How Important Are Non-Default Factors for CDS Valuation? A Non-parametric Analysis

Biao Guo*, Qian Han**, and Doojin Ryu***

* School of Finance & China Financial Policy Research Center, Renmin University of China, No. 59 Zhong Guan Cun Street, Haidian District, Beijing, 100872, China, Email: tigerguob@gmail.com.

** Room A402, Economics Building, Wang Yanan Institute for Studies in Economics, Xiamen University, Xiamen, China 361005, China, Email: hanqian@gmail.com.

*** School of Economics, College of Business and Economics, Chung-Ang University, Dongjak-gu, Seoul 156-756, Korea, Email: doojin.ryu@gmail.com

How Important Are Non-Default Factors for CDS Valuation? A Non-parametric Analysis

Abstract

To which extent can non-default components explain CDS (Credit Default Swap) spreads is under debate in the literature. Unlike other research applying conventional structural or reduced-form models, this study investigates this issue by conducting a principal component analysis and a non-parametric local linear regression using corporate CDS data during the period from 2001 to 2011, which includes the recent global financial crisis. A model with two market state factors, approximated by the first two components, are found to outperform a model with only one factor. The first component capturing the variation in the overall level of CDS spreads can explain 90.82% of the variation in the data, resulting in a 21.58 basis points (bps) root mean square error (RMSE). The second component, orthogonal to the first component by construction, is barely explained by variables implied by models of default and explains an additional 5.06% of the variation, helping to reduce the RMSE to 8.99 bps. The out-of-sample tests support the in-sample analysis, finding that a default-factor model performs much worse after the beginning of the financial crisis in 2007. The study provides support for the recent findings that liquidity and counterparty risk are priced in CDS spreads besides credit risk and sheds light on CDS valuation especially for volatile periods.

Keywords: Credit Default Swap (CDS), Local Linear Regression, Principal Component, non-default

JEL classification: C13, C14, G13, G14

1. Introduction

A CDS is a credit derivative in which the protection buyer makes a series of payments (CDS spreads) to the protection seller and, in exchange, receives a payoff in the event of a default. Understanding the variation in CDS spreads and its determinants has become increasingly important for investors because of the substantial size of the CDS market[1], the common practice

[1] Although CDSs have been around since the early 1990s, the CDS market has expanded sharply since 2003, increasing to USD 62.2 trillion (in the notional amount) by the end of 2007 and stabilizing to USD 30.4 trillion by the end of 2009. In comparison to CDSs, the amount of total equity derivatives was USD

of using CDSs to hedge against defaults, and the recent 2007-2009 financial crisis. Many default-risk models have been developed and applied to understand the dynamics of CDS spreads[2]. However, only recently researchers start to realize that non-default components may also play significant roles in explaining CDS spreads and the question of how non-default components explain CDS spreads is still under debate. Researches supporting pure default components include Schueler and Galletto (2003), in which the authors assume that CDS spreads are contract values without liquidity risk, and Longstaff et al. (2005), in which the authors argue that CDS is a contract that can always be created, thus it is not subject to a liquidity constraint and has no liquidity premium. Other research, however, suggests that variables for non-default risk are critical in CDS spreads valuation and may explain the well documented credit spread puzzle. Lin et al. (2011) find that non-default spreads attribute 13% to CDS spreads. Tang and Yan (2007) regress CDS spreads on non-default variables that capture expected liquidity and liquidity risk, and find that illiquidity produces higher spreads. Bongaerts et al. (2011) show strong evidence of liquidity spread using an equilibrium asset pricing model and conclude that liquidity risk has a significant effect on CDS spreads. Pu et al. (2010) argue that both liquidity and counterparty related variables determine CDS spreads.

While significant advances have been made in previous studies, two flaws remain. First, the empirical methods employed in previous studies fail to quantify the importance of non-default factors for CDS valuation. In other words, does a default model suffice to value CDS contracts for both in-sample and out-of-sample? How much performance improvement we can achieve by adding non-default factors? Second, existing studies examine the determinants of CDS spreads either by adding a factor into a reduced-form pricing model and checking its significance as in Lin et al. (2011) or by regressing CDS spreads on selected default and non-

---

10.0 trillion for 2007 and USD 6.8 trillion for 2009. http://www.isda.org/statistics/pdf/ISDA-Market-Survey-annual-data.pdf.

2 Previous studies have proposed two approaches for default risk. The structural models are based on the idea that a firm defaults when its value drops below a certain threshold. Early important theoretical work includes Black and Cox (1976), Merton (1974), Geske(1977), Longstaff and Schwartz (1995), and many others. Reduced form models, in comparison to structural models in which the credit spread is endogenously determined by the issuer's balance sheet, assume that there are exogenously specified stochastic processes for factors driving the movement of credit spreads. Influential work in this area includes, among others, Das (1995), Das and Sundaram (1998), Duffie and Singleton (1999), Hull and White (2000a, 2000b), Jarrow and Turnbull (1995), Lando (1998), Pierides (1997), and Schonbucher (2000). See Arora, et al. (2006), Jarrow (2011) and Jarrow and Protter (2004) for comparisons between the structural and the reduced-form models.

default variables as in Pu et al. (2010). However, these methods may suffer the issue of model misspecification either due to an imposed reduced or structural functional form, or due to omitted explanatory variables, since there is no consensus on what the non-default variables should be.

In this study we address these two issues by conducting a non-parametric analysis on both in-sample and out-of-sample model performances. Specifically, we adopt a non-parametric method called local linear regression with state variables approximated by principal components (PCs) extracted from CDS spreads. The non-parametric estimation method resolves the issue of model misspecification due to either an imposed reduced or structural functional form and has been widely applied in derivative valuation. Li and Zhao (2009) apply it to estimate the state-price densities implicit in interest rate caps and demonstrate the high accuracy of this method via a simulation study. Ait-Sahalia and Duarte (2003) and Li and Zhang (2010) adopt this non-parametric method for S&P 500 index options and show its good fittingness. Analogous to Li and Zhang (2010), in this study we first extract state variables with principal component analysis (PCA), an orthogonal transformation to convert a set of correlated CDS spreads into a set of linearly uncorrelated PCs.[3] We then fit CDS spreads non-parametrically as a function of the PCs. Since by construction the first PC explains the largest variance of the CDS spreads and the second PC explains the second largest variance and is uncorrelated with the first PC, we posit that the first PC is attributable for the default components and the second PC captures the non-default components.[4] Using the fitted CDS spreads, we can examine the performance of a model with and without a non-default factor.

We apply the methodology to weekly data from January 2002 to November 2011 of USD-denominated senior unsecured corporate CDSs with 1-, 2-, 3-, 5-, 7-, and 10-year time to maturity. We divide CDSs to four groups (AA/AAA, A, BBB, and below BBB (BBB-)) by credit ratings, take the median of each group on each observation date as the spread of that group for

---

[3] Li and Zhang (2010) propose the use of nonlinear principal components (NPCs) for determining the number of state variables for implied volatility modeling. Our separate study using NPCs for CDSs suggests that the use of NPCs does not influence the results in this study qualitatively.

[4] An alternative way of viewing this issue is since the first PC is for the default factor, the residuals from the regression of CDS spreads on the first PC should then capture the non-default components. We extract a common factor from the residuals by a PCA method again, and find that this common factor is highly correlated with the second PC with a 96.83% correlation. Both methods yield similar conclusions for this study.

that day. PCs are extracted from grouped CDSs spreads of various credit ratings and maturities[5]. We find that the first PC accounts for 90.82% of the total variation in the CDS spreads and the second PC explains an additional 5.06%. An OLS (ordinary least square) regression analysis indicates that the change of the first PC is well explained by the change of default variables implied by structural model of default: risk-free rate, yield spread, business cycle, jump magnitude and the square of risk-free rate, while the change of the second PC is barely related to these variables but can be explained by variables related to liquidity and counterparty risk, suggesting the rationale of representing the first and second PCs as the default and non-default state variables. Our non-parametric results show that the root mean squared error (RMSE) is 21.58 bps when only the first PC is included in the model and decreases to 8.99 bps when the second PC is included. Furthermore, a bootstrap test strongly rejects the null hypothesis that the models with one and two PCs perform equally in CDS valuation. Lastly, we conduct an out-of-sample analysis to relieve the concern of possible over-fitting caused by adding a variable. Our out-of-sample results indicate that a model with two PCs outperforms in all sample periods but the year of 2006. The improvement of adding the second PC becomes more pronounced after 2007, the beginning of the recent financial crisis. Overall, both the in-sample and out-of-sample tests lead to the conclusion that adding a non-default factor into CDS valuation helps to better understand the dynamics of CDS spreads, and this benefit is larger for short-term CDSs and in a volatile market. The rising importance of non-default risk after the crisis is similar to the findings by studies on other assets, for instance, Gefang et al. (2011) model the spreads between the short-term London Interbank Offered Rate (Libor) and overnight index swap (OIS) and find in the financial crisis most major events are more linked to liquidity risk than credit risk. Cassola and Morana (2012) examine the Euro financial market and find supporting evidence that beyond credit risk, liquidity risk is relevant during the financial crisis.

[5] It is worth noting that our whole analysis is conducted from the viewpoint of a CDS portfolio manager whose daily work involves controlling and diversifying the CDS-specific risks and of a financial modeler who tries to determine the number of state factors in her CDS pricing equation. To that end, in this study we focus on the number of market state variables by classifying CDSs into 24 sub-groups and take the median of CDS spreads for each sub-group, in contrast to the analysis on individual CDS spreads as in Cont and Kan (2011). While it is widely known that individual CDS spreads are subject to idiosyncratic risks, CDS portfolios can average them out and allow us to focus on market factors. A similar group classification is done by Li and Zhang (2010) for option prices and by Bongaerts et al. (2011) for CDS contracts.

The rest of the paper is organized as follows: Section 2 describes the data. Section 3 introduces the methodology of principal component analysis and non-parametric valuation. Empirical results are discussed. Section 4 concludes this study.

2. Data: Senior Unsecured CDS

We obtain weekly mid-quotes for USD-denominated senior unsecured CDSs with 1-, 2-, 3-, 5-, 7-, and 10-year time to maturity for the period from January 2002 to November 2011. Weekly data is used to reduce the noise in daily data and to provide a larger sample size than monthly data. We include only those CDSs satisfying the following three screening criteria: first, the CDS must have at least one-year trading data; second, it must have quotes for all maturities on each observation date; and third, it must have a modified restructuring (MR) clause. The first criterion excludes any CDS that disappears soon after being listed or is issued recently, the second criterion chooses only those CDSs with enough liquidity, and the third criterion is applied because a restructuring clause can change the recovery rate in the event of a default and thus, various clauses may have differential effects on the CDS spread valuation method. The above screening returns 308,202 quotes issued by 913 reference entities. We classify CDSs within each maturity category according to their credit ratings[6] into four groups: AA/AAA, A, BBB, and below BBB (BBB-). We then take the median[7] of all CDS quotes in each group as its CDS spread on each observation date.

Table 1 shows the average CDS spreads and their standard deviations for the four rating groups. Both the average and standard deviation (S.D.) of spreads are generally larger for CDSs with lower credit rating and longer time to maturity. For instance, the 1-year CDS spread for AA/AAA is 18.68 bps (S.D.=18.48 bps), whereas that for BBB- is 176.95 bps (S.D.=153.84 bps). The A-rated spread for 1-year maturity is 25.17 bps, whereas that for 10-year is 61.09 bps. Figure 1 visualizes the average CDS spreads for the four rating groups. The charts indicate clear paradigm shifts in the behavior of these spreads after June 2007. CDS spreads were stable before 2007 but have been quite volatile since then, peaking around the middle of 2009 regardless of credit ratings. An additional observation is that the spreads of CDSs with different maturity dates tend to move together.

---

[6] We compute averaged rating for any entity with multiple rating records for the sample period.

[7] Median instead of mean is used to reduce the impact of outlier.

3. Methodology and Empirical Results

3.1 Principal Component Analysis

In this section, we conduct the PCA on the aforementioned CDS datasets. The whole datasets are divided into 24 groups according to their credit rating and number of years to maturity. This classification guarantees enough data in our sample since not many CDSs have long enough common trading dates, and furthermore, the grouping averages out the CDS-specific risks and allows us to focus on market factors (Li and Zhang 2010, Bongaerts et al. 2011). A standard PCA is then conducted and the first four PCs are extracted using these 24 time series. PCA uses an orthogonal transformation to convert a set of correlated CDS spreads into a set of linearly uncorrelated PCs. The transformation is defined in such a way that the first PC explains the largest variance of the data; the second PC is orthogonal to the first PC and explains the second largest variance, and so on. As reported in Table 2, the first four PCs explain 90.82%, 5.06%, 2.68%, and 0.77%, respectively, of the total variation in the 24 average CDS spreads. The eigenvalues of the first four PCs are 21.80, 1.21, 0.64, and 0.19, respectively. According to the simple stopping rule suggested by the Kaiser-Guttman method, PCs sufficient for data analysis are those with eigenvalues bigger than one, which implies that only the first two PCs warrant our attention. With this finding and the major motivation to investigate the roles of default and non-default state variables in CDS valuation, we focus on the first two PCs in our subsequent analysis.

Figure 2 plots the eigenvectors for the first two PCs. The coefficients for the first eigenvector are all positive and similar in magnitude, suggesting that the first PC captures the variation in the overall level of CDS spreads. The second eigenvector for short-term CDSs has negative or small positive coefficients, whereas that for long-term CDSs shows large positive coefficients, suggesting that the second PC captures the variation in the slope along the maturity dimension.

Figure 3 shows the time series plots of the first two PCs. Not surprisingly, the first PC exhibits a pattern similar to that of the average CDS spreads in Figure 1, which is consistent with the large percentage of the total variation it explains and its reflection of the overall level of CDS spreads. On the contrary, the second PC increases gradually with a less clear pattern.

3.2 Regression Analysis on PCs

Given the importance of representing the default and non-default state variables with the first and the second PC, in this section we conduct OLS regression analyses to examine the economic meaning of the first two PCs.

We first adopt the regression model suggested by Collin-Dufresne et al. (2001) and Ericsson et al. (2009)[8],

$$\Delta p_t = \beta_0 + \beta_1 \Delta r_t + \beta_2 \Delta \text{VOL}_t + \beta_3 \Delta r_t^2 + \beta_4 \Delta \text{slope}_t + \beta_5 \Delta \text{jump}_t + \beta_6 \Delta \text{S\&P}_t + \varepsilon_t \qquad (1)$$

where β_0 is an intercept, Δp_t, Δr_t, ΔVOL_t represents the change in PC, risk-free rate, and volatility, respectively, at *t*. The square of risk-free rate, r_t^2 is included to capture the nonlinear relationship between default spreads and risk-free rates; slope_t represents the difference between the long-term (10-year) and short-term (2-year) risk-free rate in order to reflect the magnitude of the instantaneous short rate; $\Delta \text{S\&P}_t$ is the return of the S&P 500 to reflect the overall state of the economy; jump_t is a proxy for jumps in market value to control for the effect of a jump on credit spread. Following Collin-Dufresne et al. (2001) and Ericsson et al. (2009), we approximate risk-free rate as the ten-year Treasury yield; volatility as the VIX index, a measure of the implied volatility of S&P 500 index option; jump as the slope of the smirk of implied volatilities σ of European put options on the S&P 500 index, which reflects the probability of extreme moves. The slope measures the steepness of volatility smirk and is an indicator of jump magnitude of an asset. The larger the slope, the steeper the smirk and the higher probability of a jump in an asset's value. Define moneyness $m_i = \ln\left(\frac{K_i}{S}\right) / \sqrt{T_i}$, where K_i is the strike price, S is the S&P 500 index value, T_i is the time to maturity of a European option *i* on date t. The slope of the smirk b for date t is then estimated via an ordinary linear regression $\sigma(m_i) = a + b m_i + \varepsilon$.

Panel A of Table 3 reports the results of regression (1) for the first and second PC. The first PC is highly explained by those variables with 18.14% adjusted R^2; all variables are significant except the VIX and yield spread. The signs of coefficient estimates are consistent with the common understanding of the determinants of default spreads: smaller risk-free rate and S&P 500 return, or larger volatility, yield spread and jump magnitude lead to a wider change of

[8] We omit the leverage ratio variable as an explanatory variable because it is at an individual CDS firm level while our regression analysis is at an aggregate level, which unlikely changes our results given the strong significance shown in Table 3.

default spread (first PC). On the contrary, these variables barely explain the second PC with only 3.13% adjusted R^2; the only two significant variables are yield spread and the S&P 500 return,; but the sign of S&P 500 return becomes less intuitive.

We further investigate the explanatory power of non-default market variables on the two PCs. We follow Pu et al. (2010)'s argument about the role of liquidity and counterparty risk priced in CDS.

$$\Delta p_t = \beta_0 + \beta_1 \Delta r_t + \beta_2 \Delta \mathrm{VOL}_t + \beta_3 \Delta r_t^2 + \beta_4 \Delta \mathrm{slope}_t + \beta_5 \Delta \mathrm{jump}_t + \beta_6 \Delta \mathrm{S\&P}_t + \beta_7 \Delta \mathrm{LiborRepo}_t + \beta_8 \Delta \mathrm{Gamma}_t + \beta_9 \Delta \mathrm{DebtIssue}_t + \beta_{10} \Delta \mathrm{MMMF}_t + \varepsilon_t \quad (2)$$

where LiborRepo represents the spread between the 3-month Libor and Repo rate. Pu et al. (2010) use the LiborRepo to measure the aggregate counterparty risk as it describes the spread between secured and unsecured loan rates. DebtIssue is the total dollar volume of corporate debt issued in the fixed income market, its change reflects the new debt in the market and thus affects the liquidity of the financial market; MMMF is the total money market mutual fund assets, a variable whose changes capture the inflow and outflow of funds to the money market and is typically associated with the market liquidity situation; our final variable for liquidity is Gamma, a liquidity measure proposed by Bao et al. (2011),

$$\gamma = \mathrm{cov}(\Delta s_t, \Delta s_{t-1})$$

where s_t is the CDS spread on date t. It measures the covariance between consecutive CDS spread changes[9]. We use positive covariance instead of the negative sign in Bao et al. (2011) for corporate bonds since, by definition, CDS spread is approximately the difference between bond yield and risk-free rate and its change is, therefore, negatively correlated with corporate bond price return. Higher γ indicates stronger illiquidity. We first calculate γ for each CDS and then use the cross-sectional median γ as the aggregate γ liquidity measure for the CDS market, analagous to Bao et al. (2011).

We re-organize the data at a monthly frequency as only monthly DebtIssue and MMMF are available. Panel B of table 3 presents the results of regression (2). First, the estimates of those default-related variables on the first PC are largely consistent with the results in regression (1),

[9] Based on the model in Bao et al. (2011), an asset's return Δs consists of two components $\Delta s_{\mathrm{t}} = \Delta \mathrm{f_t} + \Delta \mathrm{u_t}$, where the first component $\mathrm{f_t}$ represents the fundamental value without any friction and follows a random walk, the second component $\mathrm{u_t}$ is a transitory term uncorrelated with $\mathrm{f_t}$ and represents the impact of illiquidity. The covariance $\mathrm{cov}(\Delta S_{\mathrm{t}}, \Delta S_{\mathrm{t-1}})$ thus depends only on the transitory component and captures its magnitude.

suggesting the robustness of the above findings after controlling for additional variables; second, only one non-default factor, MMMF, is significant in explaining the first PC, a stark contrast to the estimates for the second PC, where LiborRepo, Gamma and MMMF are all significant, indicating the strong explanatory power of liquidity and counterparty risk for the second PC.

In summary, Table 3 indicates that the first PC and the second PC represent fairly well the default and the non-default components of CDS yield.

3.3 Non-parametric Estimation

After extracting the PCs as state variables and examining their economic meanings, in this section we fit a non-parametric model with those extracted PCs. Let $s_{i,J,T}$ be the spread of a CDS with T years to maturity issued by a reference entity i with credit rating J, the stochastic process of $s_{i,J,T}$ is governed by a M vector of state variables $x=\{x_1, x_2, ..., x_M\}$ following a Markov process, and the valuation function of the CDS spread can be formally expressed as $s=f(x, J, T)$[10], where f is a linear or nonlinear function related to the payoff structure of the CDS.

To avoid model misspecifications, we choose for the CDS valuation a non-parametric estimation method called local linear regression, which has been widely applied in the field of finance, including the valuation of the interest rate cap (Li & Zhao (2009)) and the pricing of S&P 500 options (Ait-Sahalia & Duarte (2003)，Li & Zhang (2010)). Let p_k be the k^{th} extracted PC, and $k=\{1, 2\}$. The CDS spread s is a function of (p_k, J, T), the coefficients α and β and thus the estimator of s are estimated by minimizing the following local linear regression equation:

$$\sum_{i=1}^{N}\left[s_i - \alpha - \sum_{k=1}^{K}\beta_k(p_{k,i} - p_k) - \beta_{K+1}(J_i - J) - \beta_{K+2}(T_i - T)\right]^2 \times \prod_{j=1}^{K}\frac{1}{h_p}G\left(\frac{p_{k,i}-p_k}{h_p}\right)\frac{1}{h_J}G\left(\frac{J_i-J}{h_J}\right)\frac{1}{h_T}G\left(\frac{T_i-T}{h_T}\right) \qquad (3)$$

where s_i is the observed CDS spread, N is the number of observations, $G()$ is a kernel function, and h is the associated bandwidth for the kernel function, $K=1$ for the model with one PC and $K=2$ for the model with two PCs. It is well known that the choice of the kernel function has little effect on the estimation, whereas that of the bandwidth h determines the accuracy of the final outcome. Thus, we choose the widely used second-order Gaussian kernel

---

[10] The recovery rate measures the amount that a creditor can receive upon a default. In this paper, we assume an equal and constant recovery rate for all CDSs regardless of their credit ratings, as in Longstaff et al. (2005) and others.

$$G(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

and *h* by the least squares cross-validation method (Li and Racine, 2004).

We can then estimate the RMSE[11] for the case k=1,2 and calculate $R_k^2 = 1 - \frac{RMSE_k^2}{RMSE_{k-1}^2}$ following Li and Zhang (2010) to gauge the improvement in the performance of adding the second PC. If the two-PC model performs better than the one-PC model, $RMSE_2^2 < RMSE_1^2$ and $0 < R_2^{2<} < 1$. Obviously, R_2^2 can be negative when $RMSE_2^2 > RMSE_1^2$ and a negative value implies that the model with two PCs performs worse than that with one PC.

Table 4 presents the results using the local linear regression equation (3). The RMSE is 21.58 bps when the first PC is included in the model, but the addition of the second PC reduces the total RMSE to 8.99 bps. The 83% partial R^2 suggests a significant performance improvement of the model with the addition of the second PC. Table 4 also reports the subtotal RMSE by the rating group and time to maturity. Across the rating groups, BBB- shows the largest RMSE, which is as expected given the large variation in CDS spreads for this group (Table 1).To be more specific, for the model with one PC, the value of RMSE for the BBB- group is 5.76 times higher than that for the AA/AAA group. The ratio increases to 7.09 after the second PC is priced in, suggesting that a non-default factor contributes more to the AA/AAA rated CDSs than to the BBB- rated CDSs. The ratio of RMSE of the two-PC model to that of the one-PC model for the 1-year CDS is 33.54%, the lowest among maturities, indicating the relative insufficiency of explaining short-term CDS spread with a default factor alone. This finding is consistent with other studies on the fallacy of default models for short-term defaultable assets, for example, by Huang and Huang (2002).

Figure 4 shows the time series of average residuals. The residuals are large in 2003 and 2008-2010, two volatile periods caused by the accounting scandal and the financial crisis, and they decrease when the second PC is added. The decrease is especially prominent during the financial crisis, suggesting the rising importance of non-default components for CDS spreads.

In sum, we find that the model with two PCs improves the performance of the model with one PC both cross-sectionally and over time. The first PC, representing default components,

---

[11] We also estimate the mean absolute percentage error (MAPE), a measure of accuracy specifically in trend estimation and obtain similar conclusions.

is unable to capture the full information of CDS spreads, while the role of non-default components, represented by the second PC, become more crucial for short-term CDSs and during the crisis. The conclusions are similar for both the RMSE and residual analysis.

3.4 Performance Bootstrap Test

In this section, we conduct a rigorous test by using a bootstrap procedure to investigate whether two models are the same in terms of their valuation performance. Specifically, we consider the following hypotheses:

$$H_0: s(p_1, J, T)=s(p_2, J, T)$$
$$H_1: s(p_1, J, T)\neq s(p_2, J, T)$$

The intuition behind the test is that if the unrestricted model with two PCs shows a larger improvement in valuation performance than the restricted model with one PC, then CDSs valued under these two models should be statistically different from each other, rejecting H_0. By contrast, H_0 cannot be rejected if one model performs only marginally better than the other model.

We adopt a two-point wild bootstrap method (see Li & Wang (1998), Li & Zhang (2010)) for the test. Li and Wang (1998) demonstrate that this test has good finite-sample properties. We first estimate CDS spreads with the restricted model $s(p_1, J, T)$ and compute the residuals as $\varepsilon_i = s_i - s_i(p_1, J, T)$, where s_i is the market-observed CDS spread. We then construct the two-point wild bootstrap residuals as $\hat{\varepsilon}_i = (\frac{1-\sqrt{5}}{2})\varepsilon_i$ with probability $p = \frac{1+\sqrt{5}}{2\sqrt{5}}$, and as $\hat{\varepsilon}_i = (\frac{1+\sqrt{5}}{2})\varepsilon_i$ with probability[12] $1-p$. The bootstrap samples are generated as $\hat{s}_i = s_i(p_1, J, T) + \hat{\varepsilon}_i$. We then calculate new partial $\hat{R}_k^2$ for each set of bootstrap samples. By comparing the original partial R_2^2 with the $\hat{R}_2^2$ from many sets of bootstrap samples, we can compute the *p*-value for the null hypothesis. For example, if $R_2^2 > \hat{R}_2^2$ for more than 90% of total sets of bootstrap samples, we can conclude that the *p*-value is 10% and reject the null hypothesis at the 10% significance level. The *p*-value in Table 4 reports the outcome of 100 sets of bootstrapped samples. The zero *p*-value suggests a significant difference in performance between the model with one PC and the one with two PCs. This result is not surprising considering the sharp decreases in the RMSE and residuals.

---

[12]This construction guarantees that the bootstrap residuals satisfy the following conditions: $\mathrm{E}[\hat{\varepsilon}_i] = 0, \mathrm{E}[\hat{\varepsilon}_i{}^2] = \varepsilon_i{}^2$, and $\mathrm{E}[\hat{\varepsilon}_i{}^3] = \varepsilon_i{}^3$, where E[.] is the expectation operator.

3.5 Out-of-sample Tests

We have shown that two PCs help to understand the CDS spreads. A potential concern is that a model with multiple variables may suffer over-fitting and actually underperform a parsimonious model. To determine if this is the case, we conduct an out-of-sample analysis on the model with two PCs, against the model with one PC only. We test the performance of each year in our sample in order to capture better the capabilities of models under different market conditions. We first fit parameters by using all data other than the given year and then use the fitted parameters to value the CDS spreads in that year. For example, in order to compare model out-of-sample performances for year 2003, we use data from year 2002, and 2004 to 2011 for local linear regression parameter estimation and compute the model spreads for year 2003 with the estimated parameter set.

Out-of-sample RMSE is reported in Table 5 as a comparison criterion, together with the partial R^2. Two important observations arise: first, the model with two PCs outperforms the model with one PC for all years except 2006, when the partial R^2 becomes negative. The overall improvement is consistent with the in-sample results shown in Table 4. Second, the significance of the second PC becomes stronger after 2007, as with a larger partial R^2 than previously. The last column makes the differences clearer by showing the ratio of $RMSE_{2PC}/RMSE_{1PC}$, a smaller percentage indicating a more dramatic decrease of RMSE, and hence a more important role played by the second PC. Not surprisingly, the percentage during the financial crisis is much smaller than other periods, consistent with the indispensable role the non-default factor plays in volatile periods as shown in section 3.3.

4. Conclusion

This paper investigates the question of whether non-default components are significant for CDS spreads valuation and in particular how the former explains the latter. We group the CDSs by credit ratings and maturities to average out idiosyncratic risk factors. To avoid the issue of model misspecification, we use a non-parametric estimation method to value CDSs. We proxy market state factors by the PCs extracted from historical CDS spreads. Our results show clearly and quantitatively that non-default factors, represented by the second PC, are significant in valuing CDSs. A model with default and non-default factors generally outperforms a model with only a default factor, as indicated by a lower RMSE value. The improvement is more pronounced for short-term CDSs and during the financial crisis. The rigorous bootstrap test, together with the out-of-sample tests, provides further support for this conclusion. Overall, our results are not only revealing for CDS valuation, but also instructive to those CDS portfolio risk managers who need to implement default risk control on a regular basis.

References

Ait-Sahalia, Y. and Duarte, J. (2003)."Nonparametric option pricing under shape restrictions." Journal of Econometrics **116**(1-2): 9-47.

Arora, N., Bohn, J. R., and Zhu, F. (2006). "Reduced Form vs. Structural Models of Credit Risk: A Case Study of Three Models." Journal of Investment Management **3**.

Bao, J., Pan, J. and Wang, J. (2011), The illiquidity of corporate bonds, Journal of Finance, 66(3), 911–946.

Black, F., and Cox J. (1976). "Valuing corporate securities: Some effects of bond indentures provisions." Journal of Finance 31:351–367.

Bongaerts, D., De Jong, F., and Driessen, J. (2011). "Derivative Pricing with Liquidity Risk: Theory and Evidence from the Credit Default Swap Market." Journal of Finance 66(1): 203-240.

Brunnermeier, M. K., and Pedersen, L. (2009)."Market Liquidity and Funding Liquidity." Review of Financial Studies. 22(6): 2201-2238.

Cassola, N., Morana, C. (2012). “Euro Money Market Spreads during the 2007 - ? Financial Crisis.” Journal of Empirical Finance 19: 548-557.

Collin-Dufresne, P., Goldstein, R. S., and Martin, J. S. (2001). "The determinants of credit spread changes." Journal of Finance 56(6): 2177-2207.

Cont, R., and Kan, Y., (2011), Statistical Modeling of Credit Default Swap Portfolios. Working paper.

Das, S. R. (1995). "Credit risk derivatives." Journal of Derivatives 2:7–23.

Das, S. R., and Sundaram, R. K. (1998). "A direct approach to arbitrage-free pricing of credit derivatives." NBER Working paper NO. 6635.

Duffie, D., and Singleton, K. J. (1999). "Modeling term structures of defaultable bonds." Review of Financial Studies 12(4): 687-720.

Ericsson, J., Jacobs, K. and Oviedo, R. (2009). "The Determinants of Credit Default Swap Premia." Journal of Financial and Quantitative Analysis 44(01): 109-132.

Gefang, D., Koop, G., and Potter, S., M. (2011). "Understanding Liquidity and Credit Risks in the Financial Crisis". Journal of Empirical Finance 18(5): 903-914.

Geske, R. (1977). "Valuation of Corporate Liabilities as Compound Options." Journal of Financial and Quantitative Analysis 12(4): 541-552.

Huang, J.Z. and Huang, M. (2002). "How much of the corporate-treasury yield spread is due to credit risk?" NYU Working Paper No. S-CDM-02-05.

Hull, J. C., and White, A. (2000a). "Valuing credit default swaps I: No counterparty default risk." Journal of Derivatives 8: 29–40.

Hull, J. C., and White, A. (2000b). "Valuing credit default swaps II: Modeling default correlations." Journal of Derivatives 8: 12–22.

Jarrow, R. A. (2011). "Credit market equilibrium theory and evidence: Revisiting the structural versus reduced form credit risk model debate." Finance Research Letters8(1): 2-7.

Jarrow, R. A. and Protter, P. (2004). "Structural Versus Reduced Form Models: A New Information Based Perspective." Journal of Investment Management2(2).

Jarrow, R. A., Turnbull, S. M. (1995). "Pricing Derivatives on Financial Securities Subject to Credit Risk." Journal of Finance50(1): 53-85.

Lando, D. (1998). "On Cox processes and credit risky securities." Review of Derivatives Research2(2): 99–120.

Li, G. and Zhang, C. (2010). "On the Number of State Variables in Options Pricing." Management Science56(11): 2058-2075.

Li, H., and Zhao, F. (2009). "Nonparametric Estimation of State-Price Densities Implicit in Interest Rate Cap Prices." Review of Financial Studies, 22(11): 4335-4376.

Li, Q., and Racine, J. (2004). "Cross-validated local linear nonparametric regression." Statistica Sinica14(2): 485-512.

Li, Q., and Wang, S. (1998). "A simple consistent bootstrap test for a parametric regression function." Journal of Econometrics87(1): 145-165.

Lin, H., Liu, S., and Wu C. (2011). " Dissecting Corporate Bond and CDS Spreads." The Journal of Fixed Income 20(3): 7-39.

Longstaff, F. A., Mithal, S., and Neis, E. (2005). "Corporate yield spreads: Default risk or liquidity ? New evidence from the credit default swap market." Journal of Finance60(5): 2213-2253.

Longstaff, F. A. and Schwartz, E. S. (1995). "A Simple Approach to Valuing Risky Fixed and Floating Rate Debt." Journal of Finance50(3): 789-819.

Merton, R. C. (1974). "Pricing of Corporate Debt - Risk Structure of Interest Rates." Journal of Finance29(2): 449-470.

Pierides, Y. A. (1997). "The pricing of credit risk derivatives." Journal of Economic Dynamics and Control21:1579–1611.

Pu, X., Wang, J., and Wu, C. (2011). "Are Liquidity and Counterparty Risk Priced in the Credit Default Swap Market?" Journal of Fixed Income 20(4).

Schonbucher, P. J. (2000). "The pricing of credit risk and credit derivatives." Unpublished manuscript, University of Bonn.

Schueler, M., and Galletto, R., (2003), "Basis Trading." Working Paper, JP Morgan Credit Derivatives Marketing.

Tang, D., and Yan, H. (2007). “Liquidity and credit default swap spreads”, working paper, University of Hong Kong.

Table 1: CDS spreads in the sample (in bps)

Average	1Y	2Y	3Y	5Y	7Y	10Y
AA/AAA	18.68	23.33	27.25	35.62	39.75	44.64
A	25.17	30.83	36.76	47.94	53.83	61.09
BBB	47.61	57.20	67.05	85.03	92.79	102.28
BBB-	176.95	214.10	246.84	295.89	306.62	315.46
Standard Deviation						
AA/AAA	18.48	20.98	22.44	25.79	25.72	25.83
A	21.47	22.75	24.48	27.17	27.03	26.94
BBB	41.52	42.99	44.77	47.11	45.72	44.64
BBB-	153.84	158.83	161.73	163.83	155.29	146.30

This table shows the average CDS spreads and standard deviations for four rating groups (with 1, 2, 3, 5, 7, and 10 years to maturity).

Table 2: Variance explained by the first four PCs

	PC1	PC2	PC3	PC4
Variance explained	90.82%	5.06%	2.68%	0.77%
Cumsum	90.82%	95.88%	98.57%	99.34%

The first row is the percentage of the variance explained by each PC, and the second row is the cumulative percentage of the variance explained.

Table 3: Regression of the first two PCs on selected explanatory variables

	PC1	PC2
Panel A: Default		
Constant	0.0058	0.0073
VIX	0.0125	0.0068
10Y Treasury	-2.7593***	-0.2365
Yield Spread	0.4446	0.3568***
Yield Square	0.2559**	0.0178
Jump	2.3540**	0.2091
S&P 500	-0.0047***	0.0017***
Adjusted R2	0.1814	0.0313
Panel B: All		
Constant	-0.0234	0.0353
VIX	0.0285	-0.0029
10Y Treasury	-8.0075***	0.3928
Yield Spread	1.4463**	0.1186
Yield Square	0.8071***	-0.0364
Jump	-1.2850	1.0538
S&P 500	-0.0078***	0.0012
LiborRepo	-0.2596	0.2970**
Gamma	-0.0207	0.0727*
DebtIssue	0.0000	0.0000
MMMF	0.0056**	-0.0013**
Adjusted R2	0.5591	0.1359

Panel A of this table shows the results of the weekly OLS regression of the first two PCs on the following explanatory variables: VIX, the 10-year Treasury yield, the spread between the 10-year and 2-year Treasury yield, the square of 10-year Treasury yield, the jump magnitude and the S&P 500 index return:

$$\Delta p_t = \beta_0 + \beta_1 \Delta r_t + \beta_2 \Delta \mathrm{VOL}_t + \beta_3 \Delta r_t^2 + \beta_4 \Delta \mathrm{slope}_t + \beta_5 \Delta \mathrm{jump}_t + \beta_6 \Delta \mathrm{S\&P}_t + \varepsilon_t.$$

Adjusted R^2 results for explanatory power are shown in the last row. Panel B presents the monthly regression results after additional variables are included:

$$\Delta p_t = \beta_0 + \beta_1 \Delta r_t + \beta_2 \Delta \mathrm{VOL}_t + \beta_3 \Delta r_t^2 + \beta_4 \Delta \mathrm{slope}_t + \beta_5 \Delta \mathrm{jump}_t + \beta_6 \Delta \mathrm{S\&P}_t + \beta_7 \Delta \mathrm{LiborRepo}_t + \beta_8 \Delta \mathrm{Gamma}_t + \beta_9 \Delta \mathrm{DebtIssue}_t + \beta_{10} \Delta \mathrm{MMMF}_t + \varepsilon_t,$$

where LiborRepo, Gamma, DebtIssue and MMMF represent the spread between 3-month Libor and Repo rate, Gamma liquidity measure, total dollar volume of corporate debt issued in the fixed income market, and fund in the money market, respectively. ***Significant at the 1% level. **Significant at the 5% level. *Significant at the 10% level.

Table 4: Valuation performance of the first two PCs based on local linear regression

	Total			Decomposition									
PC	RMSE (bps)	Partial R^2	p-value	AA/AAA	A	BBB	BBB-	1Y	2Y	3Y	5Y	7Y	10Y
1	21.58			7.09	6.16	10.37	40.83	26.66	21.70	18.02	18.87	20.68	22.45
2	8.99	0.83	0.00	2.44	2.54	3.35	17.30	8.94	9.32	8.61	8.43	8.98	9.58

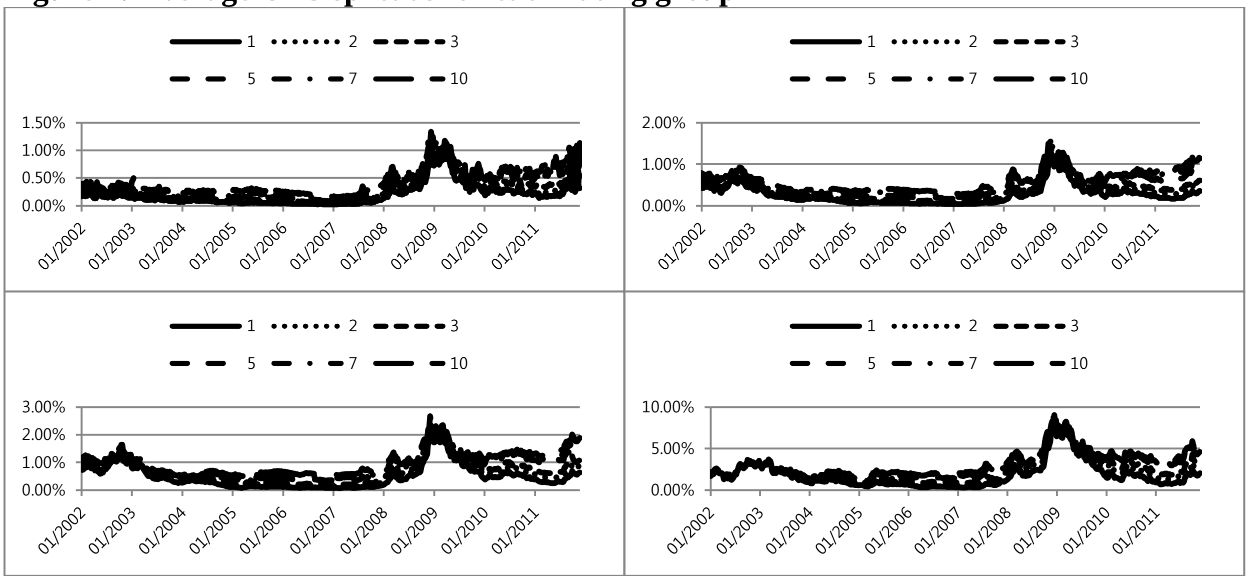
This table shows the results for valuation performance based on local linear regression. "Total" shows the statistics for the model with one or two PCs of the whole sample. The associated partial $R^2 = 1 - \frac{RMSE_2^2}{RMSE_1^2}$ is estimated to gauge the relative performance of adding the second PC and *p*-value is for the null hypothesis that the model with one PC and the model with two PCs generate equal CDS spreads. Decomposition reports the subtotal RMSE in bps by rating group and year to maturity.

Table 5: Out-of-sample performance of the first two PCs based on local linear regression

Year	RMSE (bps)		Partial R^2	Percentage
	PC1	PC2		
2002	36.72	13.02	0.87	0.35
2003	23.58	14.35	0.63	0.61
2004	9.34	5.80	0.61	0.62
2005	6.59	6.58	0.00	1.00
2006	3.18	3.27	-0.06	1.03
2007	17.58	6.10	0.88	0.35
2008	22.32	10.86	0.76	0.49
2009	25.14	8.79	0.88	0.35
2010	18.85	10.30	0.70	0.55
2011	30.32	3.98	0.98	0.13
Overall	21.66	9.10	0.82	0.42

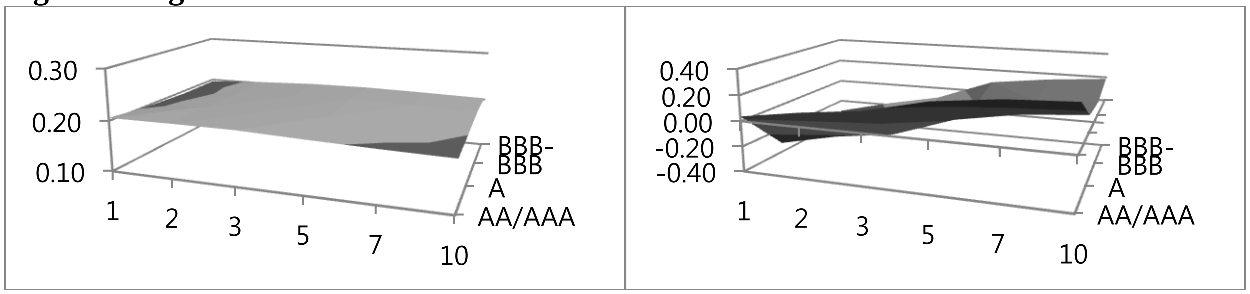
This table shows the results for out-of-sample performance based on local linear regression. For each year, model is first fitted using CDS spreads for the other nine years, forecast is then made for that year using the fitted model. RMSE in bps for the model with one or two PCs is shown for each year. The associated partial $R^2 = 1 - \frac{RMSE_2^2}{RMSE_1^2}$ is estimated to gauge the relative performance of adding the second PC. Percentage measures the ratio of RMSE of the model with two PCs to that of one PC.

Figure 1: Average CDS spreads for each rating group



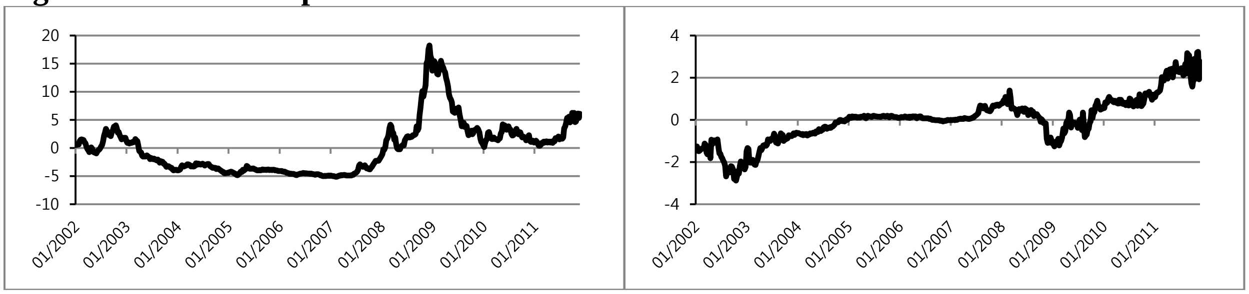
Average spreads for CDSs with 1, 2, 3, 5, 7, and 10 years to maturity for each rating group, plots from top left to bottom right are for AA/AAA, A, BBB, and BBB-, respectively. Y-axis is for CDS spreads.

Figure 2: Eigenvector results for the first two PCs



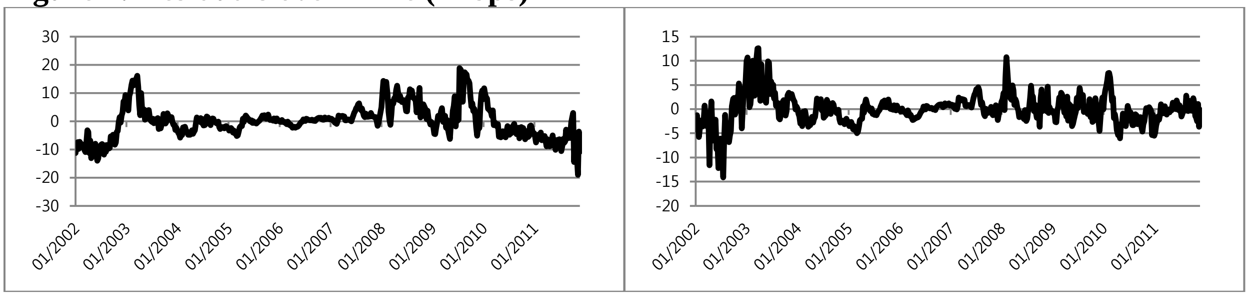
This figure shows the eigenvector results for the eigenvectors of the first two PCs for CDSs with 1, 2, 3, 5, 7, and 10 years to maturity and for rating AA/AAA, A, BBB and BBB-. Left is for the first PC and right is for the second PC. Z-axis is for eigenvector values.

Figure 3: Time series plots for the first four PCs



Left is for the first PC and right is for the second PC. Y-axis is for PC values.

Figure 4: Residuals over Time (in bps)



This figure plots average residuals from local linear regression over time. The performance for the model with one and two PCs is shown from the left to the right. Y-axis is for CDS spreads residuals.