

2019

# Probability of Default White Paper

The Credit Research Initiative (CRI)  
National University of Singapore

*First version: March 2, 2017, this version: January 8, 2019*

## ABSTRACT

Probability of Default (PD) is the core credit product of the Credit Research Initiative (CRI). The CRI system is built on the forward intensity model developed by Duan *et al.* (2012, *Journal of Econometrics*). This white paper describes the fundamental principles and the implementation of the model. Details of the theoretical foundations and numerical realization are presented in [RMI-CRI Technical Report \(Version 2017\)](#). This white paper contains five sections. Among them, Sections II & III describe the methodology and performance of the model respectively, and Section IV relates to the examples of how CRI PD can be used.

## CONTENT

I. OVERVIEW .....	2
II. METHODOLOGY .....	3
III. MODEL PERFORMANCE .....	8
IV. APPLICATIONS .....	11
V. CONCLUSION .....	14
ABOUT THE CREDIT RESEARCH INITIATIVE .....	15

---

\* Please cite this document in the following way: “The Credit Research Initiative of the National University of Singapore (2019), Probability of Default (PD) White Paper”, Accessible via [https://www.rmici.org/en/white\\_paper/](https://www.rmici.org/en/white_paper/).

## I. OVERVIEW

Probability of Default (PD) is the core credit measure of the CRI corporate default prediction system built on the forward intensity model of Duan *et al.* (2012)<sup>1</sup>. This forward intensity model is governed by two independent doubly stochastic Poisson processes, operating on forward time instead of spot time. This enables the model to produce forward-looking PD-term structures of public firms based on dynamic learning from the macrofinancial and firm-specific data.

The key features of the CRI model are:

- Combines a reduced-form model (based on a forward intensity construction) and a structural model (using Distance-To-Default as one of its input covariates)
- Accommodates the two risks that a listed firm might encounter; namely default risk and risk of other types of corporate exits (i.e. mergers and acquisitions)
- Uses forward probabilities of default and other types of exits as building blocks to construct the PD term structure in a consistent manner
- Employs sixteen input covariates (or default predictors) from both market-based and accounting-based firm-specific attributes, as well as macro-financial factors (except for China which have fourteen covariates, for more information please refer to [Addendum 9 to the CRI Technical Report \(2017\) Update 1](#))

### Coverage

In July 2010, the CRI began to release daily updated PD of around 17,000 public firms in 12 Asian economies. As of January 2018, the coverage of the CRI has expanded to over 67,000 exchange-listed firms in 128 economies worldwide with prediction horizons from 1 month to 5 years. Out of those firms, over 34,000 are currently actively listed and have their PD updated on a daily basis. Furthermore, historical PD series are refreshed on a yearly basis as part of the CRI commitment to scientific pursuit and to account for retroactive information.

---

<sup>1</sup> Duan, J. C., Sun, J., and Wang, T. (2012). "Multiperiod Corporate Default Prediction – A Forward Intensity Approach", *Journal of Econometrics*, 179, pages 191-209.

## II. METHODOLOGY

### Model Description

The building block of the CRI corporate default prediction model is the conditional forward probability. As Figure 1 illustrates, when firm  $i$  is at time  $t$  facing the future,  $p_{i,t}(3)$  is the probability that the firm defaults in the fourth month, **conditional** on its survival up to the third month.

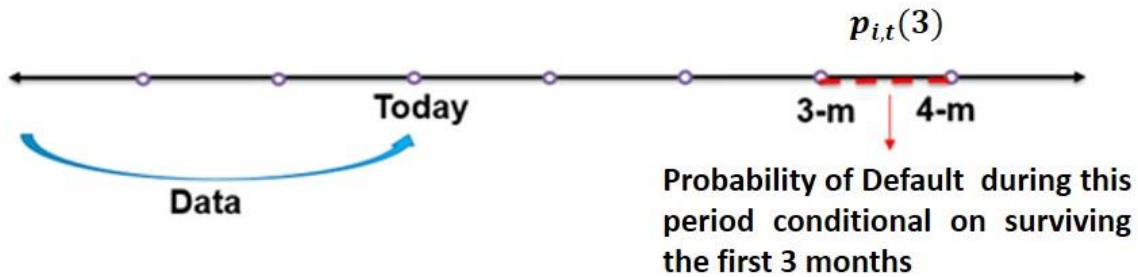


Fig 1. Forward probability in the CRI model

Formally, for each forward period  $\tau$ ,  $p_{i,t}(\tau)$  is constructed on a forward intensity function, whose inputs include the state of the economy (macrofinancial risk factors  $X_t$ ) and the vulnerability of individual obligors (firm-specific attributes  $Y_{i,t}$ ):

$$p_{i,t}(\tau) = P_{\tau}(X_t, Y_{i,t})$$

With  $p_{i,t}(\tau)$  in place, the multi-period default probabilities with different term structures can be obtained through the typical survival-exit formula. The underlying forward intensity functions are parameterized, and the parameters are estimated for each calibration groups on a monthly basis as new information accumulated in the CRI database.

## Input Covariates

Following the notation above, firm  $i$ 's input covariates at time  $t$  are represented by 1) the vector  $X_t$  that is common to all firms in the same economy<sup>2</sup>, and 2) a firm-specific vector  $Y_t$  with components constructed from the firm's financial statements and market capitalizations. The CRI corporate default prediction model employs four macrofinancial variables and twelve firm-specific variables, as presented in Table 1 below.

**Table 1. Input covariates for the CRI PD model**

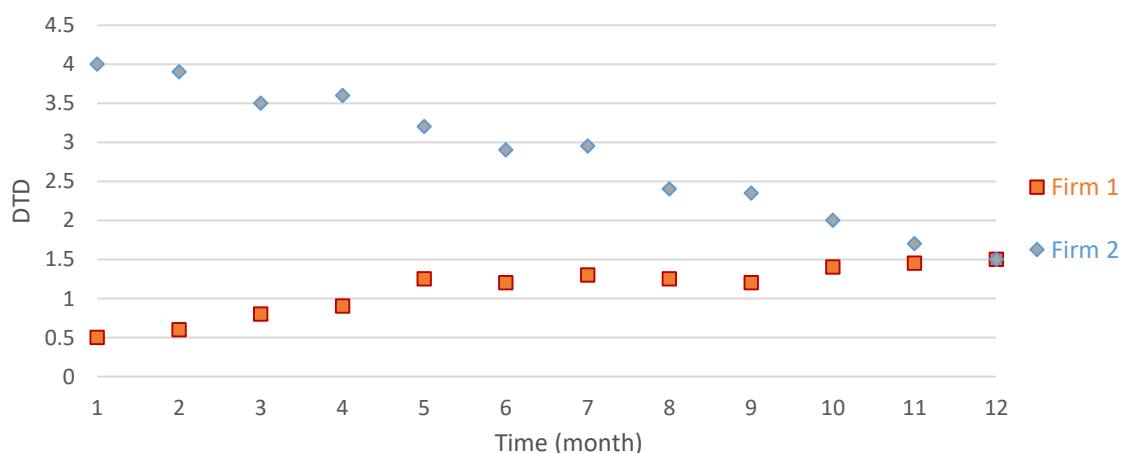
	Model Inputs	Description
Macro-Financial Factors	Stock Index Return	Trailing 1-year return of the prime stock market, winsorization and currency adjusted
	Short-term Risk-Free Rate	Yield on 3-month government bills
	Economy-level Distance-To-Default for financial firms	Median Distance-to-Default of financial/non-financial firms in each economy/country inclusive of those foreign firms whose primary stock exchange is in this economy/country (Not applicable to China)
	Economy-level Distance-To-Default for non-financial firms	
Firm-Specific Attributes	Distance-to-Default (level)	Volatility-adjusted leverage based on Merton (1974) with <b>special</b> treatments
	Distance-to-Default (trend)	
	Cash/Total Assets (level)	For financial firm's liquidity - Logarithm of the ratio of each firm's sum of cash and short-term investments to total assets
	Cash/Total Assets (trend)	
	Current Assets/Current Liabilities (level)	For non-financial firm's liquidity - Logarithm of the ratio of each firm's current assets to current liabilities
	Current Assets/Current Liabilities (trend)	
	Net Income/Total Assets (level)	Profitability - Ratio of each firm's net income to total assets
	Net Income/Total Assets (trend)	
	Relative Size (level)	Logarithm of the ratio of each firm's market capitalization to the economy's median market capitalization over the past one year
	Relative Size (trend)	
	Relative Market-to-Book Ratio	Individual firm's market misvaluation/ future growth opportunities relative to the economy's median level of market-to-book ratio
	Idiosyncratic Volatility	1-year idiosyncratic volatility of each firm, computed as the standard deviation of its residuals using the market model

<sup>2</sup> Firms which are listed on the stock exchanges of that economy.

In the table above, “*level*” is computed as the 12-month moving average (a minimum of six observations in the 12-month range are required, otherwise level variables will bear missing values), and “*trend*” is computed as the current value minus the “*level*” value (if the current month value is missing, the trend variable is set to be the last valid value in the previous month).

The “*trend*” measure captures the momentum effect and gives a hint about the direction of future movements. Duan *et al.* (2012) show that using the level and trend of the measures for some input covariates significantly improves the predictive power of the model, particularly for shorter horizons.

In order to understand the momentum effect, consider the case of two firms that have the same current value of Distance-To-Default (DTD). Firm 1 reaches its current value of DTD from a lower level, while Firm 2 reaches the same current value of DTD as Firm 1 but from a higher level, as shown in Figure 2. If only the current value of the DTD is employed for default prediction, the impact of the DTD on the PD would be identical for both firms. Intuitively however, one would expect the DTD of Firm 1 to have an upward momentum whereas that of Firm 2 to continue its decline. In order to account for such momentum effects, the CRI uses both level and trend attributes in the computations of its PD.



**Fig 2. DTD momentum effect**

Firms with lower default risk will have higher DTD.

DTD has long been recognized as an important indicator of a firm's credit quality, and is employed by the CRI as a default predictor in the forward intensity model.

Typically, the DTD for each firm is estimated using Merton's<sup>3</sup> structural model with the same assumptions on the debt maturity and size as in the KMV implementation, i.e.,

$$DTD_t = \frac{\log\left(\frac{V_t}{L}\right) + \left(\mu - \frac{\sigma^2}{2}\right)(T - t)}{\sigma\sqrt{T - t}}$$

where  $V_t$  is the asset value following a geometric Brownian motion with drift  $\mu$  and volatility  $\sigma$ ,  $L$  is the default point with value equal to short-term liabilities plus half of long-term liabilities, and  $\sqrt{T - t}$  is set to 1 year.

However, to improve the traditional DTD measure, the CRI implements some special treatments on its own DTD calculation to overcome several drawbacks that have been identified in the literature.

The key treatments are:

- Follow Duan (2010)<sup>4</sup> to add a fraction ( $\delta$ ) of other liability to the KMV default point  $L$
- Set  $\mu = \frac{\sigma^2}{2}$  to improve the stability of DTD estimation.
- Standardize the firm's market value by its book value to handle the scale change due to any major investment and financing action

The parameters required for DTD estimation are estimated by the maximum likelihood method described in Duan (1994<sup>5</sup>, 2000<sup>6</sup>)

<sup>3</sup> Merton, R. C. (1974). "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates". *The Journal of Finance*, 29(2), 449-470.

<sup>4</sup> Duan, J. C. (2010). "Clustered Defaults". *Risk Management Institute Working Paper*.

<sup>5</sup> Duan, J. C. (1994). "Maximum Likelihood Estimation Using Price Data Of The Derivative Contract". *Mathematical Finance*, 4(2), 155-167.

<sup>6</sup> Duan, J. C. (2000). "Correction: Maximum Likelihood Estimation Using Price Data of the Derivative Contract". *Mathematical Finance*, 10(4), 461-462.

A brief expression of the CRI's version of DTD can be rewritten as:

$$\text{DTD}_t = \frac{\log\left(\frac{V_t}{L}\right)}{\sigma\sqrt{T-t}}$$

where the default point is set to  $L = \text{Current Liabilities} + \frac{1}{2} \text{Longterm Liabilities} + \delta \times \text{Other Liabilities}$ , and  $\delta \in [0,1]$  is shared by firms in the same sector (financial vs. non-financial) in each calibration group.

## Calibration

Currently, the data for the CRI corporate default prediction system comes from various international data distributors (e.g. Thomson Reuters Datastream, Bloomberg Backoffice License etc.). It is worth noting that there are few or no credit default events in certain economies due to limited number of listed firms in those countries, which means that calibrating models for individual economy would not be statistically meaningful. In view of this, public companies around the world are segregated into six calibration groups according to certain similarities in the stage of economic development and geographic locations of their primary exchanges. These calibration groups are North America, Europe, Asia-developed economies, Emerging Markets, China, and India.

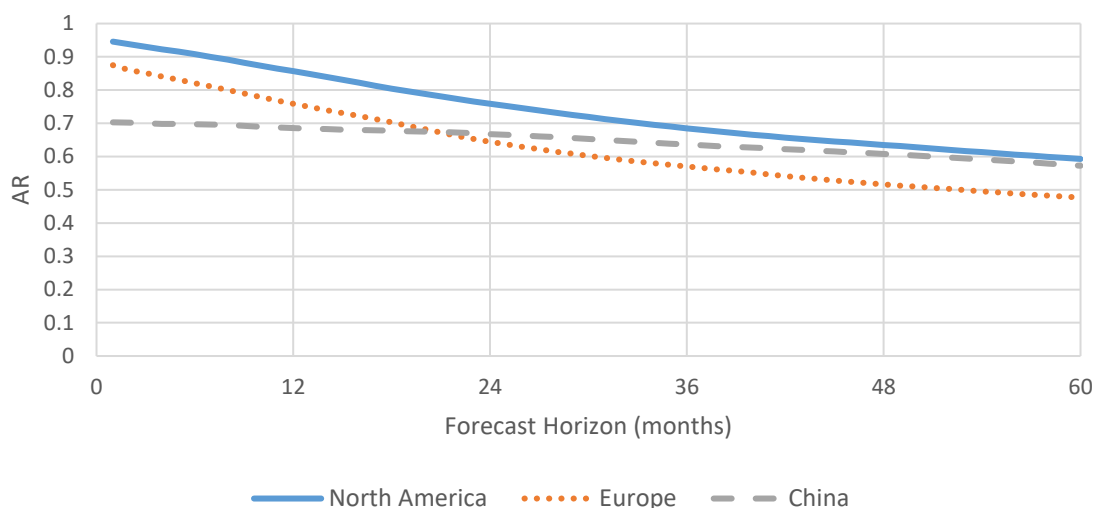
The CRI PD measures of companies within the same calibration group share the same set of parameters, (except for some covariates in some special circumstances). To overcome optimization difficulties caused by high dimensionality of the parameters (i.e. 16 covariates in general and 14 covariates in the case of China for 60 monthly prediction horizons), the CRI system employs the Nielson-Siegel term structure function and relies on sequential Monte Carlo for the model's estimation. Details of the procedure can be found in the [RMI-CRI Technical Report \(Version 2017\)](#).



### III. MODEL PERFORMANCE

The Accuracy Ratio (AR) is a popular quantitative measure for evaluating the discriminatory power of a default prediction system. It is the ratio of (a) the differential of the performance of the evaluated system and the random system over (b) the differential of the performance of the perfect system and the random system. A totally non-informative model will yield an AR of zero. The interpretation of AR is that if defaulted firms have been assigned among the highest credit risks before they defaulted, then the model has discriminated properly between the safe and risky firms. The CRI corporate default prediction model achieves high AR scores for all its covered regions and economies, indicating its good performance.

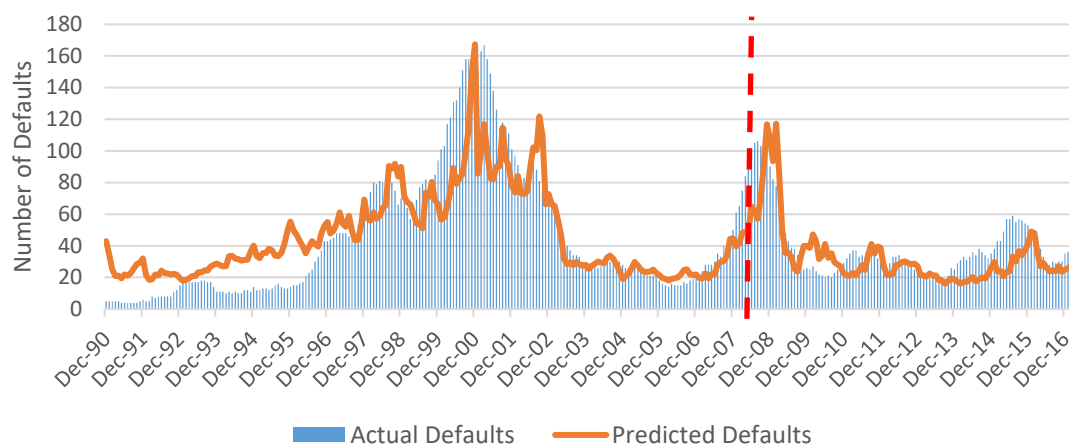
Figure 3 illustrates the AR of the CRI corporate default prediction model for North America, Europe and China for horizons from 1 month to 5 years.



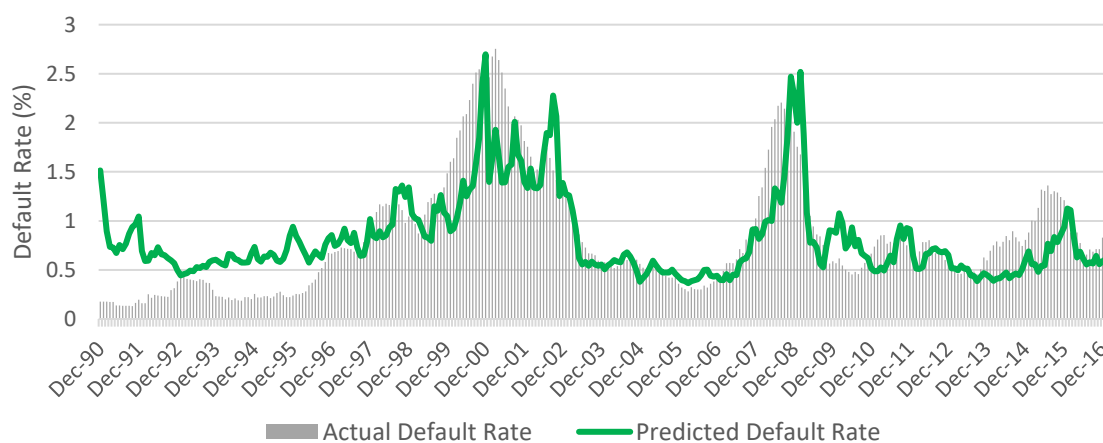
**Fig 3. Accuracy Ratio of the CRI PD model**

As of March 2018.

A more straightforward alternative to AR for judging the performance of a default prediction model is comparing the actual number of realized defaults against the number of defaults predicted by the model. The followings figures 4a to 4d present such comparison using the CRI PD model and realized defaults within one year in the US and China on a monthly frequency.

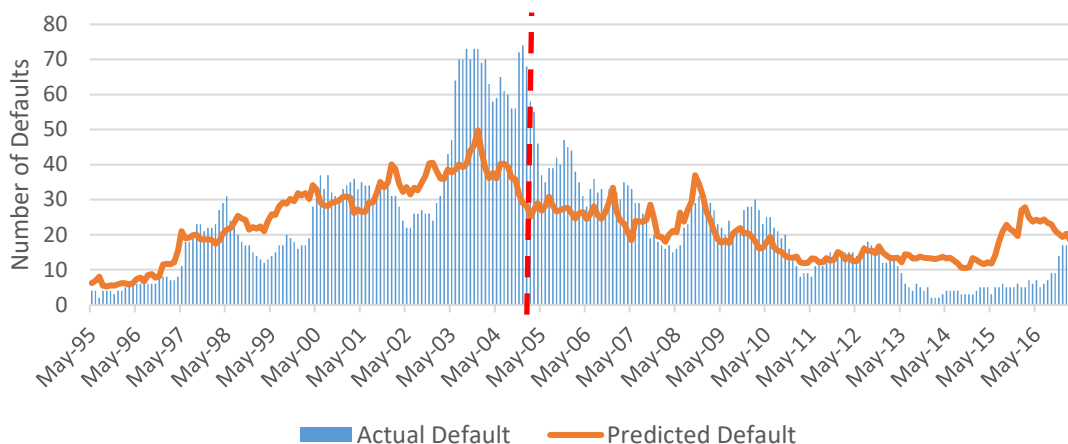


**Fig 4a. Realized vs. predicted number of defaults within 1 year for the United States**  
With structural break in September 2008 (dotted red line). Source: CRI, May 2018.



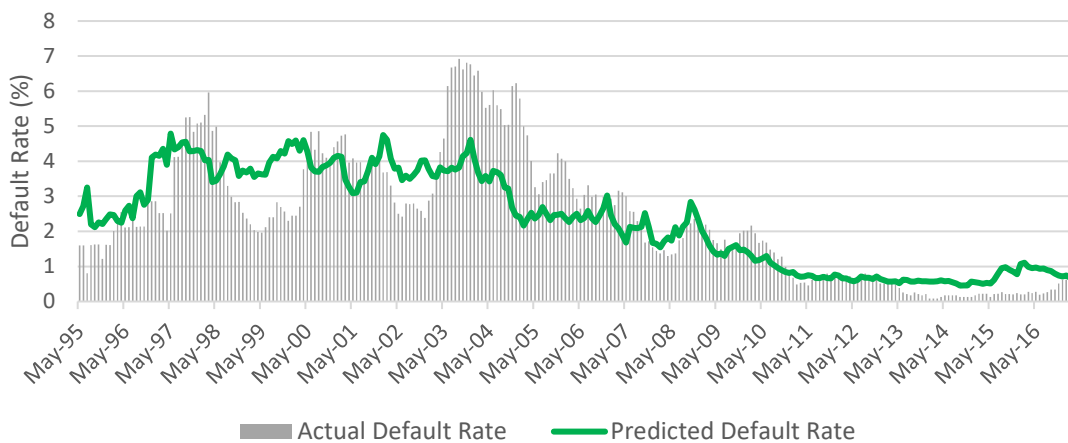
**Fig 4b. Realized vs. predicted default rate within 1 year for the United States**  
Source: CRI, May 2018.

For the North American calibration group, the CRI PD model introduces a dummy variable on the intercept for financial firms to account for differences that have not been reflected through the other covariates. A structural break in the financial-sector intercept dummy variable has also been introduced to address the dramatic changes after Lehman Brothers' default in September 2008. Figures 4a through 4d show that the CRI PD model performs well in predicting default numbers and rates for US firms.



**Fig 4c. Realized vs. predicted number of defaults within 1 year for China**

With structural break in December 2004 (dotted red line). Source: CRI, May 2018.



**Fig 4d. Realized vs. predicted default rate within 1 year for China**

Source: CRI, May 2018.

For China, the CRI PD model also accounts for a structural break, occurring in December 2004, to reflect the significant drop in default rates after that point. By simply changing the behavior of two parameters (i.e., coefficients for the intercept and DTD level) after December 2004, the performance of the CRI PD model for Chinese firms has significantly improved.

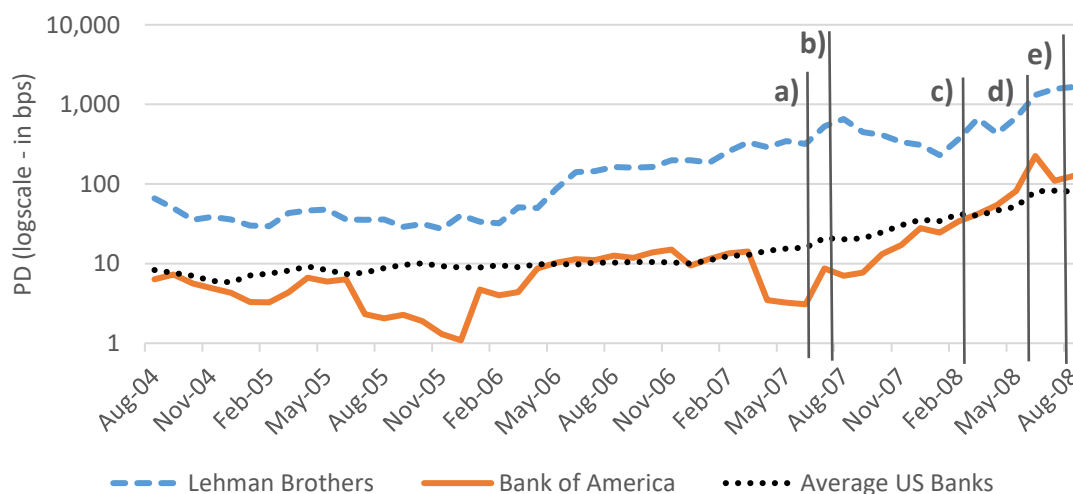
For more information, please refer to the [Addendum 9 to the CRI Technical Report \(2017\) Update 1](#)

## IV. APPLICATIONS

### Predicting Default of a Firm

Lehman Brothers was the fourth-largest investment bank in the US at the time of its collapse. In 2006, this bank securitized \$146 billion of mortgages or a 10% increase from the previous year. The bank in effect had shifted its business model from an investment bank to a real estate hedge fund. The US subprime mortgage crisis erupted in Q1 2007 when the number of defaults on mortgages underlining those mortgage backed securities surged to a seven-year high. Heavily relying on mortgage securitization and sale, Lehman Brothers reported substantial losses in Q1 and Q2 2008 and eventually filed for Chapter 11 bankruptcy protection on September 15, 2008.

Figure 5 presents the evolution of Lehman Brothers' one-year CRI PD two years before the firm filed for bankruptcy. The CRI PD for Bank of America and the average CRI PD for American banks have been added to this graph for comparison purposes.

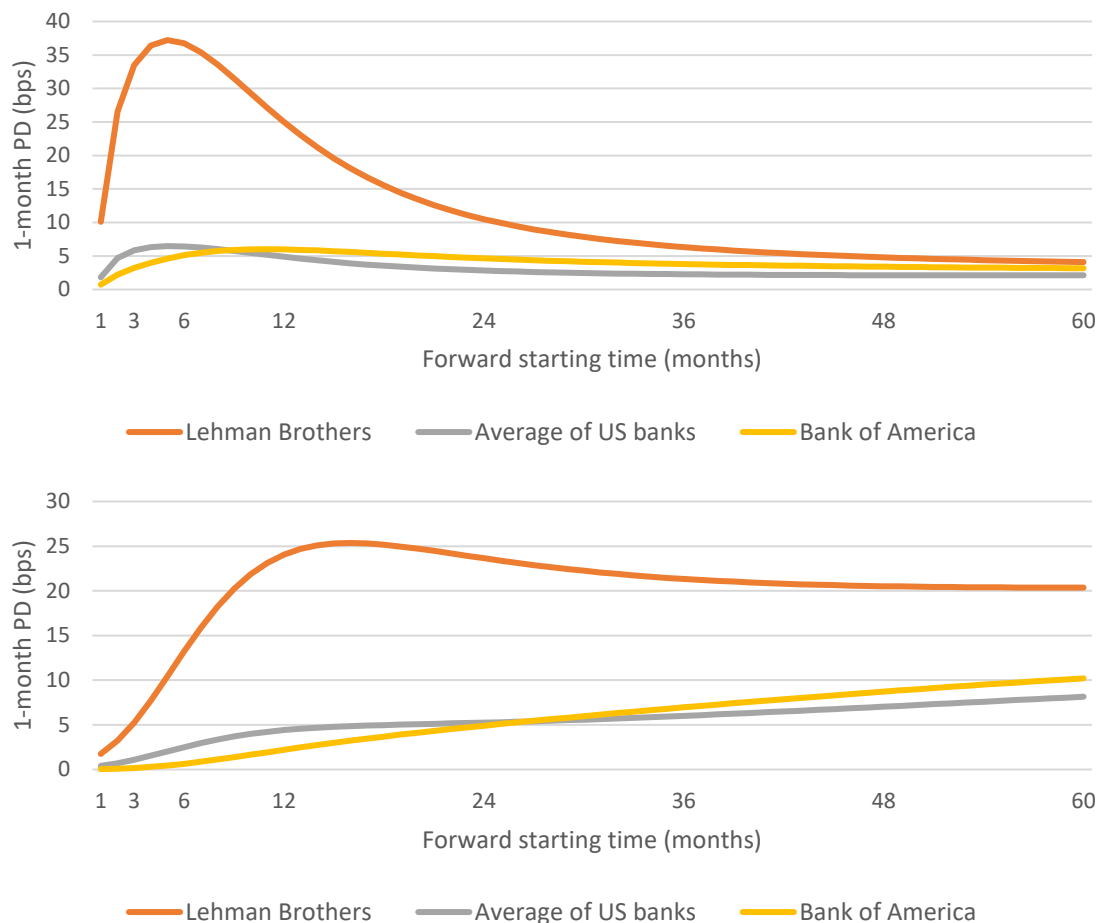


**Fig 5. Historical time series of 1-year CRI PD for Lehman Brothers, Bank of America, and the average of US banks**

4 year before Lehman Brothers bankruptcy (August 2004 to August 2008). Source: CRI 2018. Key events:

- a) **July 2007:** Collapse of two subprime Bear Stearns hedge funds
- b) **August 2007:** Lehman quarterly filings reveal \$79.6 billion of mortgage exposure, major CRA cut ratings
- c) **March 2008:** Demise of Bear Stearns due to the subprime mortgage crisis in the US
- d) **June 2008:** Lehman Brothers announced a loss of \$2.8 billion
- e) **August 2008:** Lehman Brothers announced a loss of \$3.9 billion, Lehman Brothers filed for Chapter 11

Figure 6 below shows the risk profile of Lehman Brothers compared to the profiles of Bank of America and the US bank average using the term structure of forward-looking one-month PD. Lehman Brothers' credit worthiness on a forward-looking basis has consistently been below the US bank average whether 3 months or 24 months preceding its demise.



**Fig 6. Risk profiles of major US banks 3 months (top) and 24 months (bottom) before Lehman Brothers' bankruptcy (September 2008)**

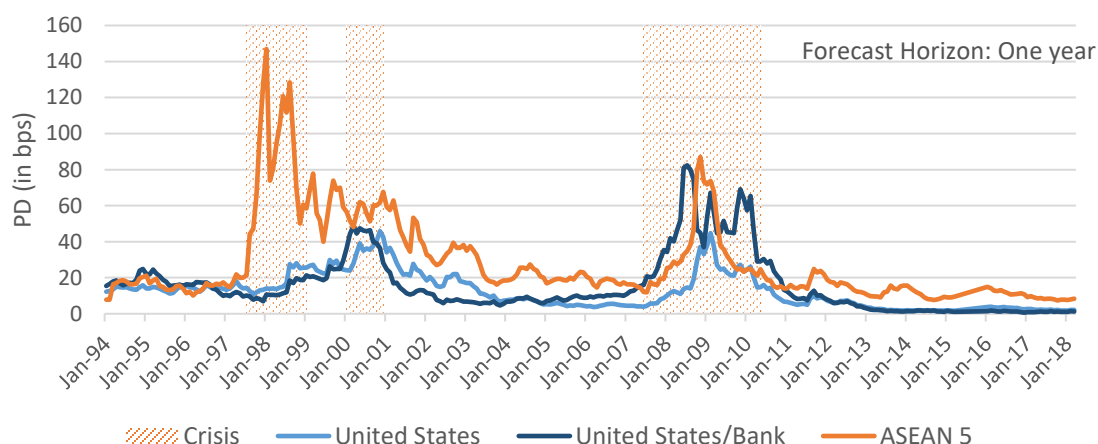
(top) Parameters calibrated with data up to June 2008

(bottom) Parameters calibrated with data up to September 2006.

Source: CRI, May 2018.

## Assessing the Aggregate Credit Risk of a Portfolio

Because the CRI computes PD on an individual firm-level basis, the CRI PD of all firms within a specific region and/or sector can easily be aggregated<sup>7</sup> to deliver an overview of the credit environment of that portfolio at a certain point in time. Figure 7 depicts the aggregate (median) CRI 1-year PD for the US, the banking sector in the US, and the ASEAN 5 (Malaysia, Thailand, the Philippines, Indonesia and Singapore).



**Fig 7. Historical time series of aggregate 1-year CRI-PD**

Median CRI PD for three selected groups. Source: CRI, May 2018.

The aggregate CRI PD well reflects the known financial crises through rises in PDs in time of crisis. For instance, the 1997 Asian financial crisis particularly affected the credit environment of the ASEAN 5 countries, whereas the internet bubble in 2000-2001 caused wide spread defaults among US firms and the global financial crisis of 2008-09 severely impacted the US banking sector first before spreading to the rest of the world including the ASEAN 5 countries.

## Other Examples of Application

<sup>7</sup> The aggregate CRI PD is a simple median of individual PDs across all firms within a region and/or sector. The domicile location of a firm follows the country of its headquarters defined by our data provider. Dual-listed companies (for example, Rio Tinto) exist as a single corporation but retain two different legal identities. They may have two different sets of PDs, due to two exchange listings for separate entities but sharing the same domicile. In such cases, we will override the entity's default domicile country to follow its stock exchange's location.

Aside from assessing the credit risk of a single corporate, and aggregate credit risk of a group or portfolio, CRI PD can be used for a variety of credit management related applications; from benchmarking, to fixed income investment, and credit portfolio management, from model validation to credit research reports.

## V. CONCLUSION

The CRI PD estimates the default risk of publicly listed firms by quantitatively analyzing their financial statements, stock market data and macro-financial factors retrieved from various international data sources. Unlike credit models that utilize letter-grade ratings, the CRI PD is a more granular gauge for credit risk and is available in term structure ranging from 1 month to 5 years. The CRI PD also captures default correlations and can be further aggregated to reflect credit cycles, among others.

The CRI currently provides daily updated PD of over 34,000 active exchange-listed firms globally. The CRI also distributes historical time series of PD of over 33,000 inactive firms due to bankruptcy, corporate consolidation or delisting for other reasons.

## ABOUT THE CREDIT RESEARCH INITIATIVE

The Credit Research Initiative (CRI) was launched by Professor Jin-Chuan Duan in July 2009 at the Risk Management Institute of the National University of Singapore. Aiming at “Transforming Big Data into Smart Data”, the CRI covers over 67,000 public firms and produces daily updated Probabilities of Default (1-month to 5-year horizon), Actuarial Spreads (1-year to 5-year contract) and Probability of Default implied Ratings on over 34,000 currently active, exchange-listed firms in 128 economies. The CRI also distributes historical time series of over 33,000 inactive firms due to bankruptcy, corporate consolidation or delisting for other reasons. In addition, the CRI produces and maintains Corporate Vulnerability Indices (CVI), which can be viewed as stress indicators, measuring credit risk in economies, regions and special portfolios.

As a further step, the CRI converts smart data to actionable data to meet the customized demands of its users and offers bespoke credit risk solutions leveraging on its expertise in credit risk analytics. A concrete example is our development of the BuDA (Bottom-up Default Analysis) toolkit in collaboration with the IMF. BuDA is an automated analytic tool based on the CRI PD system, enabling IMF economists to conduct scenarios analyses for the macro-financial linkage.

The CRI publishes Weekly Credit Brief and Quarterly Credit Report, highlighting key credit-related events, offering insights based on the CRI PD of the entities involved, and providing useful statistics on credit risk of economies and sectors.



**©2019 NUS Risk Management Institute (RMI). All Rights Reserved.**

The content in this white paper is for information purposes only. This information is, to the best of our knowledge, accurate and reliable as per the date indicated in the paper and NUS Risk Management Institute (RMI) makes no warranty of any kind, express or implied as to its completeness or accuracy. Opinions and estimates constitute our judgment and are subject to change without notice.

NUS Risk Management Institute (RMI) – Credit Research Initiative  
Address: 21 Heng Mui Keng Terrace, I<sup>3</sup> Building, Level 4, Singapore 119613  
Tel: (65) 6516 3380 Fax: (65) 6874 5430  
Website: <http://rmicri.org/>