, the statistical inference coincides with all the three measures of standard errors except for t-statistic of $\hat{\lambda}$ for $Level_t$ using the distance to default decile model. That is, with any measure of standard errors, the model cannot be rejected at 10% level, $\hat{\lambda}$ for $Level_t$ is statistically insignificant and $\hat{\lambda}$ for $Slope_{t+1}$ is statistically significant. The only difference is that $\hat{\lambda}$ for $Level_t$ using the distance to default decile model becomes marginally significant. Overall, the distance to default decile model seems to produce standard errors that are too small. This may be due to the assumption that the sensitivity of the individual error terms to the industry level shock is restricted to be one. If there is significant heterogeneity in an industry such that the assumption is unrealistic, the resulting standard errors might be wrong. Nonetheless, it is comforting that the two factor model is still far from being statistically rejected.

On the other hand, the one factor model produces very similar result to the clustered standard errors. Overall, with alternative measures of standard errors, I did not find significant evidence against clustered standard errors.