Tail return analysis of Bear Stearns’ credit default swaps

Liuling Li a, Bruce Mizrach b,⁎

a Nankai University, P.R. China
b Department of Economics, Rutgers University, United States

ARTICLE INFO

JEL classification:
C11
G13

Keywords:
Bear Stearns
Credit default swaps
Bayesian analysis
Exponential power distribution

ABSTRACT

We compare several models for Bear Stearns’ credit default swap spreads estimated via a Markov chain Monte Carlo algorithm. The Bayes Factor selects a CKLS model with GARCH–EPD errors as the best model. This model best captures the volatility clustering and extreme tail returns of the swaps during the crisis. Prior to November 2007, only four months ahead of Bear Stearns’ collapse though, the swap spreads were indistinguishable statistically from the risk-free rate.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

The first major investment bank to fail during the global financial crisis was Bear Stearns. Bear Stearns was an 85-year old institution that until the summer of 2007 had never had a losing quarter. Nine months later, it was gone, absorbed into J.P. Morgan Chase in a shotgun marriage. This paper looks at credit default swap (CDS) prices to see if they shed any light on this story.

The academic literature on credit risk suggests that the bulk of price discovery is going on in the CDS market. Blanco et al. (2005) study 27 single name CDS, including 15 financial firms, and find that, on average, the CDS market contributes 80% of price discovery. They attribute this to more informed trading in the CDS market. A second factor is the deep liquidity in the CDS market. In the first half of 2008, the Bank for International Settlements reported that notional volume of single name CDS outstanding peaked at $33.4 trillion. There was another $24 trillion in multi-name instruments.

Prior to the financial crisis, the CDS spreads on investment grade debt narrowed substantially. The benchmark 5-year investment grade CDX index from Markit fell from 60 to 20 basis points between 2004 and early 2007.¹ For this reason, we first consider the swap spread as a risk-free rate, using the model of Chan et al. (1992).

As the crisis unfolded though, spreads gradually widened. We find that the volatility in the series requires adding GARCH errors. As Bear Stearns flirts with bankruptcy, an exponential power distribution is needed to capture these statistically improbable events. Wu (2006) has proposed using a similar model for stock returns and argues that they better approximate the tail behavior of financial assets.

We begin our analysis in Section 2 by describing the cash flows in a credit default swap, and then proceed to analyze the Bear Stearns series. In Section 3, we consider progressively more general models for the swap spread. We start with the CKLS model, then add GARCH errors, and finally, add an exponential power distribution to describe the fat tails that remain after GARCH modeling. Estimation via Markov chain Monte Carlo is described in Section 4. In Section 5, we report results and find that the Bayes Factor selects the general CKLS–GARCH–EPD model as the best. We test for structural breaks in Section 6, and confirm that the swap spread is statistically similar to a risk-free process prior to November 2007. Section 7 concludes.

2. Data

2.1. Credit default swaps

Credit default swaps are derivative securities that pay off in the case of a credit event by the reference entity, typically a default. The protection seller agrees to provide any missing cash flows from the reference obligation to the buyer, including interest and principal. The protection buyer generally pays an up-front fixed fee and a swap spread that varies with the market’s assessment of the credit risk of the firm.

After signing the CDS contract, the buyer makes periodic payments, generally quarterly, to the seller until the maturity of the CDS or until a credit event occurs. The payment is calculated using the swap spread. Quoted in basis points (bp), or 0.01%, a spread of 180 bp, for example, implies that a protection buyer will pay $18,000 per year

KPUB-9993$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.econmod.2010.07.023
to insure $1 million worth of par value. A higher spread, holding other factors constant, indicates a greater likelihood of default.

In our setting, Bear Stearns is the reference entity. The reference obligations are 5-year bonds which represent, according to Jorion and Zhang (2007), 85% of the CDS market. Although Bear Stearns came under severe stress, their takeover by J.P. Morgan prevented any credit events for bond holders. In contrast, the Lehman Brothers CDS wound up paying $91.375 for every $100 of par value insured following their October 2008 bankruptcy.

2.2. The story behind Bear Stearns’ swap spreads

We use a daily time series of Bear Stearns’ 5-year CDS spreads. It spans from April 2006 to March 2008 which is the month of the takeover by JP Morgan. There are 501 observations. This data was purchased from GFI, a major international broker dealer with a strong presence in the over-the-counter derivative markets.

The CDS spreads are plotted in Fig. 1.

It is straightforward to map the changes in CDS spreads into the time line of news events in Table 1. These headlines were collected from the Wall Street Journal, the Financial Times, and Bloomberg.

The swap spreads traded at 30 bp or less until February 2007, just days before Bear Stearns reported its first ever loss on the High Grade Structured Credit Strategies Fund (SCSF). This was the less heavily leveraged of two Bear Stearns’ hedge funds with exposure to the subprime mortgage market. Mizrach (2010) notes that SCSF had gone 40 straight months without a loss, producing a cumulative 50% return.

Surprisingly though, spreads narrowed and fell back below 30 bp until June 2007 when Bear Stearns had to engineer a $3.2 billion bailout of its own funds. Spreads crossed 100 bp just before the two funds filed for bankruptcy in August 2007. Bear Stearns’ credit risk further deteriorated when Warren Spector, head of the fixed income division and co-president (with Alan Schwartz) resigned on August 6, 2007.

Problems began to spread beyond Bear Stearns at that point. BNP Paribas suspended redemptions in several of its funds, and this was soon followed by liquidity injections from both the European Central Bank (ECB) and the Federal Reserve.

<table>
<thead>
<tr>
<th>Date</th>
<th>Events</th>
</tr>
</thead>
<tbody>
<tr>
<td>01-Mar-2007</td>
<td>Bear Stearns (BSC) reports first ever loss on the High Grade Structured Credit Fund</td>
</tr>
<tr>
<td>14-Jun-2007</td>
<td>BSC reports a 10% decline in quarterly earnings.</td>
</tr>
<tr>
<td>18-Jun-2007</td>
<td>Merrill Lynch seizes collateral from BSC hedge funds.</td>
</tr>
<tr>
<td>22-Jun-2007</td>
<td>BSC commits $3.2 billion to High Grade Structured Credit Fund.</td>
</tr>
<tr>
<td>17-Jul-2007</td>
<td>BSC tells clients that assets in Enhanced Leverage Fund are essentially worthless.</td>
</tr>
<tr>
<td>01-Aug-2007</td>
<td>BSC hedge funds file for bankruptcy.</td>
</tr>
<tr>
<td>06-Aug-2007</td>
<td>Warren Spector, Co-President resigns.</td>
</tr>
<tr>
<td>20-Sep-2007</td>
<td>BSC reports 68% drop in quarterly income.</td>
</tr>
<tr>
<td>26-Sep-2007</td>
<td>Rumor that Warren Buffet may buy 20% stake in BSC.</td>
</tr>
<tr>
<td>22-Oct-2007</td>
<td>BSC announces deal with Citic.</td>
</tr>
<tr>
<td>01-Nov-2007</td>
<td>Wall Street Journal article about CEO Cayne accuses him of smoking marijuana.</td>
</tr>
<tr>
<td>14-Nov-2007</td>
<td>CFO Molinaro says BSC will write down $1.62 billion and take a 4th quarter loss.</td>
</tr>
<tr>
<td>28-Nov-2007</td>
<td>BSC lays off another 4% of its staff.</td>
</tr>
<tr>
<td>20-Dec-2007</td>
<td>BSC takes $1.9 billion write-down. Cayne says he will skip his 2007 bonus.</td>
</tr>
<tr>
<td>07-Jan-2008</td>
<td>CEO Cayne retires under pressure. Schwartz takes over.</td>
</tr>
<tr>
<td>22-Jan-2008</td>
<td>Fed cuts rates 75 bps.</td>
</tr>
<tr>
<td>14-Feb-2008</td>
<td>UBS writes down $2 bn in Alt-A which BSC was long (paired with subprime short).</td>
</tr>
<tr>
<td>03-Mar-2008</td>
<td>Thornburg Mortgage fails to meet margin calls.</td>
</tr>
<tr>
<td>10-Mar-2008</td>
<td>Rumors of BSC liquidity problems begin to surface.</td>
</tr>
<tr>
<td>11-Mar-2008</td>
<td>Goldman Sachs e-mails clients that it will not do derivative deals with BSC.</td>
</tr>
<tr>
<td>14-Mar-2008</td>
<td>BSC announces $30 billion in funding from JP Morgan (JPM), via the Federal Reserve.</td>
</tr>
<tr>
<td>17-Mar-2008</td>
<td>JPM announces acquisition of BSC for $2 a share.</td>
</tr>
<tr>
<td>24-Mar-2008</td>
<td>JPM raises bid for BSC to $10 a share.</td>
</tr>
</tbody>
</table>

In retrospect, October of 2007 looks like the eye of a hurricane. The stock market rallied, and the Dow Jones Index reached an all-time high of 14,164 on October 9, 2007. Bear Stearns’ swap spreads fell back to 70 bp.

Rumors about Bear Stearns’ subprime exposure and liquidity needs persisted though and concerns about the firm were raised again in November 2007 when a Wall Street Journal article portrayed James
Cayne as a distracted, drug-using CEO. Swap spreads did not fall back below 100 bp from that point on.

The market’s fears about Bear Stearns proved justified. The firm announced its first quarterly loss ever on November 14, 2007 and by the end of the month, swap spreads crossed 200 bp. Cayne resigned early in 2008, and despite further monetary injections by the Fed in February 2008, the upward trend in swap spreads remained intact.

To appreciate the velocity of Bear Stearns’ collapse though, you have to recall that spreads did not reach 300 bp until March 3 when the takeover, and eventually, J.P. Morgan agreed to raise the acquisition price to $10 a share. With that offer in place on March 24, 2008, swap spreads fell back below 150 bp.

2.3. Descriptive statistics

We begin to move to a more formal analysis by providing some descriptive statistics in Table 2.

The skewness is significantly positive at 2.41, and the kurtosis is 10.77, which is greater than 3, indicating fat tails in the data. The probability density function of the first difference of CDS spreads is plotted in Fig. 2.

Compared to the normal distribution, both the left tail and the right tail of CDS spreads are longer.

3. Models

The short-term interest rate process is a fundamental input into a variety of asset pricing models. Econometric modeling of the short-rate remains an active area of research. We append the GARCH model to Eq. (1). The CKLS–GARCH model is given by

\[ r_t = a + b_1 r_{t-1} + r_{t-1}^2 \epsilon_t, \]

\[ \epsilon_t = \sigma_t \epsilon_t, \]

where \( \epsilon_t \) is distributed as Normal.

3.1. CKLS

The paper by Chan et al. (1992) is the beginning of our search for an appropriate model. They proposed a differential equation, which we have discretized, that nust a number of popular models in the literature,

\[ r_t = a + b_1 r_{t-1} + r_{t-1}^2 \epsilon_t, \]

\[ \epsilon_t = \sigma_t \epsilon_t, \]

where \( \epsilon_t \) is distributed as Normal.

### Table 2

<table>
<thead>
<tr>
<th></th>
<th>Obs.</th>
<th>Mean</th>
<th>Std. dev.</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Descriptive statistics for Bear Stearns’ 5-year CDS spreads.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>501</td>
<td>82.57</td>
<td>96.20</td>
<td>19</td>
<td>660</td>
<td>2.41</td>
<td>10.77</td>
<td></td>
</tr>
</tbody>
</table>

Notes: There are 501 observations spanning April 3, 2006 to March 31, 2008.

Setting \( c = 1/2 \), you obtain the important Cox et al. (1985) model. Setting \( c = 0 \), which removes the effect from the level of the interest rate on volatility, you have the Ornstein–Unlenbeck process used by Vasicek (1977).

3.2. CKLS–GARCH model

In many financial time series, we observe unconditional fat tails and conditional volatility. Volatility clustering refers to the phenomenon that there are periods of high and low variances. That means large changes of variance tend to be followed by large changes, and small changes by small changes. Engle (1982) shows conditional heteroskedasticity may cause the fat tails in the unconditional distribution and suggests an ARCH model to capture the conditional heteroskedasticity. Bollerslev (1986) extends the ARCH model to a GARCH model, which is now the benchmark model for volatility clustering.

We append the GARCH model to Eq. (1). The CKLS–GARCH model is given by

\[ r_t = a + b_1 r_{t-1} + r_{t-1}^2 \epsilon_t, \]

\[ \epsilon_t = \sigma_t \epsilon_t, \]

where \( \epsilon_t \) follows the normal distribution. Research on CKLS–GARCH model can be found in Brenner et al. (1996), Koedijk et al. (1997), and Demirtas (2006). Brenner et al. (1996) show that the CKLS–GARCH model outperforms the CKLS model in maximum likelihood estimation of the 3-month Treasury bill.

3.3. CKLS–GARCH–EPD model

Many papers examining the time varying volatility in financial time series have questioned the conditional normality of the GARCH model. Typically, the standardized GARCH residuals \( \epsilon_t / \sigma_t \) are not normally distributed. Bollerslev (1987), for example, suggests a conditional \( \epsilon \)-distribution for \( \epsilon_t \). Haas et al. (2006) have also utilized the stable Paretian density.
In this paper, we propose to capture the outlying volatility shocks \( \varepsilon_t \) using an exponential power distribution (EPD) with the probability density function (PDF)

\[
f(\varepsilon_t) = \frac{1}{\lambda^{2^{1/\alpha}} \Gamma(1 + 1/\alpha)} \exp\left\{ -\frac{1}{2} \frac{|\varepsilon_t|^\alpha}{\lambda} \right\}
\]

(3)

where \( \lambda \) is a normalizing constant to make the variance of \( \varepsilon_t \) equal to unity:

\[
\lambda = \sqrt{\frac{2^{-2/\alpha} \Gamma(1/\alpha)}{\Gamma(3/\alpha)}}
\]

In Fig. 3, we plot a class of PDFs for the exponential power distribution. If parameter \( \alpha = 2 \), the exponential power distribution will be the Normal distribution. If parameter \( \alpha = 1 \), it will be the Laplace distribution, which has fatter tails than Normal distribution. As \( \alpha \) increases from 2 to 3, the PDF will become more platykurtic. As \( \alpha \) decreases from 2 to 1, the PDF will become more leptokurtic. The exponential power distribution in our model can be used to capture the jump component in the swap spread. For more information of PDFs for the exponential power distribution (EPD), one can refer to Nelson (1991) and Bali and Wu (2006).

The kurtosis of \( \varepsilon_t \) is defined as follows (see Nadarajah, 2005):

\[
\text{Kurtosis}(\varepsilon_t) = \frac{\Gamma(1/\alpha) \Gamma(5/\alpha)}{\Gamma^2(3/\alpha)}
\]

The CKLS–GARCH–EPD model is given by

\[
\begin{align*}
\varepsilon_t &= a + b_t r_{t-1} + \varepsilon_{t-1} \\
\sigma_t^2 &= \alpha_0 + \sum_{j=1}^{s} \alpha_j \varepsilon_{t-j}^2 + \sum_{j=1}^{s} \beta_j \sigma_{t-j}^2 \\
\alpha_0 &> 0, \quad \alpha_j \geq 0, \quad j = 1, \ldots, r, \quad \beta_j \geq 0, \quad j = 1, \ldots, s
\end{align*}
\]

(4)

where \( \varepsilon_t \) is drawn from Eq. (3).

4. Methodology

We estimate these models by the Bayesian method. Markov chain Monte Carlo algorithms are used. For example, the posterior PDF of the CKLS–GARCH–EPD model is given by

\[
p(\Delta | \text{data}) \propto p(\Delta) \prod_{t=1}^{n} \frac{r_{t-1}^c - r_{t-1} \sigma_{t-1}^{-2}}{2^{1-\alpha}} \exp\left\{ -\frac{1}{2} \frac{\hat{\varepsilon}_t^\alpha}{\lambda} \right\}
\]

(5)

where

\[
\Delta = \{a, b, c, \alpha_0, \Gamma, \Lambda, \alpha\} \\
\Gamma = (\alpha_1, \ldots, \alpha_r)^T \\
\Lambda = (\beta_1, \ldots, \beta_s)^T
\]

and

\[
\hat{\varepsilon}_t = \frac{r_t - a - b_t r_{t-1}}{\alpha_0 \sigma_{t-1}} , \quad t = 1, \ldots, n, \quad \hat{\alpha}_0 = - = \hat{\alpha}_{-q} = 0.
\]

As the prior, we set

\[
p(\Delta) = N(\pi, \sigma^2_\pi) \times N(\bar{\sigma}_a^2, \sigma^2_{\sigma_a}) \times N(\sigma_c, \sigma^2_{\sigma_c}) I(\alpha_0 > 0)
\]

\[
\times N(\Gamma, \sigma^2_{\Gamma}) I(\Gamma > 0) \times N(\Lambda, \sigma^2_{\Lambda}) I(\Lambda > 0) \times p(c) \times p(\alpha)
\]

where "*" denotes the prior parameters, \( p(c) = 1, p(\alpha) = 1, \) and \( I(\cdot) \) is an indicator function. We set all \( \sigma^2 \) sufficiently large and all the prior mean parameters zero. Similarly, we can derive the posterior densities for other models.

The MCMC algorithms for the CKLS–GARCH–EPD model include the following blocks (see Li et al. (2009)):

1. draw \( a, b, c \); 2. draw \( \alpha_0 \); 3. draw \( \Gamma \); 4. draw \( \Lambda \); 5. draw \( c \); 6. draw \( \alpha \).

The convergence of the MCMC algorithm is judged by the Kolmogorov–Smirnov tests (KST) that are explained in Goldman et al. (2008). The robustness of our MCMC algorithm is checked by changing both priors and initial values, respectively.

5. Results

5.1. CKLS

The CKLS model is estimated in the first column of Table 3.

We calculate the estimated spreads from the CKLS model as follows:

\[
\hat{r}_t = a + \hat{b}_t r_{t-1}
\]

and the standardized residuals are:

\[
\hat{\varepsilon}_t = \frac{r_t - \hat{r}_t}{\hat{\sigma}_t}
\]

The Lagrange multiplier (LM) test with null hypothesis of no ARCH effects is applied to the standardized residuals. The LM test statistic is 15.88 with a \( p \)-value of 0, which means we can reject the null hypothesis at the 5% significance level. Hence, we conclude that ARCH effects exist in the residuals of the CKLS model.
effects in the residuals of the CKLS

0.31 with a

and the standardized residuals are:

$r_t = \hat{\alpha} + \beta_1 r_{t-1}$

and the standardized residuals are:

$\hat{\epsilon}_t = \hat{\alpha}_t / \hat{\sigma}_t$.

where

$\hat{\sigma}_t^2 = \hat{\alpha}_0 + \sum_{j=1}^{s} \hat{\alpha}_j \hat{\epsilon}_{t-j}^2 + \sum_{j=1}^{s} \hat{\beta}_j \hat{\sigma}_{t-j}^2$  

$\hat{\epsilon}_{t-j} = \hat{\alpha}_t / \hat{\sigma}_t$

The LM test statistic calculated for the standardized residuals is 0.31 with a p-value of 58%. That means under 5% significance level we accept the null hypothesis and conclude that there are no ARCH effects in the residuals of the CKLS–GARCH model.

The estimates for parameter $\alpha_t$ and parameter $\beta_t$ are 0.7222 and 0.1673. The sum, $\alpha_t + \beta_t = 0.8895$, is well inside the integrated GARCH boundary.

We perform a Jarque–Bera test of the null hypothesis that the residuals come from a normal distribution. The p-value for the Jarque–Bera test is 0, so we conclude the standardized residuals are not normally distributed at the 5% significance level.

5.3. CKLS–GARCH–EPD

We allow the GARCH error to follow an EPD and report estimation results in the third column of Table 3. We get the estimated spread from the CKLS–GARCH model as follows:

$\hat{r}_t = \hat{\alpha} + \beta_1 r_{t-1}$

and the standardized residuals are:

$\hat{\epsilon}_t = \hat{\alpha}_t / \hat{\sigma}_t$.

We calculate the LM test for the standardized residuals. The LM test statistic is 1.97 with a p-value of 16%, so we can accept the null hypothesis at the 5% significance level. That is, there are no ARCH effects in the residuals of the CKLS–GARCH–EPD model.

We perform a Jarque–Bera test of the null hypothesis that the standardized residuals come from a normal distribution. The p-value for the Jarque–Bera test is 0, so we conclude the residuals are not normally distributed at the 5% significance level.

Next, we generate a sample from the EPD with $\hat{\alpha} = 0.2826$ and plot the empirical CDF in Fig. 4.

Compared to the empirical CDF of the standardized residuals, we can see they are very close to each other. For comparison, we also plot the CDF of standard normal. We conclude that the standardized residuals are closer to the EPD than to the normal.

The p-value of sign test for the standardized residuals and the sample of EPD with $\hat{\alpha} = 0.2826$ is 7.39%, higher than 5% significance level. So we conclude that there is no difference between the standardized residuals and the sample of EPD with $\hat{\alpha} = 0.2826$.

Following Bollerslev (1987), we compare the conditional kurtosis of EPD with that of the standardized residuals. The conditional kurtosis of EPD is calculated using the formula in Nadarajah (2005).

Kurtosis($\epsilon_t$) = \frac{\Gamma(1 / \hat{\alpha}) \Gamma(5 / \hat{\alpha})}{\Gamma^2(3 / \hat{\alpha})}.

The estimated conditional kurtosis of EPD is 234, which is very similar to the sample analogue of $(\hat{\epsilon}_t / \hat{\sigma}_t)^2$, $K = 279$. This result is similar to that of Bollerslev (1987).

The estimates for $\alpha_t$ and $\beta_t$ are 0.6778 and 0.1744. $\alpha_t + \beta_t = 0.8522$, which again lies comfortably inside the integrated GARCH frontier.

5.4. Model selection by Bayes Factor

The Bayes Factor is used as model selection criterion (see Appendix A). The result shows the CKLS model with GARCH–EPD error terms provides a better fit. The critical values of log Bayes Factor are reported in Table 4. For example, for the CKLS–GARCH–EPD model (denoted as $M_3$), the Bayes Factor of $M_3$ over $M_2$ is 3110, which is greater than 2. That means, the evidence supporting $M_3$ is decisive.
which provides significant evidence against the null hypothesis. If one is not highly convinced of the probability of the null hypothesis, if one is not highly convinced of the probability of the null hypothesis, then log10 BF12 is 0.5, then log10 BF12 = −0.5 since log10 BF12 = log10 (1/BF12).

<table>
<thead>
<tr>
<th>log10BF12</th>
<th>Evidence support M1</th>
<th>log10BF21</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 to 0.5</td>
<td>Weak</td>
<td>−0.5 to 0</td>
</tr>
<tr>
<td>0.5 to 1</td>
<td>Substantial</td>
<td>−1 to −0.5</td>
</tr>
<tr>
<td>1 to 2</td>
<td>Strong</td>
<td>−2 to −1</td>
</tr>
<tr>
<td>&gt;2</td>
<td>Decisive</td>
<td>≤−2</td>
</tr>
</tbody>
</table>

Notes: The values of log Bayes Factor can be negative. For example, if log10 BF12 = 0.5, then log10 BF21 = −0.5 since log10 BF21 = log10 (1/BF21).

We quantify the evidence of the Bayes Factor by the “energy” of changing the prior probability of the null hypothesis to a posterior probability of the null hypothesis. If one is not highly convinced of the probability of the null hypothesis, if one is not highly convinced of the probability of the null hypothesis, then log10 BF12 is 0.5, then log10 BF12 = −0.5 since log10 BF12 = log10 (1/BF12).

<table>
<thead>
<tr>
<th>Table 4</th>
<th>Critical values for log Bayes Factor.</th>
</tr>
</thead>
<tbody>
<tr>
<td>log10BF12</td>
<td>Evidence support M1</td>
</tr>
<tr>
<td>0 to 0.5</td>
<td>Weak</td>
</tr>
<tr>
<td>0.5 to 1</td>
<td>Substantial</td>
</tr>
<tr>
<td>1 to 2</td>
<td>Strong</td>
</tr>
<tr>
<td>&gt;2</td>
<td>Decisive</td>
</tr>
</tbody>
</table>

Notes: The values of log Bayes Factor can be negative. For example, if log10 BF12 = 0.5, then log10 BF21 = −0.5 since log10 BF21 = log10 (1/BF21).

Table 5
<table>
<thead>
<tr>
<th>Evidence for model</th>
<th>logBF1</th>
<th>Change in probability of the null hypothesis(model 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>From (%)</td>
<td>To (%)</td>
<td></td>
</tr>
<tr>
<td>Decisive 2</td>
<td>75</td>
<td>86</td>
</tr>
<tr>
<td>Decisive 3110</td>
<td>75</td>
<td>100*</td>
</tr>
<tr>
<td>Decisive 975</td>
<td>75</td>
<td>100</td>
</tr>
<tr>
<td>Decisive 3</td>
<td>3</td>
<td>99</td>
</tr>
<tr>
<td>Decisive 975</td>
<td>3</td>
<td>99</td>
</tr>
</tbody>
</table>

Notes: The table reports the posterior probability of the null hypothesis after observing the Bayes Factors with given prior probability. Prior odds = Bayes Factor > Prior odds. Prior probability = odds/(1+ odds). For example, the number marked by an asterisk in Table 5 is calculated as follows: Prior odds = Bayes Factor × Prior odds = BF12 × 75/(1 − 75%) = 9330. Prior probability = odds/(1 + odds) = 9330/(1 + 9330) = 100x.

Table 6
<table>
<thead>
<tr>
<th>Posterior means and t-statistics for the subsample of CDS spreads.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Models</td>
</tr>
<tr>
<td>--------</td>
</tr>
<tr>
<td>a</td>
</tr>
<tr>
<td>b</td>
</tr>
<tr>
<td>c</td>
</tr>
<tr>
<td>d</td>
</tr>
<tr>
<td>e</td>
</tr>
<tr>
<td>f</td>
</tr>
<tr>
<td>g</td>
</tr>
<tr>
<td>h</td>
</tr>
<tr>
<td>i</td>
</tr>
<tr>
<td>j</td>
</tr>
<tr>
<td>k</td>
</tr>
<tr>
<td>l</td>
</tr>
<tr>
<td>m</td>
</tr>
<tr>
<td>n</td>
</tr>
</tbody>
</table>

Notes: We analyze the spread models over the period April 3, 2006 to October 31, 2007, just prior to the structural break identified by the Zivot–Andrews test. M1 is the CKLS model in Eq. 1. M2 is the CKLS-GARCH model in Eq. (2). M3 is the CKLS-GARCH-EPD model in Eq. (4).
model selection criterion. We find that the CKLS model with GARCH volatility and exponential power distribution errors provides the best fit. This establishes that level effects, volatility clustering and jumps are statistically significant components of CDS spreads.

Our analysis also documents the calm before the storm. As late as October 2007, when policy makers and industry participants were assuring us that the subprime crisis was contained, Bear Stearns’ CDS spreads were evolving smoothly like most other investment grade debt. While Bear Stearns’ March 2008 collapse was ultimately less leveraged institution, events would eventually spin out of control just six months later.

Acknowledgment

We would like to thank Shiliang Li for assistance with the MCMC algorithms.

Appendix A. Bayes Factor

Appendix A.1. Bayes Factor, posterior odds ratio and likelihood ratio

Given data Y, we want to compare models M₁ and M₂, which are with parameter sets Θ₁ and Θ₂, respectively. The likelihood function for model M₁ is \( p(Y|\Theta_1, M_1) \).

The marginal likelihood function for model M₀ or the integrated likelihood or the evidence for model M₁ is:

\[
p(Y|\Theta_1, M_1) = \int p(Y, \Theta_1 | M_1)d\Theta_1 = \int p(Y|\Theta_1, M_1)p(\Theta_1|M_1)d\Theta_1.
\]

The last equation is obtained by applying Bayes’ theorem.

The posterior probability of model i given data Y, \( p(M_i|Y) \), can be derived by Bayes’ theorem,

\[
p(M_i|Y) = \frac{p(M_i | Y)}{p(Y)} = \frac{p(Y | M_i)p(M_i)}{p(Y)}.
\]

Hence, the posterior odds ratio for model M₁ against model M₂ is:

\[
\frac{p(M_1 | Y)}{p(M_2 | Y)} = \frac{p(M_1 | Y)}{p(Y | M_2)} \times \frac{p(M_1)}{p(M_2)} = \frac{p(Y | M_1)p(M_1)}{p(Y|M_2)p(M_2)}.
\]

i.e.,

\[
\text{posterior odds ratio} = \text{Bayes Factor} \times \text{prior odds ratio}.
\]

Solving for the Bayes Factor, you find

\[
P(Y|M_1) = \frac{p(Y|M_1)p(M_1)}{p(Y | M_2)} = \frac{p(M_1 | Y)}{p(M_1 | Y)}.
\]

The Bayes Factor differs from the posterior odds ratio by eliminating the effect of priors.

In general, prior odds ratio is set to be 1 i.e., \( p(M_1) = p(M_2) = 0.5 \). In this case, Bayes Factor equals the posterior odds ratio. That’s the reason why sometimes these two terminologies can be inter-exchangeable. Further, if these two models are assumed with no parameters, then there will be no integration with respect to parameters. In this case, the Bayes Factor is just the likelihood ratio.³

Appendix A.2. Model selection via Bayes Factor

In a model selection problem, we have to choose between M₁ and M₂ on the basis of the data Y. In theory, Bayes Factor BF₁₂ is given as follows:

\[
BF_{12} = \frac{P(Y|\Theta, M_1)}{P(Y|\Theta, M_2)} = \frac{\int p(Y, \Theta_1 | M_1)d\Theta_1}{\int p(Y, \Theta_2 | M_2)d\Theta_2} = \frac{\int p(Y|\Theta_1, M_1)p(\Theta_1|M_1)d\Theta_1}{\int p(Y|\Theta_2, M_2)p(\Theta_2|M_2)d\Theta_2}.
\]

where \( p(Y|M_i) \) is the marginal likelihood for model i, \( p(Y|\Theta, M_i) \) is the likelihood for model i. And \( \Theta \) is the parameter set in model i.

To calculate the Bayes Factor from the MCMC algorithms, we use following method (Kass and Raftery (1995), p.779),

\[
BF_{12} = \frac{\int p(Y|\Theta_1, M_1)p(\Theta_1|M_1)d\Theta_1}{\int p(Y|\Theta_2, M_2)p(\Theta_2|M_2)d\Theta_2} \approx \frac{1}{n} \sum_{i=1}^{n} p(Y|\Theta_1^{(i)}, M_1) - 1\sum_{i=1}^{n} p(Y|\Theta_2^{(i)}, M_2)
\]

where j is the jth draw of parameter \( \Theta \) from the MCMC algorithms. \( n \) is the number of the accepted draws in the MCMC algorithms.

The critical regions in Table 4 are from page 777 of Kass and Raftery (1995). For example, if the value of \( \log_{10} BF_{12} \) falls into the interval of [0.5, 1], we conclude that there is substantial evidence supporting model 1.

Followed Goodman (1999), we quantify the evidence of Bayes Factor by the “energy” of changing the prior probability of the null hypothesis to a posterior probability of the null hypothesis.

For instance, if one is highly convinced of model 1 (75% prior probability of the null hypothesis) before analyzing the data, a Bayes Factor of 0.001 will convince that the null hypothesis is not true (3% posterior probability of the null hypothesis). In other words, to achieve 5% posterior probability of the null hypothesis with a Bayes Factor of 0.001, one needs to have an 85% prior probability of the null hypothesis.

A Bayes Factor of 2 is strong enough to move one from being 75% sure of the null hypothesis to being 85% sure (see Table 5).

References


