

## **The Number of State Variables for CDS Pricing**

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

\* Finance & Accounting Division, Business School, Jubilee Campus, University of Nottingham,  
Nottingham, NG8 1BB, UK, tigerguob@gmail.com.

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

\*\*\* School of Economics, Chung-Ang University, Dongjak-gu, Seoul, South Korea 156-756; E-  
mail: doojin.ryu@gmail.com

# **The Number of State Variables for CDS Pricing**

## **Abstract**

This study investigates the number of state variables needed for CDS pricing by conducting a principal component analysis using CDS data for the 2006-2009 period. Two state variables, approximated by the first two components, are found sufficient for pricing CDS spreads. The first component corresponds to a market level factor and can explain over 97% of the variation in the data, resulting in a 20.04 bps root mean square error (RMSE). The second component corresponds to a liquidity level factor and can explain an additional 1.7% of the variation, helping to reduce the RMSE to 5.29 bps. A rigorous bootstrap test, together with two robustness tests, on the model performance improvement corroborates our conclusions. The study sheds light on CDS pricing and provides support for the most recent findings that liquidity risk is priced in CDS spreads.

*Keywords:* Credit Default Swap, Liquidity, Local Linear Regression, Principal Component

*JEL classification:* C13, C14, G13, G14

## **1. INTRODUCTION**

A credit default swap (CDS) is a credit derivatives in which the protection buyer makes a series of payments (often referred to as CDS spreads) to the protection seller and, in exchange, receives a payoff in the event of a default. Understanding the variation in CDS spreads has become increasingly important for investors because of the substantial size of the CDS market<sup>1</sup>, the

---

<sup>1</sup> 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, total equity derivatives amount to USD 10.0 trillion for 2007 and USD 6.8 trillion for 2009. <http://www.isda.org/statistics/pdf/ISDA-Market-Survey-annual-data.pdf>.

common practice of using CDSs to hedge against defaults, and the recent 2007-2009 financial crisis. Previous studies have proposed two approaches for pricing credit derivatives. 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 (1999), Duffie and Singleton (1999), Hull and White (2000a, 2000b), Jarrow and Turnbull (1995), Lando (1998), Pierides (1997), and Schonbucher (2000)<sup>2</sup>.

The above two approaches are also used to value corporate yield spreads since both of the bond yield and the CDS spread reflect credit risk of a specific reference entity<sup>3</sup>. However, although numerous studies have investigated the determinants of corporate bond yields<sup>4</sup>, it is only recently that researchers start to realize that factors other than those affecting the bond yield spreads play important roles in pricing credit derivatives<sup>5</sup>. In particular, liquidity risk has attracted a lot of attention. Longstaff et al. (2005) find that non-default components in corporate bond spreads are strongly related to liquidity measures. Tang and Yan (2007) regress CDS spreads on variables that capture expected liquidity and liquidity risk, and find that illiquidity produces higher spreads. Buhler and Trapp (2006, 2008) explicitly incorporate a liquidity intensity rate process into their reduced form models. Pan and Singleton (2008) examine the term structure of sovereign CDS spreads and claim that the second principal component (possibly

---

<sup>2</sup> See Arora, Bohn and Zhu (2006), Jarrow (2011) and Jarrow and Protter (2004) for comparisons between the structural and the reduced-form models.

<sup>3</sup> Until recently, empirical studies have equated the CDS spread with the yield spread of the corresponding corporate bond. Duffie and Singleton (1999) and Hull and White (2000a) provide theoretical arguments that the two should be equivalent under no arbitrage conditions. Blanco, Brennan, and Marsh (2005) test this relationship with CDS data from 2001 to 2002 and provide empirical support for the hypothesis.

<sup>4</sup> Typical variables considered in the literature include the risk-free rate and default intensity rate (Longstaff & Schwartz (1995)), the issuing firm's value volatility (Campbell & Taksler (2003)), value jumps and the overall business climate (Collin-Dufresne et al. (2001)).

<sup>5</sup> For example, Houweling and Vorst (2005) show that a simple reduced form model outperforms bond spreads in forecasting the CDS spreads, implying that CDS spreads may contain components missing in bond spreads.

related to the liquidity spread) is needed to explain the severe mispricing of one-year contracts. More recently, Bhanot and Guo (2011) show that the deviations between the CDS spread and the corporate bond spread can be explained by funding liquidity and asset-specific liquidity. Pu et al. (2011) conduct an in-depth empirical analysis and suggest that market-wide liquidity, CDS-specific liquidity, and market-wide counterparty risk are important factors to be included in pricing models. Bongaerts et al. (2011) find a strong evidence of the liquidity spread and conclude that liquidity risk has a significant effect on CDS spreads. Other variables are also considered in the literature. Jacobs and Li (2008) add in their credit risk model a stochastic volatility factor of the default rate. Jankowitsch et al. (2008) consider an additional variable for monitoring the cheapest-to-deliver (CTD) option embedded in CDSs. Ericsson et al. (2009) add firm value leverage as a pricing component.

A research question then naturally arises: how many state variables suffice for accurately valuing CDS contracts? Moreover, is liquidity a priced factor? The present study answers these questions by first conducting a principal component (PC) analysis using CDS quotes issued on a broad range of reference entities. We find that the first two PCs account for more than 99% of the total variation in the CDS spreads. A regression analysis indicates that the first component corresponds to a market level factor influencing all CDS contracts similarly, and is significantly related to market volatility, market index level, and the 10-year Treasury yield. Altogether they explain about 95% of the variation of the first component. The second component reflects the variation across the rating categories. The spread between the three-month Libor rate and the three-month treasury yield, a proxy for market-wide liquidity risk,<sup>6</sup> can explain a significant portion (44%) of the second component. The third component captures the curvature of the ratings, whereas the fourth component seems to reflect the term structure characteristics of the CDS data.

To further investigate the number of state variables for CDS valuation, we use a non-parametric method called local linear regression to price the CDS contracts. The non-parametric estimation method resolves the issue of model misspecification due to either an imposed reduced or structural functional form. We allow the risk-free rate, approximated by the US Treasury yield,

---

<sup>6</sup> For example, see Bhanot and Guo (2011), Brunnermier and Pedersen (2009).

to be one of the state variables because risk-free rate is widely accepted as a determinant of the CDS spread. The remaining state variables are approximated by principal components (PCs) extracted from the time series of CDS spreads to circumvent inappropriate definitions of state variables. Our results show that the root mean squared error (RMSE) decreases from 189.37 bps to 20.04 bps when the first PC is included in the pricing model and further decreases to 5.29 bps when the second PC is included. By contrast, the third and the fourth PCs only slightly decrease the RMSE to 4.68 bps and 4.90 bps, respectively. An examination of the model residuals leads to the similar conclusion that two state variables are sufficient for CDS valuation.

To corroborate our findings, we conduct a rigorous bootstrap test on the model performance improvement. Consistent with the results using PC analysis and the non-parametric pricing model, significant pricing performance improvement is discovered when the first and the second PC are added to the model. Adding the third and/or the fourth PC, however, does not statistically improve the pricing performance. Two alternative robustness tests based on an ordinary least square (OLS) regression confirm our conclusion. Overall, these results shed light on the CDS pricing by demonstrating that two state variables are sufficient for modeling and that liquidity should be considered in CDS pricing<sup>7</sup>.

The rest of this paper is organized as follows: Section 2 describes the data. Section 3 introduces the methodology of principal component analysis, non-parametric valuation and the bootstrap test used in this study. Empirical results are discussed. Section 4 regresses the first two principal components against several market variables and discusses their implications. Section 5 concludes this study.

## 2. DATA

---

<sup>7</sup> This paper is also related to previous research on the number of state variables for option pricing. Li and Zhang (2010) propose the use of nonlinear principal components (NPCs) for determining the number of state variables for implied volatility modeling. However, the question of whether the relationship between CDS spreads of different expiration dates is nonlinear or not is a subject of debate, and thus, we focus on linear PCs. Our separate study using NPCs suggests that the use of NPCs does not influence the results.

We obtain daily mid-quotes for USD-denominated senior unsecured CDSs with 1, 2, 3, 5, 7, and 10 years time to maturity for the period from January 2006 to August 2009. We exclude 4-year CDSs because we could not obtain daily 4-year U.S. Treasury yields. We include only those CDSs satisfying the following two screening criteria: First, the CDS must have at least one-year trading data; and second, it must have a modified restructuring (MR) clause. The first criterion excludes any CDS that disappears soon after being listed or is issued recently, and the second criterion is the same as that in Pu et al. (2011) 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, and the MR clause is common in the U.S. market. The above screening leaves 703,153 quotes issued by 892 reference entities. We then classify CDSs according to their credit ratings<sup>8</sup> into three groups: above BBB (BBB+), BBB, and below BBB (BBB-).<sup>9</sup> In addition, we take the average of all CDS quotes in each group. We obtained 1-, 2-, 3-, 5-, 7-, and 10-year Treasury yields as the proxy for the risk-free rate from the Federal Reserve.

Table 1 shows the average CDS spreads and their standard deviations for the three rating groups. Both the mean and standard deviation (S.D.) spreads are larger for lower credit rating CDS. For instance, the average 1-year CDS spread for BBB+ is 83.55 bps (S.D.=101.99 bps), whereas that for BBB- is 541.48 bps (S.D.=713.80 bps). Short-term CDS spreads are generally narrower and more volatile than long-term spreads. For example, the average 1-year CDS spread for BBB is 116.28 bps (S.D. =140.66 bps), whereas the 10-year average CDS spread for BBB is 150.27 (S.D. =101.35 bps). Figure 1 visualizes the average CDS spreads for the three rating groups. The charts indicate clear changes in the behavior of these spreads after June 2007, the beginning of the recent subprime mortgage crisis. CDS spreads are stable and narrow before 2007 but are volatile in the latter half of 2007, peaking around the middle of 2009 regardless of credit ratings. In addition, the spreads of CDSs with different maturity dates tend to move together.

### 3. METHODOLOGY AND EMPIRICAL RESULTS

<sup>8</sup> We compute averaged rating for any entity with multiple rating records for the sample period.

<sup>9</sup> We classify CDSs into three groups such that the numbers of reference entities in each group are close. There are 252, 350 and 290 reference entities for BBB+, BBB, and BBB-, respectively.

### 3.1 Principal Component Analysis

We classify CDS data sets into 18 groups according to the credit rating and the number of years to maturity. A standard principal component analysis (PCA) is conducted using these 18 time series. As shown in Table 2, the first four PCs explain 97.79%, 1.72%, 0.34%, and 0.10%, respectively, of the total variation in the 18 average CDS spreads. The first PC explains a dominant percentage of the total variation, and the first and second PCs together explain 99.5% of the total variation. However, the addition of the third and fourth PCs has little improvement in explanation.

Figure 2 plots the eigenvectors for the first four 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 BBB- CDSs shows positive coefficients, whereas that for BBB+ CDSs shows negative coefficients, suggesting that the second PC captures the variation in the slope along the dimension of credit ratings. The third eigenvector for BBB CDSs shows positive coefficients, whereas that for both BBB- and BBB+ CDSs shows negative coefficients, suggesting that the third PC captures the variation in the curvature along the dimension of credit ratings. Finally, regardless of credit ratings, the fourth eigenvector for short-term CDSs shows negative coefficients, whereas that for long-term CDSs shows positive coefficients, suggesting that the fourth PC captures the variation in the slope along the maturity dimension.

Figure 3 shows the time series plots of the first four 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. The second and third PCs are relatively stable before 2008 but fluctuate afterward, and the fourth PC vacillates over the whole period, showing no clear pattern.

### 3.2 Non-parametric Estimation

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  $x=\{x_1, x_2, \dots,$

$x_M$  following a Markov process, and the pricing 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  $r$  be the risk-free rate,  $p_k$  be the  $k^{\text{th}}$  extracted PC, and  $k=\{0, 1, 2, 3, 4\}$ ,<sup>11</sup> with  $k=0$  being the pricing equation without a PC. The CDS spread  $s$  is a function of  $(p_k, r, J, T)$ , and the coefficients  $\alpha$  and  $\beta$  and thus the estimator of  $s$  are estimated by minimizing the following local linear regression equation:

$$\mathbf{1}_i(\mathbf{p} = \mathbf{1})^T N \mathbf{H}_i \mathbf{s}_i \mathbf{1} \propto \mathbf{1}_i(k = \mathbf{1})^T K \mathbf{H}_i \mathbf{k} \begin{bmatrix} p_k(k, t) \mathbf{1}_{p_k k} \\ \beta_1(K+1) (J_i \mathbf{1} \mathbf{1} J) \\ \beta_1(K+2) (T_i \mathbf{1} \mathbf{1} T) \\ \beta_1(K+3) (r_i \mathbf{1} \mathbf{1} r) \end{bmatrix} \mathbf{1}_i \mathbf{2} \times \mathbf{1}_i(\mathbf{p} = \mathbf{1})^T K \mathbf{H}_i \mathbf{H}_i \mathbf{p} \quad G((p_k(k, t) \mathbf{1}_{p_k k}) / \mathbf{1}_i \mathbf{p}) \parallel \mathbf{1}_i \mathbf{H}_i \quad G((J_i \mathbf{1} \mathbf{1} J) / \mathbf{1}_i \mathbf{J})$$

(1)

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. 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

and choose  $h$  by the least squares cross-validation method.<sup>12</sup>

We can then estimate the RMSE for the case  $k=0, 1, 2, 3, 4$  and calculate following Li and Zhang (2010) to gauge the improvement in the performance of adding the  $k^{\text{th}}$  component. If the

<sup>10</sup> 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.

<sup>11</sup> We test up to the 4<sup>th</sup> PC, but the inclusion of additional PCs does not improve the results.

<sup>12</sup> See Li and Racine (2004).



$$0 < R_k^2 < 1$$

k-PC model performs better than the (k-1)-PC model, then  $RMSE_k^2 < RMSE_{k-1}^2$  and

Obviously,  $R_k^2$  can be negative when and a negative value implies that the model with  $k$  PCs performs worse than that with  $k-1$  PCs because of problems such as overfitting.

Table 3 presents the results using the local linear regression equation (1). Panel A shows the total RMSE and the associated partial  $R_k^2$  for each  $k=0, 1, 2, 3, 4$ . When no PC is included in the pricing model, the total RMSE is 189.37 bps, which is rather large but not surprising because CDSs are valued only with the risk-free rate (i.e., no other state variables are considered). The RMSE decreases sharply to 20.04 bps when the first PC is included in the model, and the partial value of 98.88%  $R_1^2$  demonstrates a substantial improvement in the performance of the model when adding the first PC. The addition of the second PC further reduces the total RMSE to 5.29 bps, and the 93.03%  $R_2^2$  suggests additional significant performance improvement of the pricing model with the second PC. However, when the third PC is included in the model, the RMSE decreases only marginally from 5.29 bps to 4.68 bps with a 21.61%. Finally, the fourth PC results in an RMSE of 4.90 bps, and the -9.6%  $R_4^2$  suggests that the model with four PCs performs worse than that with three PCs. Ericsson et al. (2009) use a standard structural model of default and find an average RMSE of around 30 bps, which exceeds our results for all models with  $k=1, 2, 3, 4$  and is closest to the model with only the first PC. Given that we use a non-parametric estimation method, our errors are naturally smaller than theirs. In addition, the result that our model with the first PC has the closest RMSE as their default model is consistent with the discussion in Section 3.1, that is, the first PC may capture the variation in the overall level of CDS spreads and reflect the default risk.

Panel B of Table 3 reports the subtotal RMSE by the rating group and time to maturity. Among 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 without any PC, the value of RMSE for the BBB- group is 3.67 times higher than that for the BBB+ group. The ratio increases to 9.73, 6.29, 8.86, and 10.07 after the first, second, third, and fourth PCs, respectively, are priced in. This distinct pricing performance suggests that CDSs with the

high ratings tend to have greater pricing improvement than those with the low ratings when more state variables are included in the model. The 5-year CDS shows the lowest RMSE value, possibly due to the fact that 5-year CDSs are the most actively traded corporate CDSs and thus their quotes should be the most accurate ones and have the least number of outliers. The RMSE value for the 1-year CDS is 0.88, 1.70, 2.00, 1.92, and 1.82 times higher, and the RMSE value for the 10-year CDS is 2.19, 0.90, 1.48, 1.71, and 1.54 times higher than that for the 5-year CDS, for the model with 0, 1, 2, 3, 4 PCs, respectively.

Figure 4 visualizes the values of RMSE from the pricing equation (1) for models with different number of PCs. Each 3-D plot shows the pricing performance of the model with 0, 1, 2, 3, or 4 PCs against the rating group and time to maturity. First, the model with only the first PC performs considerably better than that with no PC regardless of the credit rating and time to maturity. The second PC further reduces the pricing errors, but the third and the fourth PCs make only marginal contributions. Second, the pricing errors “smile” along the maturity dimension, with the turning point around five years. This pattern is particularly pronounced for CDSs with low ratings and when the second PC is included in the model.

Figure 5 plots the RMSE for each rating group over time. All RMSE values become more volatile after 2007, particularly for the BBB- group. Nevertheless, the following performance pattern is clear and holds for all groups: The model with the first PC performs substantially better than that with no PC, and the second PC further enhances the performance of the model. However, the third PC increases performance only slightly, and the fourth PC worsens the pricing performance.

Figure 6 shows the time series of average residuals for pricing models. The residuals are large when the pricing model includes no PC (the plot in the top left corner), and they decrease when the first and second PCs are added. The third PC further reduces the magnitude of residuals and smoothes the few spikes observed in the model with the first and second PCs. The fourth PC does not improve pricing performance, and the model with the fourth PC even generates larger residuals than that with only the first three PCs.

In sum, these results suggest that the model with the first two PCs can capture a substantial portion of the variation in CDS spreads both cross sectional and over time, whereas the model

with more than two PCs performs only marginally better or even worse. The results remain for both the RMSE and residual analysis.

### 3.3 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 pricing performance. Specifically, we consider the following hypotheses:

$H_0$ :

$H_1$ :

where  $k=1, 2, 3$ , and  $4$ . The intuition behind the test is that if the unrestricted model with  $k$  state variables shows a larger improvement in pricing performance than the restricted model with  $k-1$  state variables, 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

$$s(p_{k1}, r, J, T).$$

first estimate CDS spreads with the restricted model and compute the residuals as

, where  $s_i$  is the market-observed CDS spread. We then construct the two-point

wild bootstrap residuals as with probability  $p = \frac{1 + \sqrt{5}}{2\sqrt{5}}$ , and as

$\varepsilon_i = \left(\frac{1 + \sqrt{5}}{2}\right) \varepsilon_i$  with probability <sup>13</sup> . The bootstrap samples are generated as . We then calculate new partial  $R_k^2$  for each set of bootstrap samples. By comparing the original partial  $R_k^2$  with the  $R_k^2$  from many sets of bootstrap samples, we can compute the  $p$ -value for the null hypothesis. For example, if  $R_k^2 > \hat{R}_k^2$  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 last row of Panel A of Table 3 reports the  $p$ -values for 100 sets of bootstrapped samples. The zero  $p$ -value for the model with one PC suggests a significant difference in pricing performance between the model with one PC and that with no PC. In addition, there is a significant performance difference between the model with one PC and that with two PCs. Thus, these results strongly reject the null hypothesis for  $k=1, 2$ . However, the null hypothesis that there is no difference in pricing performance between the models with the first two PCs and with the first three PCs cannot be rejected at the 10% significance level, suggesting that including the third PC in the pricing model results in no significant difference. This result is not surprising in that Section 3.2 shows only marginal decreases in the RMSE and residuals. Furthermore, the 0.38  $p$ -value for the model adding the fourth PC indicates that the pricing performance is statistically equal to the model with the first three PCs.

### 3.4 Alternative Robustness Tests

Since our conclusion that two state variables suffice for CDS valuation is based on the non-parametric local linear regression in section 3.2 and the bootstrap test in section 3.3, it may be a concern whether the conclusion is strongly depending on the applied method or test. In this section we adopt two alternative tests as a robustness check in order to relief the concern. The first test is described in Connor and Korajczyk (1993) and a more recent test is proposed in Bai and Ng (2002), both tests investigate the number of factors in approximate factor models.

<sup>13</sup> This construction guarantees that the bootstrap residuals satisfy the following conditions:  $\mathbf{E}[\varepsilon_i] = 0$ ,  $\mathbf{E}[\varepsilon_i^2] = \sigma_i^2$ , and  $\mathbf{E}[\varepsilon_i^3] = \sigma_i^3$ , where  $\mathbf{E}[\cdot]$  is the expectation operator.

Let  $s_{i,t}$ , where  $s_{i,t}$  is the market-observed CDS spread<sup>14</sup>,  $p_k$  is the  $k=\{0,1,2,3,4\}$  PCs,  $r$  is the risk free rate. A simple ordinary least square (OLS) regression returns the error term  $\varepsilon_i$  as the unexplained part of CDS spread by the first  $k$  PCs and the risk free rate. Connor and Korajczyk (1993) argue that if  $k$  is the correct number of factors then there should be no significant decrease of the variance of  $\varepsilon_i$ . By comparing the variance of  $\varepsilon_i$  for the model with  $k$  and  $k+1$  PCs and calculating the associated t-statistics, we are able to determine the number of sufficient state variables. Bai and Ng (2002) select the number of factors that minimizes the information criteria defined as  $IC(k) = \ln(V(k)) + k*((N+M)/N*M) * \ln(N*M/(N+M))$ , where  $V(k)$  is the average variance of  $\varepsilon_i$ ,  $N$ ,  $M$  is the number of groups and sample periods, respectively.

Table 4 reports the results for the two tests. Panel A is the t-statistics and associated p-values for Connor and Korajczyk test, both the zero p-values for  $k=0$  and  $k=1$  suggest that the null hypothesis that there is no significant decrease of the variance of  $\varepsilon_i$  between the model with  $k$  PCs and with  $k+1$  PCs is rejected. However, we can't reject the null hypothesis when  $k=2$ , suggesting that adding the third PCs doesn't decrease the variance significantly. Consistent with Connor and Korajczyk test, the result for Bai and Ng test shown in Panel B indicates that the model has the smallest information criteria when  $k=2$ .

Overall our robustness tests are coherent with the bootstrap test in section 3.3, indicating that two state variables, approximated by the first two components, are enough for CDS pricing, regardless of the regression methods.

#### 4. INTERPRETATIONS OF PRINCIPAL COMPONENTS

The empirical results reported in Sections 3 indicate that two state variables approximated by the first and second PCs are sufficient for CDS pricing. A further interesting question is what these two PCs represent in the real economic world. In Section 3.1, we observe from the eigenvectors that the first PC captures the variation in the overall level of CDS spreads and that the second PC captures the variation in the slope along the dimension of credit ratings. To better examine the

---

<sup>14</sup> We depress time  $t$  for easier notation.

economic meaning of these two PCs, we run a linear regression of each PC on several selected explanatory variables: VIX, the 10-year Treasury yield, the level of the S&P 500 index, and the spread between the 3-month Libor rate and the 3-month Treasury yield. VIX is a measure of the implied volatility of S&P 500 index options; the 10-year Treasury yield is a proxy for the risk-free rate; the level of the S&P 500 index is a measure of the overall business climate; and the spread between the 3-month Libor rate and the 3-month Treasury yield measures overall market liquidity. Collin-Dufresne et al. (2001) employ the first three variables to explain the default risk embedded in the yield spread, Brunnermeyer and Pedersen (2009) and Bhanot and Guo (2011) use the last variable as a proxy for market-wide liquidity risk. The motivation is that this spread reflects the shadow cost of capital and thus constraints the funding liquidity of arbitrageurs, which has a direct impact on the aggregate market liquidity. The higher the Libor-Repo spread is, the larger is the shadow cost of capital and the lower aggregate market liquidity.

Table 5 presents the regression results. First, among the four regressors, VIX, the 10-year Treasury yield, and the S&P 500 index level are highly significant, suggesting that the first PC captures the market aggregate default risk. Second, only the 10-year Treasury yield and the spread between the 3-month Libor rate and the 3-month Treasury yield are significant when the second PC is the dependent variable, suggesting that the second PC may represent the level of overall market liquidity. Third, the adjusted  $R^2$  for the first PC is 95.66%, demonstrating the adequacy of our selected variables for explaining the variation in the first PC. The adjusted  $R^2$  for the second PC (40.44%) indicates that a market liquidity variable is able to explain a substantial portion of the variation in the second PC. Possible candidates to explain the left portion include CDS-specific liquidity measures, since the second PC captures the variation in the slope along the credit rating dimension, and Predescu et al. (2009) show that CDS liquidity is correlated with its credit rating.

## 5. CONCLUSION

This paper investigates the question of how many state variables are sufficient for CDS pricing. To avoid the issue of model misspecification, we use a non-parametric estimation method to price CDS spreads. We approximate state variables by using the PCs extracted from historical

CDS spreads. The results show that two state variables, together with the risk-free rate and fundamental information embedded in CDSs, are sufficient for accurately pricing CDSs. A model with two state variables generally outperforms any model with fewer or more state variables, as indicated by its lower RMSE value. The rigorous bootstrap test, together with two robustness tests, provides support for this conclusion.

We regress the first and second PCs on several explanatory variables representing the default or liquidity risk and find that the first PC is largely explained by variables for the default risk and thus may capture the overall market default level. The proxy for the liquidity risk is highly significant in explaining the second PC, consistent with the observation from the eigenvector that the second PC captures the variation in the slope along the dimension of credit ratings. Our finding of a liquidity-like state variable provides another support for the recent argument that liquidity is priced in the CDS market (e.g., Bedendo et al. (2009), Pu et al., (2011), Bongaerts et al. (2011)).

## 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**.
- Bai, J., and Ng, S. (2002). "Determining the Number of Factors in Approximate Factor Models." Econometrica, Vol. 70, No. 1, 191-221.
- Bedendo, M., Cathcart, L., and El-Jahel, L. (2009). "Market and Model Credit Default Swap Spreads: Mind the Gap!" European Financial Management.
- Bhanot, K. and Guo, L. (2011). "Types of liquidity and limits to arbitrage—the case of credit default swaps." Journal of Futures Markets.
- Black, Fischer, and John C. Cox(1976). "Valuing corporate securities: Some effects of bond indentures provisions." Journal of Finance **31**:351–367.
- Blanco, R., Brennan, S., Marsh, I.(2005). "An empirical analysis of the dynamic relationship between investment-grade bonds and credit default swaps." Journal of Finance **60**: 2255–2281.

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, Markus K., Pedersen, Lasse Heje. (2009). "Market Liquidity and Funding Liquidity." Review of Financial Studies. (2009) 22(6): 2201-2238.

Buhler, W., Trapp, M. (2006). "Credit and liquidity risk in bond and CDS market. "Working paper.

Buhler, W., Trapp, M. (2008). "Time-Varying Credit Risk and Liquidity Premia in Bond and CDS Markets." Working paper.

Campbell, J. Y., Taksler, G. B. (2003). "Equity volatility and corporate bond yields." Journal of Finance **58**(6): 2321-2349.

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

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

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

Duffie, D., 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.

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

Gregory Connor and Robert A. Korajczyk (1993). "A Test for the Number of Factors in an Approximate Factor Model." Journal of Finance, 48(4): 1263-1291.

Houweling, P., Vorst, T. (2005). "Pricing default swaps: Empirical evidence." Journal of International Money and Finance **24**:1220–1225.

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

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

Jacobs, K. and X. F. Li (2008). "Modeling the dynamics of credit spreads with stochastic volatility." Management Science **54**(6): 1176-1188.



- Jankowitsch, R., Pullirsch, R., and Veza, T. (2008). "The delivery option in credit default swaps." Journal of Banking & Finance **32**(7): 1269-1285.
- Jarrow, R. A. (2011). "Credit market equilibrium theory and evidence: Revisiting the structural versus reduced form credit risk model debate." Finance Research Letters **8**(1): 2-7.
- Jarrow, R. A. and Protter, P. (2004). "Structural Versus Reduced Form Models: A New Information Based Perspective." Journal of Investment Management **2**(2).
- Jarrow, R. A., Turnbull, S. M. (1995). "Pricing Derivatives on Financial Securities Subject to Credit Risk." Journal of Finance **50**(1): 53-85.
- Lando, D. (1998). "On Cox processes and credit risky securities." Review of Derivatives Research **2**(2): 99–120.
- Li, G. and Zhang, C. (2010). "On the Number of State Variables in Options Pricing." Management Science **56**(11): 2058-2075.
- Li, H., 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., Racine, J. (2004). "Cross-validated local linear nonparametric regression." Statistica Sinica **14**(2): 485-512.
- Li, Q., Wang, S. (1998). "A simple consistent bootstrap test for a parametric regression function." Journal of Econometrics **87**(1): 145-165.
- 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 Finance **60**(5): 2213-2253.
- Longstaff, F. A. and Schwartz, E. S. (1995). "A Simple Approach to Valuing Risky Fixed and Floating Rate Debt." Journal of Finance **50**(3): 789-819.
- Merton, R. C. (1974). "Pricing of Corporate Debt - Risk Structure of Interest Rates." Journal of Finance **29**(2): 449-470.
- Pan, J. U. N. and Singleton K. J. (2008). "Default and Recovery Implicit in the Term Structure of Sovereign CDS Spreads." The Journal of Finance **63**(5): 2345-2384.
- Pierides, Y. A. (1997). "The pricing of credit risk derivatives." Journal of Economic Dynamics and Control **21**:1579–1611.

- Predescu, M., Thanawalla, R., Gupton, G., Liu, W., Kocagil, A., and Reyngold, A. (2009). "Measuring CDS Liquidity". Fitch Solutions presentation at the Bowles Symposium, Georgia State University, February 12 2009.
- 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, Philipp J. (2000). "The pricing of credit risk and credit derivatives." Unpublished manuscript, University of Bonn.
- Tang, Dragon, and Hong Yan, 2007, Liquidity and credit default swap spreads, working paper, University of Hong Kong.

**Table 1: CDS spreads in the sample (in %)**

| <b>Average</b>            | 1Y     | 2Y     | 3Y     | 5Y     | 7Y     | 10Y    | Num |
|---------------------------|--------|--------|--------|--------|--------|--------|-----|
| BBB +                     | 0.8355 | 0.8310 | 0.8451 | 0.8920 | 0.8981 | 0.9194 | 252 |
| BBB                       | 1.1628 | 1.2095 | 1.2752 | 1.4151 | 1.4531 | 1.5027 | 350 |
| BBB -                     | 5.4148 | 5.5423 | 5.6903 | 5.9541 | 5.8455 | 5.7613 | 290 |
| <b>Standard Deviation</b> |        |        |        |        |        |        |     |
| BBB +                     | 1.0199 | 0.9481 | 0.9108 | 0.8611 | 0.7902 | 0.7292 |     |
| BBB                       | 1.4066 | 1.3395 | 1.2964 | 1.2190 | 1.1109 | 1.0135 |     |
| BBB -                     | 7.1380 | 6.3537 | 5.8620 | 5.2004 | 4.6432 | 4.1732 |     |

This table shows the average CDS spreads and standard deviations for three rating groups (with 1, 2, 3, 5, 7, and 10 years to maturity) and the number of reference entities in each rating group.

**Table 2: Variance explained by the first four PCs**

|                    | PC1    | PC2    | PC3    | PC4    |
|--------------------|--------|--------|--------|--------|
| Variance explained | 97.79% | 1.72%  | 0.34%  | 0.10%  |
| Cumsum             | 97.79% | 99.51% | 99.85% | 99.95% |

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: Pricing performance of the first four PCs based on local linear regression (in %)**

| PC                            | 0      | 1      | 2      | 3      | 4       |
|-------------------------------|--------|--------|--------|--------|---------|
| <b>Panel A: total</b>         |        |        |        |        |         |
| RMSE                          | 1.8937 | 0.2004 | 0.0529 | 0.0468 | 0.0490  |
| Partial R2                    |        | 0.9888 | 0.9303 | 0.2161 | -0.0957 |
| p-value                       |        | 0.0000 | 0.0000 | 0.1400 | 0.3800  |
| <b>Panel B: decomposition</b> |        |        |        |        |         |
| BBB+                          | 0.7639 | 0.0353 | 0.0140 | 0.0090 | 0.0083  |
| BBB                           | 1.5208 | 0.0376 | 0.0213 | 0.0101 | 0.0074  |
| BBB-                          | 2.8039 | 0.3432 | 0.0880 | 0.0800 | 0.0842  |
| 1Y                            | 1.2366 | 0.2727 | 0.0662 | 0.0607 | 0.0648  |
| 2Y                            | 1.3314 | 0.2554 | 0.0739 | 0.0509 | 0.0531  |
| 3Y                            | 1.2489 | 0.1943 | 0.0441 | 0.0363 | 0.0389  |
| 5Y                            | 1.4128 | 0.1601 | 0.0332 | 0.0317 | 0.0357  |
| 7Y                            | 2.2601 | 0.1318 | 0.0385 | 0.0405 | 0.0401  |
| 10Y                           | 3.0904 | 0.1434 | 0.0492 | 0.0542 | 0.0549  |

This table shows the results for pricing performance based on local linear regression. Panel A shows the total RMSE for the number of PCs ( $k=0,1,2,3,4$ ). The associated partial  $R^2$  is estimated to gauge the relative performance of adding each PC, and the p-value is used for a non-parametric bootstrap test to determine whether there is a significant difference in pricing performance between a model with a PC and that with no PC. Panel B reports the subtotal RMSE by rating group and year to maturity.

**Table 4: Results for two alternative robustness test**

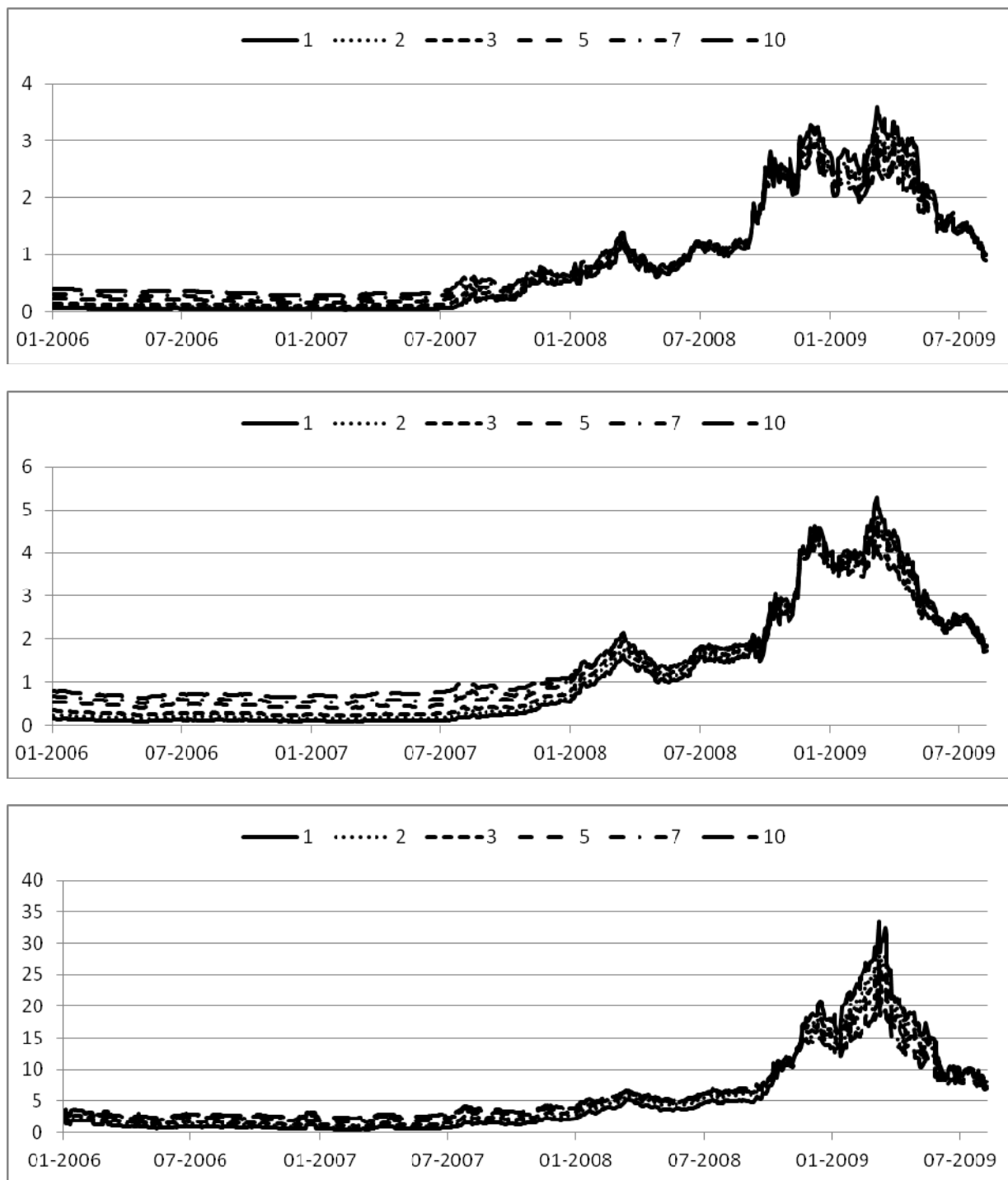
| PC                                     | 0       | 1       | 2        | 3        | 4        |
|----------------------------------------|---------|---------|----------|----------|----------|
| <b>Panel A: Connor &amp; Korajczyk</b> |         |         |          |          |          |
| t-stat                                 | 10.3415 | 5.1882  | 1.3494   | 1.7115   |          |
| p-value                                | 0.0000  | 0.0000  | 0.1779   | 0.0877   |          |
| <b>Panel B: Bai &amp; Ng</b>           |         |         |          |          |          |
| IC                                     | -6.3294 | -9.9619 | -10.2874 | -10.2226 | -10.1582 |

**Table 4: Regression of the first two PCs on selected explanatory variables**

|                         | PC1        | PC2        |
|-------------------------|------------|------------|
| Constant                | 16.4401*** | 1.0794***  |
| VIX                     | 0.0884***  | 0.0047     |
| 10Y Treasury            | -2.4643*** | -0.1197*** |
| S&P500                  | -0.0065*** | -0.0001    |
| Libor-Tbill             | -0.1218    | -0.6045*** |
| Adjusted R <sup>2</sup> | 0.9566     | 0.4044     |

This table shows the results of the linear regression of the first two PCs on the following explanatory variables: VIX, the 10-year Treasury yield, the S&P 500 index level, the spread between the 3-month Libor rate and the 3-month Treasury yield. Adjusted R<sup>2</sup> results for explanatory power are shown in the last row.

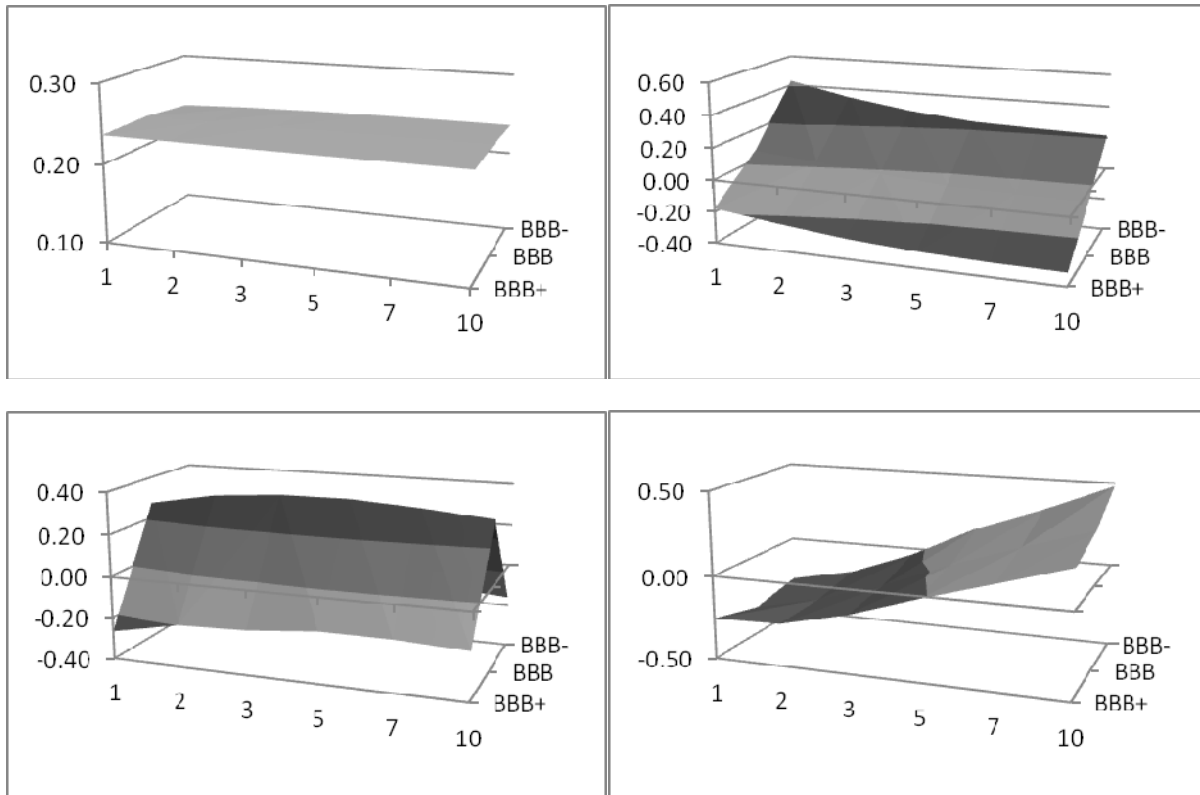
**Figure 1: Average CDS spreads for each rating group (in %)**



average spreads for CDSs with 1, 2, 3, 5, 7, and 10 years to maturity for BBB+ (the top), BBB (the middle), and BBB- (the bottom).

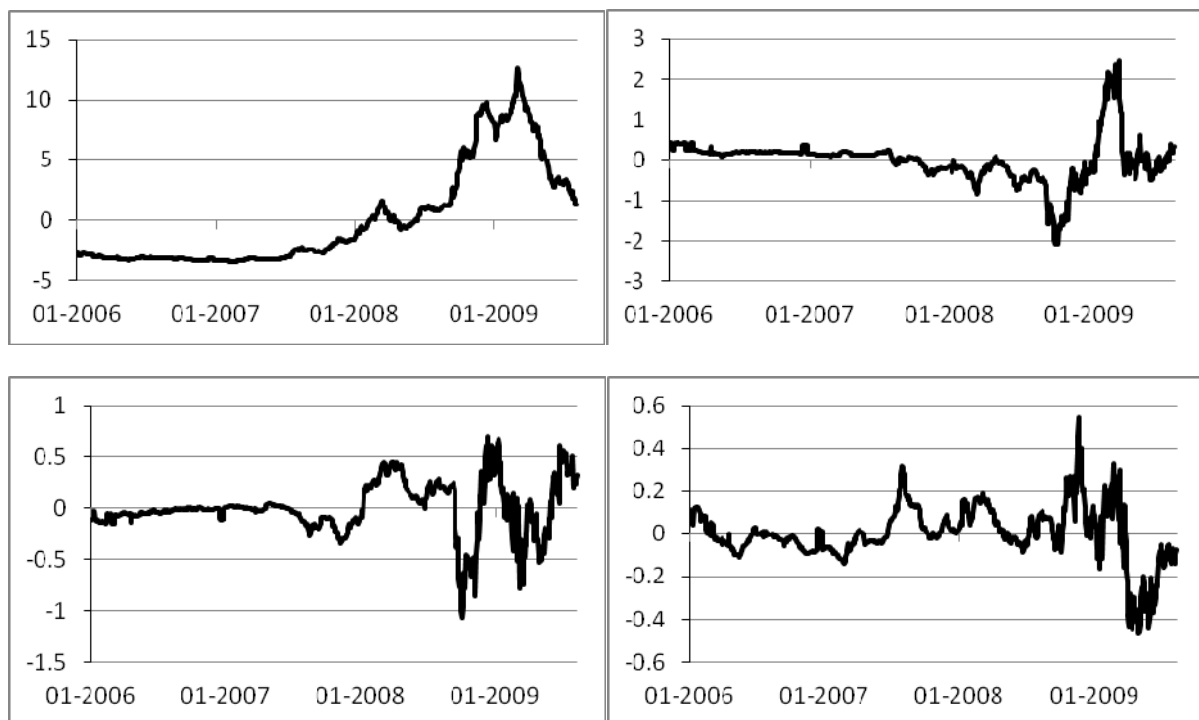


**Figure 2: Eigenvector results for the first four PCs**



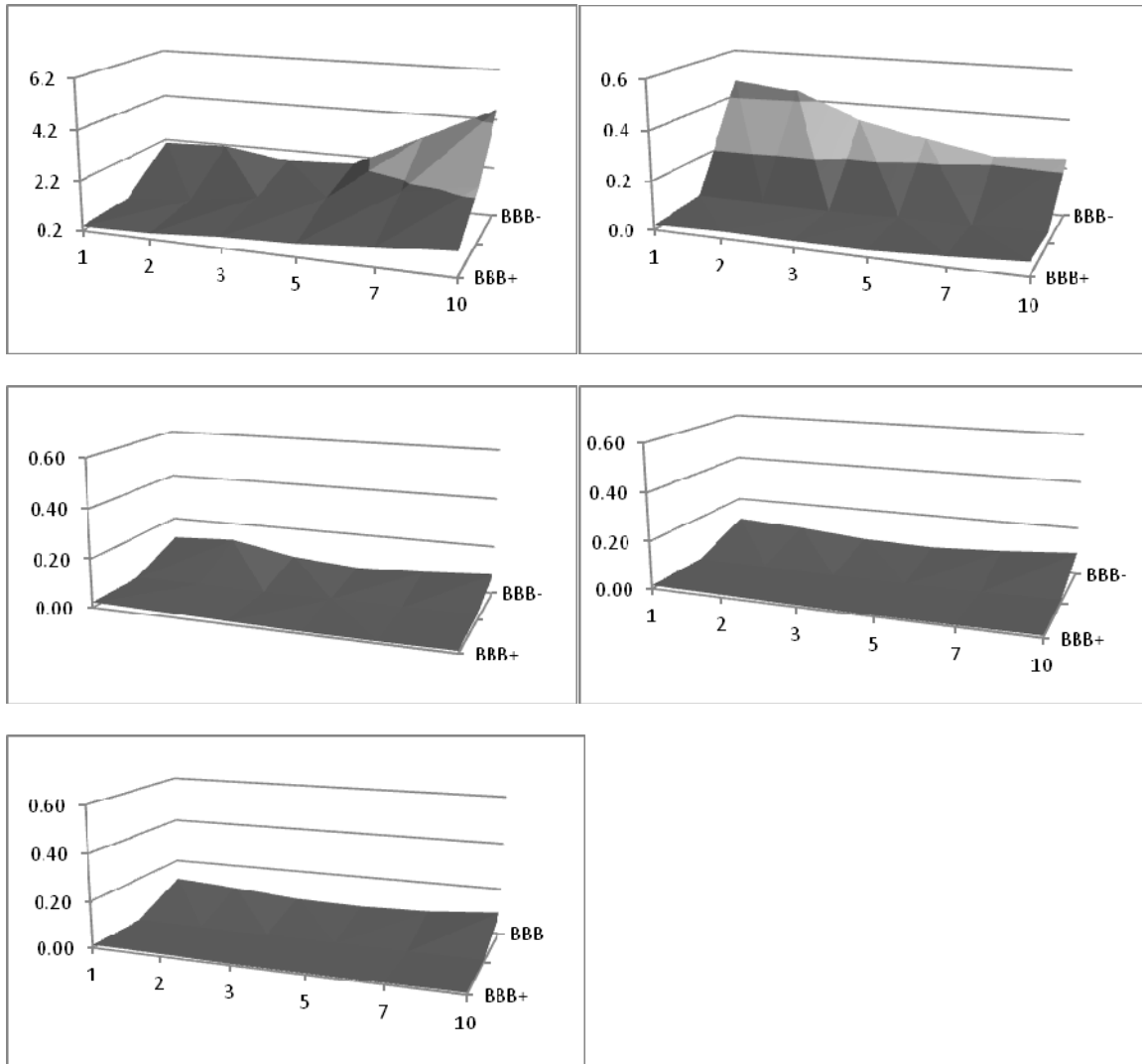
This figure shows the eigenvector results for the eigenvectors of the first four PCs for CDSs with 1, 2, 3, 5, 7, and 10 years to maturity for BBB+, BBB, and BBB-. The first, second, third, and fourth PCs are shown from the top left corner to the bottom right corner.

**Figure 3: Time series plots for the first four PCs**



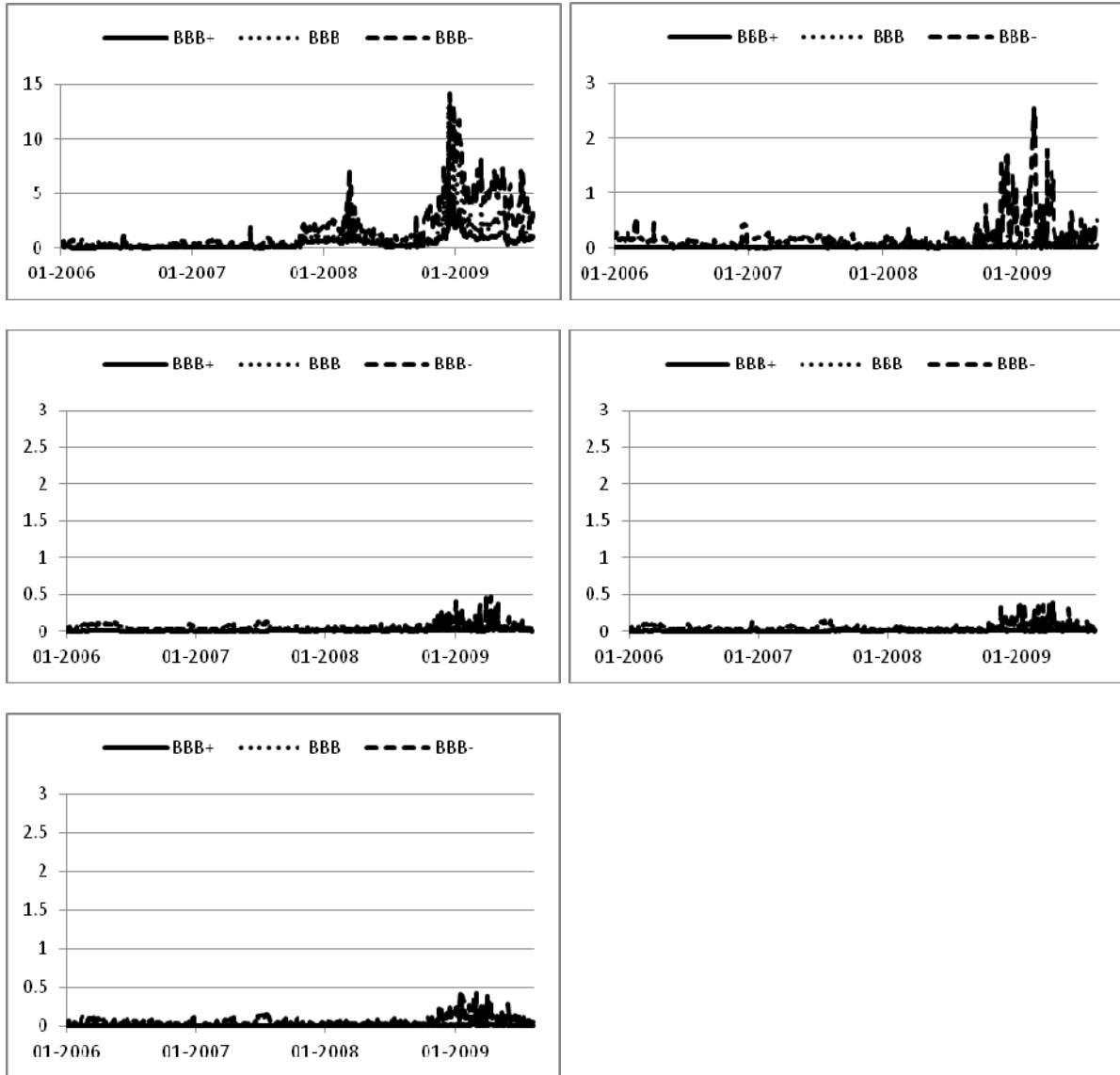
The time series plots for the first four PCs. The first, second, third, and fourth PCs are shown from the top left corner to the bottom right corner.

**Figure 4: Visualization of pricing errors for the first four PCs (in %)**



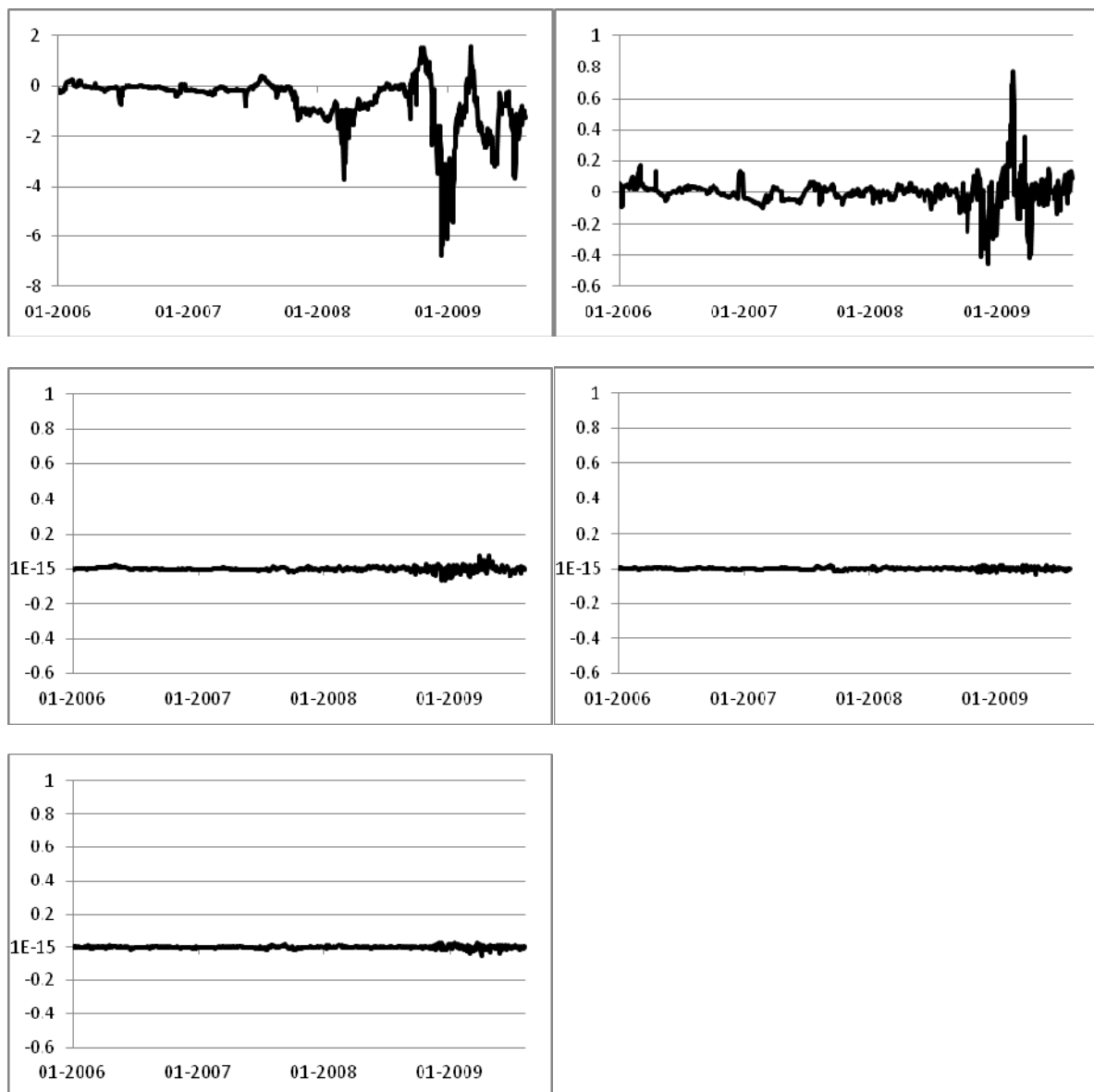
3-D plots of the RMSE from local linear regression, where the x-axis indicates the time to maturity from 1 to 10 years, and the y-axis indicates credit ratings from BBB+ to BBB-.

**Figure 5: RMSE over time (in %)**



The plots of the RMSE from local linear regression over time for each rating group. The first, second, third, and fourth PCs are shown from the top left corner to the bottom right corner.

**Figure 6: Residuals over Time (in %)**



This figure plots average residuals from local linear regression over time. The first, second, third, and fourth PCs are shown from the top left corner to the bottom right corner.