Bottom-up Default Analysis （BuDA） User Manual

A credit stress testing and scenario analysis toolkit jointly developed by the IMF and the Credit Research Initiative (CRI) team of National University of Singapore (NUS), and operationally linked to the CRI platform.

BuDA was developed by Jin-Chuan Duan and Weimin Miao of National University of Singapore (NUS) in collaboration with Jorge Chan-Lau of International Monetary Fund with active support by NUS Risk Management Institute’s Credit Research Initiative team.

Preface

There is a growing demand for practical tools and models useful for analysing the dynamics of credit risk under macro-financial scenarios. The Bottom-Up Default Analysis (BuDA) approach was conceived to meet this demand, by providing an easy-to-use interactive platform for analysing the credit risk of individual firms, sectors, or portfolios under different scenarios. With BuDA, the user can minimize time spent on calculations, and rather, use the time more wisely on scenario design and risk analysis.

This manual provides the conceptual background underlying BuDA. The aim is that readers will be able to not only operate the BuDA toolkit, implemented in Matlab, but also develop a basic understanding of the inner workings of the scenario analysis approach and underlying system and methods which produce the probabilities of default (PDs). In particular, this approach is useful for analysts interested in conducting stress testing and scenario analysis and, more generally, examining macro-financial risks.

The manual targets a general audience and assumes minimum background knowledge on credit risk modelling. That said, a basic level of competency in econometrics may help the reader to capture the economic intuition underlying the models and methods described herein. While the content may be read on a stand-alone basis, the manual is best used as the background material for hands-on courses on applied credit risk modelling and macro-financial risk analysis.

About the Credit Research Initiative at NUS-RMI

The Credit Research Initiative (CRI) is a non-profit undertaking at the Risk Management Institute (RMI) of the National University of Singapore (NUS), which seeks to promote research and development in the critical area of credit risk.

The PDs produced by CRI arose to fulfil the need for credit research in a setting where it does not conflict with profit incentives. CRI is a non-profit public good initiative with the goal of keeping the model current, evolutionary, and organic, and functions like a "selective Wikipedia".

RMI announced the CRI in July 2009 and started releasing results from its PD model in July 2010. As of the time of this document, default probabilities are produced for about 65,000 exchange-listed firms in 120 economies including Asia Pacific, North America, Europe, Latin America, the Middle East and Africa, with the prediction horizons from 1 month to 5 years. Of the 65,000 firms, over 33,000 are currently active and have their PDs updated daily. CRI's PDs are recalibrated monthly.

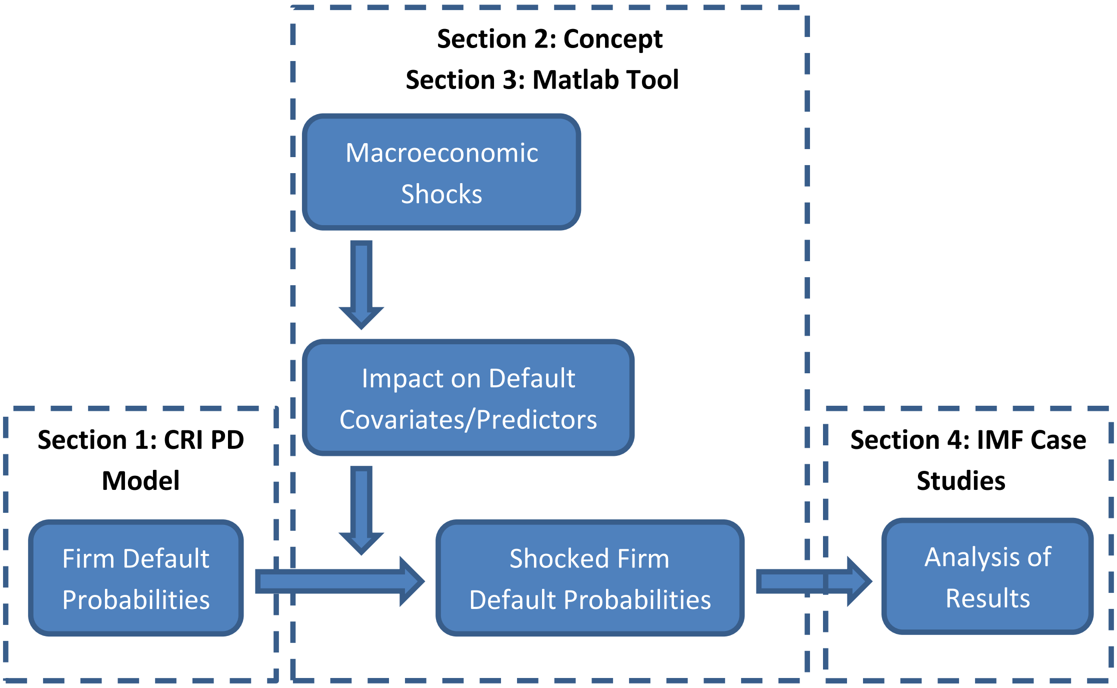
CRI produces PDs for forward horizons of 1 month up to 5 years enabling the user to obtain a perspective of the credit risk term structure. In other words, one can ask, what is the probability that a firm will default in the next month? How about over the next year, or next 5 years?

Importantly, CRI's PDs for *all* firms are freely available to users who can provide evidence of their professional qualifications to ensure that they will not misuse the data. General users who do not request global access can access a smaller list of 5,000 large corporates representing a broad geographic coverage. Data may be accessed via http://rmicri.org.

The BuDA Framework: A Conceptual Overview

BuDA is a bottom-up approach to credit stress testing and scenario analysis. The key question within the analysis is, "How does the credit quality of an economy/sector (or a group of economies/sectors) respond to shocks to macroeconomic variables?" For example, under a prolonged recession with consecutive quarters of negative GDP growth, how badly would the economy/sector suffer?

The basic elements are individual firm PDs of listed firms which, collectively, make up the corporate sector of an economy. As mentioned, CRI produces PDs for almost all of the listed firms in the world, and the work here is to translate macroeconomic shocks into impacts on the individual firm PDs, which are then aggregated into the economy/sector level.



There are four sections. The first section discusses what default events are and introduces the methodology for calculating Probabilities of Default (PDs), which was developed by researchers affiliated with the RMI-CRI team at NUS. The second section contains the conceptual foundation of the BuDA model, of how it assesses the impact which macroeconomic scenarios may have on PDs. For example, how would a country's credit risk profile be affected in the event of a recession? The third section is a how-to section, and it involves step by step instructions to facilitate hands-on practice, and get the reader ready to run its own applications. The fourth section illustrates a few applications in policy work conducted at the International Monetary Fund (IMF), which can serve as a guide to the reader.

Section 1: PD calculation, data and methodology

What are probabilities of default?

When money changes hands, the lender is concerned about whether debt repayments, including interest and principal, will take place as originally agreed with the borrower. The failure of the latter to deliver on its obligations constitutes a default event. For example, Moody's Investor Services, a U.S.-based rating agency, defines default as one of three type of credit events: a missed or delayed disbursement of interest and/or principal; bankruptcy, administration, legal receivership, or other legal blocks to the timely payment of interest and/or principal; and a distressed exchange where the lenders receive in exchange for the original obligations a diminished obligation, or the exchange had the apparent purpose of helping the borrower avoid default.

In the CRI framework, the default event falls under at least one of three categories, namely, (1) a bankruptcy filing, (2) a delisting due to bankruptcy, or (3) a default corporate action. Table 1 and Table 2 list down firm exits which are considered a default and other form of corporate exit. These default events are consistent with those of major rating agencies, and encompass similar categories of events despite differences in the legal framework across countries.

Table 1: Firm Exits Classified as Defaults

Default	
Action Type	**Subcategory**
Bankruptcy filing	Administration, Arrangement, Canadian Companies' Creditors Arrangement Act (CCAA), Chapter 7,11,15 (United States bankruptcy code), Conservatorship, Insolvency, Japanese Corporate Reogranization Law (CRL), Judicial management, Liquidation, Pre-negotiation Chapter 11, Protection, Receivership, Rehabilitation, Rehabilitation (Thailand 1997), Reorganization, Restructuring, Section 304, Supreme Court declaration, Winding up, Work out, Sued by creditor, Petition withdrawn
Delisting	Due to bankruptcy
Default corporate action	Bankruptcy, Coupon & principal payment, Coupon payment only, Debt restructuring, Interest payment, Loan payment, Principal payment, Alternative Dispute Resolution (ADR, Japan only), Declared sick (India only), Regulatory action (Taiwan only), Financial difficulty and shutdown (Taiwan only), Buyback option

Table 2: Firm Exits NOT Classified as Defaults

Other Exits	
Action Type	**Subcategory**
Delisting	Acquired/merged, Assimilated with underlying shares, Bid price below minimum, Cancellation of listing, Failure to meet listing requirements, Failure to pay listing fees, Inactive security, Insufficient assets, Insufficient capital and surplus, Insufficient number of market makers, Issue postponed, Lack of market maker interest, Lack of public interest, Liquidated, Not current in required filings, NP/FP finished, Privatized, Reorganization, Security called for redemptions, the company's request, Scheme of arrangement, Selective capital reduction of the company, From exchange to Over-the-Counter (OTC), Privatised

A default event could trigger off several important events. Company assets and liabilities may be reorganised or liquidated, depending on circumstances. Due to the lack of sufficient assets to meet liabilities, stock price will seriously decline or be completely wiped out. Likewise, bonds and debt instruments written by the company may lose value if investors believe that the obligations cannot be fulfilled completely. Credit default swaps, essentially insurance on the company's bonds, will be triggered for pay out if they are traded.

Since losses associated with defaults can be substantial, there is interest among lenders/credit investors, analysts, financial regulators, and systemic risk supervisors in knowing how likely the default of a firm or group of firms is. The most natural default risk measure of a firm is its PD. PD conveys a sense of the credit quality of a company, with a lower PD suggesting better credit quality, whereas an average or median PD conveys an assessment of credit quality on a group of companies where averaging can be either simple or value-weighted depending on the user's purpose. Fundamentally, the companies with better credit quality are those with a better balance sheet position, liquidity and profitability, and lower stock volatility.

As an illustration, let's examine the case of Lehman Brothers. Figure 1 shows the historical likelihood that Lehman Brothers will default over the next year. Starting in January 2007, the 1-year PD for Lehman Brothers rose steadily from less than 250 basis points (bps[1]) to well over 1300 bps by the end of August 2008. On September 15, 2008 Lehman Brothers filed for bankruptcy under Chapter 11 of the United States Bankruptcy Code, marking it as a default event.

[1] 100 basis points = 1%.

Figure 1: 1-year PD of Lehman Brothers

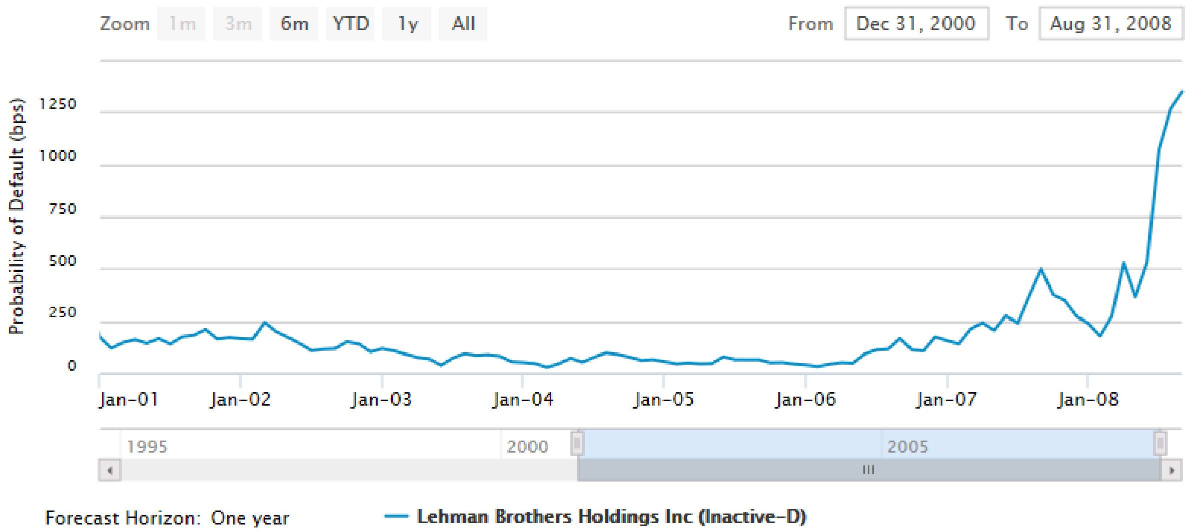


*Figure obtained from CRI website on 31 March 2017

The PD tells us how likely it is for a firm to default over a given period in future. Naturally, we are concerned when a company's PD rises beyond its historical levels or the level at which we entered into a debt or credit risk position. But our focus may go beyond a single firm. If we broaden our perspective beyond a single company, we can also study current levels of the overall PDs of specific industrial sectors or the broader economy, and their potential behaviour in response to changing economic conditions. For example, consider a macroeconomic recession scenario where GDP falls continually. This may cause the PDs of many companies to rise together, increasing the overall PD levels in the economy and raise systemic risk concerns. The BuDA approach, covered in detail in Section 2, can identify this type of situations, helping regulators, lenders, economists, and market observers to analyse and manage risks more adequately.

In comparison to most other sources of credit risk information, CRI-PDs are more responsive, being updated daily. The CRI-PD model employs a cutting-edge default econometrics with its content fully disclosed. The CRI-PD model's implementation is refined over time, and its recalibration is performed monthly to capture potentially changes in model parameters. The model incorporates market data such as the company's distance-to-default[2] (DTD), the stock market performance, and prevailing interest rate, among other information.

The CRI-PD Model: Data, Methodology and Estimation

In this section, we cover the key features of the CRI-PD model with the aim of conveying the intuition behind how it works. Technical details are available in a later section which the reader can skip if interested only in running the model.

Data

Estimating PDs is particularly challenging and difficult in part because we cannot observe their values directly. We only observe whether a firm defaults at a particular point of time, but defaults happen

[2] Measures the adequacy of a company's assets to meet its liabilities, adjusted by volatility. We will cover this in detail later.

rather infrequently. Since there is no unambiguous direct measurement of PD, one can only hope to empirically estimate PDs by applying a model on a large sample of historical records on the default/survival status of firms along with their relevant and observable attributes.

Therefore, we begin by observing actual realized defaults and other corporate exits, and identifying the circumstances under which they occur. Variables which are associated with or have predictive properties of default outcomes are called *default covariates/predictors*. Empirical studies have identified several covariates/predictors which seem particularly associated with the occurrence of defaults, which are incorporated into the CRI model. Table 3 provides a list of the covariates and a brief description. The covariates/predictors fall under two major categories, economy-wide risk factors[3], and firms-specific attributes. The first category, as the name indicates, includes covariates/predictors that tend to affect all firms in the economy and are essentially common risk factors.

Table 3: Covariates Used in Default Prediction

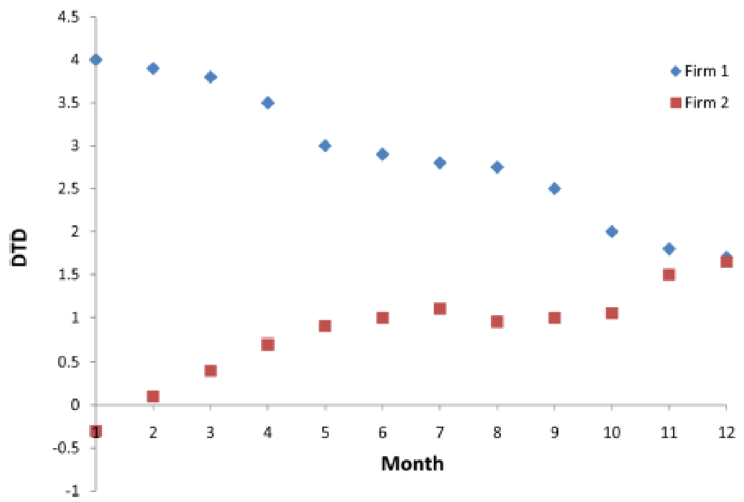
Covariate/predictor	Brief Description
Economy-wide risk factors	
Stock index	Trailing 1-year return on the stock market. A poor stock market performance is generally associated with weaker firm performance and higher probability of firm default.
Short-term interest rate	Yield on 3-month government bills. Generally speaking, if borrowing costs are higher, firms may face more funding constraints and may be more likely to default in the short-term. In the longer horizon, a higher interest rate may indicate positive economic growth and hence lower solvency risk.
Firm-specific attributes	
Distance-to-default • Level • Trend	Volatility-adjusted leverage measure based on Merton (1974), which is logarithm of the ratio between the market value of a firm's assets and its liabilities, scaled by the asset volatility. A smaller DTD increases the likelihood of default. CRI uses a modified DTD measure which we will discuss later.
Liquidity • Level • Trend	Ratio of each firm's sum of cash and short-term investments to total assets. Higher liquidity is more beneficial for the firm.
Profitability • Level • Trend	Ratio of each firms net income to total assets. Higher profitability is more beneficial for the firm.
Relative size • Level • Trend	Logarithm of the ratio of each firms market capitalization to the economy's median market capitalization. Loosely speaking, bigger firms tend to have fewer defaults, although we sometimes observe the opposite.
Market-to-book ratio	Ratio of each firms market value (market capitalization plus total book value of liabilities) to its book value (total book value of assets). It captures the misevaluation or growth opportunity effect.

[3] In the original paper, Duan, Sun and Wang (2012), economy-wide risk factors were referred to as macroeconomic risk factors. However, we reserve this term for scenario analysis/stress testing variables which we discuss in Section 2.

Idiosyncratic volatility	The variation of firm returns which cannot be attributed to the stock market index, using daily data from the past 1 year. Firms with higher idiosyncratic volatility tend to have more variable cash flows and higher probability of bankruptcy.

To increase the predictive accuracy of the PD model, it is helpful to include the trend value, in addition to level values, for some of the firm-specific attributes. Level covariates are computed as the one-year average of the measure, and the trend is computed as the current value of the measure minus the one-year average of the measure. To see why, consider Figure 2 below. Firm 1's DTD has been falling over consecutive periods, while Firm 2 has been rising. Even though they both currently sit at the same point, the statistical trends suggest that in the next period, it is quite likely that the DTD of Firm 1 will fall below Firm 2. Duan et al (2012) found that including trend significantly improves the predictive power of the model for short-term horizons.

Figure 2: DTD Trend



Major effort is put into collecting data on covariates/predictors, defaults and other corporate exits on around 65,000 firms in 120 economies. Among these firms, over 33,000 firms are currently active, and on an ongoing basis, effort has been committed to monitoring these firms and collecting the relevant data. Market-based data such as stock prices and interest rates are updated daily, while data from financial statements are checked daily and updated once available. The main sources of data are Thomson Reuters Datastream and the Bloomberg Data License Back Office Product.

Of the 120 economies, 78 of them have national stock exchanges, and for each of those, a specific, representative stock index and a short-term interest rate are chosen. For the remaining economies, CRI covers the companies which are domiciled in the economy but quoted on a foreign exchange, mostly because those economies do not have a stock exchange. BuDA currently incorporates 33 economies (Table 4 below).

Table 4: List of Economies Incorporated in BuDA

Argentina	Colombia	Ireland	Netherlands	Turkey

Austria	Finland	Italy	Peru	United Arab Emirates
Belgium	France	Jamaica	Philippines	United Kingdom
Brazil	Germany	Japan	Portugal	United States
Canada	Greece	Luxembourg	Singapore	Venezuela
Chile	India	Malaysia	Spain	
China	Indonesia	Mexico	Thailand	

Financial statement and market data for individual firms serve to compute the remaining firm-specific covariates/predictors, namely, DTD, liquidity, profitability, relative size, market-to-book ratio, and idiosyncratic volatility.

Calculating the value of the covariates/predictors is straightforward, except for the Distance-to-Default, for which we provide a more in-depth explanation next.

Distance-to-Default

KMV, now part of Moody's Analytics, first introduced the commercial usage of DTD, which serves as the foundation of its Expected Default Frequency model. Empirical studies have shown that DTD is among the best predictors of default.[4] CRI uses DTD as one of its inputs in its PD model. While conceptually similar to KMV's DTD, there are differences in the calculation of the CRI DTD to incorporate liabilities more holistically and make it applicable to financial firms, as described later in this section.

For now, to facilitate our description, consider a simplified example of two firms in Table 5, weak firm and strong firm.

Table 5: DTD – a Simplified Example

	Weak Firm	**Strong Firm**
Assets ($) (current value)	110	200
Liabilities ($) (promised amount in the future)	100	100
Volatility of Assets	20%	20%
Distance-to-Default (simplified)	10% / 20% = 0.5	100% / 20% = 5

Which firm is more likely to default? Let's take a look at Weak Firm. It has assets of $110 and liabilities of $100. Currently, the value of the assets covers the value of the liabilities (after discounting) adequately. The value of assets, however, fluctuates over time due to several reasons. Among them, adverse business conditions or ineffective collections could reduce the amount of accounts receivables. Changing business circumstances may render some of its fixed assets irrelevant and devalued. Securities held by the firm, such as bonds or stocks may lose value. Over time, there is no guarantee that the value of assets will remain at or exceed $100 at the time of liabilities coming due so that all liabilities can be fully met. This lies behind the option-theoretical

[4] Studies which have used DTD as a default covariate/predictor include Crosbie and Bohn (2001), Vassalou and Xing (2004), Duffie, Saita and Wang (2007), Bharath and Shumway (2008), and Duan, Sun and Wang (2012), to name a few.

basis for DTD in the Merton (1974) model, where the liabilities, i.e. the promised payment, serves as the strike price of a call option and the event in which the promised payment is not fully met amounts to the call option finishing out of the money. The uncertainty over how much the asset will be worth is summarized in "volatility of assets", which is 20% in this example.

DTD attempts to factor in the volatility of asset values. It measures how much headroom assets hold over liabilities per unit of asset volatility. For Weak Firm, assets exceed liabilities by 10 percent, and the volatility of assets is 20 percent, which yields a DTD value of 0.5 (or 10 percent / 20 percent). In other words, excess assets over liabilities are sufficient to buffer a 0.5 standard deviation shock to the value of its assets. Likewise, for Strong Firm, the DTD is 5 (100 percent / 20 percent) which is ten times as high as that of Weak Firm. Strong Firm's assets can absorb a shock of up to 5 times the standard deviation of its asset value. Therefore, Strong Firm is "further" away from default than Weak Firm, and correspondingly, it will have lower PD.

The above example is extremely simplified, so several clarifications are in order. First, we are looking at the market value of assets, not the book value of assets. Book values were recorded at point of entry and can be quite outdated, and fail to reflect latest valuations. Second, the actual formula for DTD used in practice such as the CRI-PD model is more complicated, involves the use of logarithms and square root scaling to the appropriate time horizon. For completeness, we show below the DTD formula at time t adopted by the CRI system for a firm whose time-t asset value is V_t, liabilities due at time T is L, and volatility rate is σ.[5] However, our aim here is to convey the intuition rather than to cover the finer points.

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

Assets and Liabilities in DTD

Let's dig one level deeper. How can we estimate the market value of assets and the promised payment, i.e., some sort of book value of liabilities for DTD purposes given the complex capital structure in reality?

To start, one must first determine what $(T-t)$ to use in practice. For no apparent reasons, the common practice has settled on setting it to one year. But obviously, liabilities for typical firms will scatter a wide maturity spectrum, and hence there is a need to apply some *ad hoc* but sensible adjustment to turn liabilities into a pseudo promised payment (referred to as default point hereafter) in 1-year time. Now, since DTD looks ahead over a 1-year horizon, we can arguably count short-term debts (due within a year) directly in the default point. However, it is only reasonable to subject long-term debts (due beyond one year) to a haircut in order to conform to the 1-year horizon. The

[5]We have purposely left out the risk-adjusted drift term, i.e. $(\mu - \frac{\sigma^2}{2})(T-t)$, in the original DTD formula to obtain a practically more informative DTD. This is because one cannot estimate μ with a reasonable precision due to the high noise-to-signal ratio inherent in typical daily stock returns, a well-known fact in the financial time series literature. For a more complete discussion on the DTD formula, we refer readers to Duan and Wang (2012).

practice advanced by KMV and adopted widely in the credit literature is to haircut long-term liabilities by 50% before adding them to the default point, reflecting the fact that long-term debts are due later than one year.

The CRI model has incorporated an *additional* component into the default point. Financial firms are notorious for being difficult to model, since the majority of their liabilities are neither in the short-term nor the long-term debt category. We refer to the additional and rather large amount of liabilities as *other liabilities*. Within this category, a predominant component for banks is, of course, customer deposits which make up the bulk of a bank's total liabilities. To substantiate the point, it is not uncommon for banks to be leveraged 10-20 times, with the majority of liabilities falling under – you guessed it – *other liabilities* as deposits. For insurance companies, policy obligations will mainly constitute other liabilities.

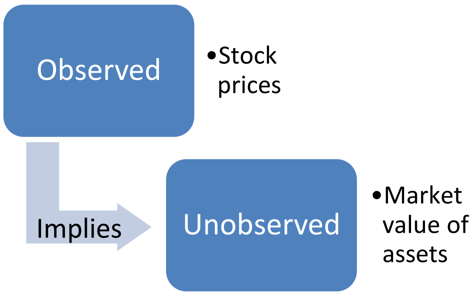
Prior work in the credit literature typically pays scant attention to the role and impact of other liabilities, which are not particularly large in size vis-a-vis short-term and long-term debts for most non-financial firms. When financial firms are conventionally excluded from default analysis, ignoring other liabilities becomes lesser of an issue.[6] In contrast to most studies, the CRI model, following the approach of Duan, Sun and Wang (2012), directly factors in other liabilities by adding to the default point formula a fraction of other liabilities where the haircut rate is treated as an unknown parameter to be estimated along with other model parameters. One way to think about this issue is that customer deposits represent obligations, a portion of which may be withdrawn on short notice while others may stay with the bank for a long term. In order to subjecting them to the same 1-year horizon, haircutting is sensible but the magnitude should not be arbitrary. Setting the haircut rate as an unknown parameter and letting the data inform us seems the most logical way of determining its value. Overall, the modification to the default point formula of KMV given below enables us to expand scope and produce PDs for financial firms in addition to non-financial firms.

$$\text{DTD Liabilities} = \text{Short-Term Liabilities} + 0.5 \times \text{Long-Term Liabilities} + \delta \times \text{Other Liabilities}$$

Moving on, our next challenge is to estimate the market value of assets. Because market values cannot be observed directly, we must rely on indirect methods to measure it. To draw an analogy, we cannot see the healthiness of a person's heart directly, not unless we surgically remove and comb through it for blockages and clots. We are not completely helpless however, as we could instead measure his cholesterol level, we could check his blood pressure, obesity level, run an ECG and interview his lifestyle habits to draw a fairly accurate diagnosis of his heart condition, all the while without looking directly at his heart (or removing it). In the same way, we can assess the market value of a firm's assets by looking at its stock price. If the stock price rises, it should mean that the market value of assets rises. Why? Because stock securities are in fact call option like claims on assets! More precisely, stock securities are claims on the residual value of assets after creditors (debt holders) have been paid off.

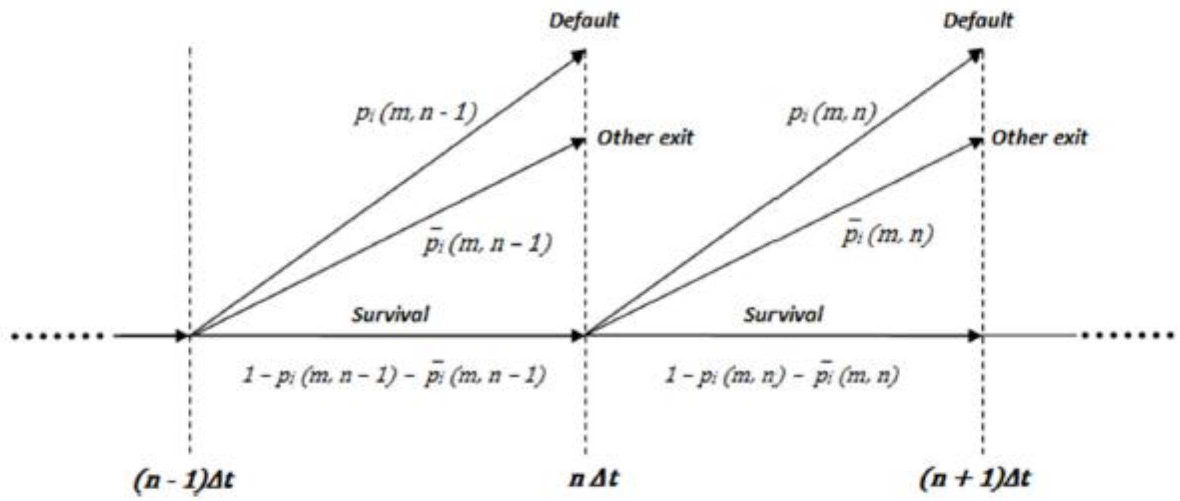
[6] At odd with the typical finding, Campbell, Hilscher and Szilagyi (2008) concluded that DTD does not help with default prediction. According to Zou (2016), their conclusion was due to the inclusion of financial firms in the sample while leaving other liabilities untreated. Zou (2016) shows that using either the DTD with the KMV default point formula on the sample excluding financial firms or the DTD with the revised default point formula factoring in other liabilities will reverse the conclusion of Campbell, Hilscher and Szilagyi (2008).

So, there is information pertaining to market value of assets within stock prices, and all we need is a suitable technique to extract it. For those of us who are familiar with options, stocks are essentially call options on the market value of assets, with the liabilities (default point in our adopted jargon) being the strike price. Consequently, exploiting this fact, we can infer market values of the assets from stock prices by inverting the option pricing formula. For finer points and references on the estimation method, we refer readers to Duan and Wang (2012). If this sounds too technical, the main takeaway here is that market values of assets, while not observable, can be extracted from stock prices. It is like looking for the smoke to find out where the fire is.



Methodology: Modelling the Default Process

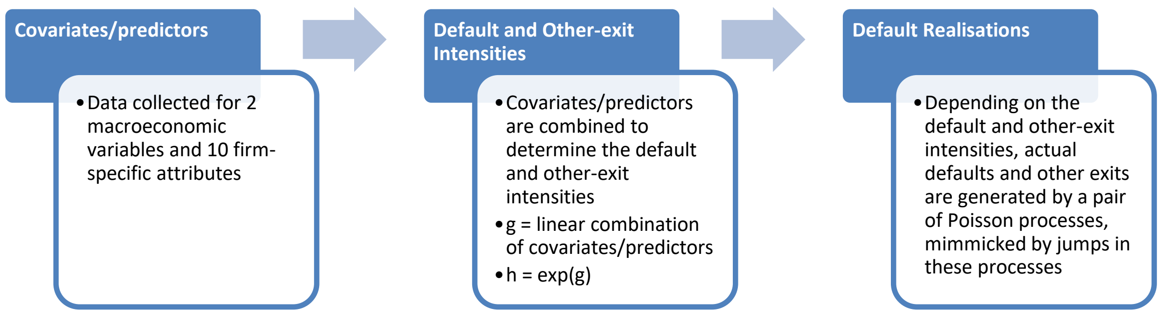
In order to utilize the data collected, we must translate the covariates/predictors into default and other-exit probabilities, keeping in mind that we observe only the actual defaults and other forms of corporate exits. Imagine for a moment that time can be split up into discrete periods. In each of these periods, one of three events may occur. First, the firm survives and continues as a viable entity in the next period. Second, the firm may default, in which case a "default" event is observed. Third, the firm may be delisted or merged with another firm, in which case we observe an "other-exit" event. We include this third category of events because they are a key component to measuring survival likelihood and vital to the estimation of default probabilities free of survival biases in a multi-period environment. In some economies as documented in Duan, Sun and Wang (2012) and the NUS-RMI CRI Technical Report (2016), exits due to reasons other than default can be up to ten times as many as exits due to default.



An attentive reader may realise that we are trying to associate the aforementioned covariates/predictors with the period by period probability of default. If the situation for the firm is unfavourable, for example DTD is low, and liquidity and profitability are poor, this translates into a higher probability for the firm to default. However, firms may still survive when the situation is bad. Being a matter of chance, the firm may or may not actually default in the current period. They are simply *more likely* to default. If the bad situation persists for a period of time, chances are that the firm will default eventually. In this manner, the evolution of the covariates/predictors over time may be related to the actual realised defaults.

This intuition can be formalised as a Poisson process, which is a common method for modelling the occurrence of a rare event (default or other exit in this case) as time evolves. The CRI model, based on Duan, Sun and Wang (2012), takes in the set of covariates/predictors and combines them into time-varying forward *default intensities*, which reflect default and other-exit events being generated via a pair of Poisson processes for each firm, and their intensities are correlated through dependency among covariates/predictors. Across firms, different pairs of intensities may also be correlated via their covariates/predictors.

Figure 3: Modelling of Defaults and Other Exits Using a Pair of Poisson Processes



Estimation

How do we combine the covariates/predictors into the forward default/other-exit intensity? Until now, we have not specified exactly how this is done. A straightforward way to combine the information coming from each of the covariates/predictors is simply to add them up, in other words, compute:

$g = \beta_0$ + $\beta_1 \times$ Stock index + $\beta_2 \times$ Short-term interest rate + $\beta_3 \times$ Distance-to-default + $\beta_4 \times$ Liquidity + $\beta_5 \times$ Profitability + $\beta_6 \times$ Relative size + $\beta_7 \times$ Market-to-book ratio + $\beta_8 \times$ Idiosyncratic volatility + $\beta_9 \times$ Distance-to-default Trend + $\beta_{10} \times$ Liquidity Trend + $\beta_{11} \times$ Profitability Trend + $\beta_{12} \times$ Relative size Trend.

Next, having a negative default intensity does not make sense, since a presently surviving firm can only default or survive in the next period; it cannot "anti-default" as there is no such interpretation. Therefore, the default intensity is set to $h = \exp(g)$, or taking an exponential of the weighted sum of individual covariates/predictors to ensure that the intensity is positive.

Estimating the model is then a matter of choosing the parameters $\beta_0, \beta_1, \beta_2, \dots, \beta_{12}$ so that the model best matches the data (observations of actual defaults/other exits). Briefly, this involves writing the likelihood of the model given the data, and maximizing it over the parameters. To grasp the intuition, consider the parameter β_5 for profitability. Since defaults tend to occur when profitability is low, we expect β_5 to be negative. A positive β_5 means that the default intensity increases with profitability, and in reality disagrees with what the data says. In other words, it is not very likely for β_5 to be positive. The optimisation method picks β_5 and all the other β s simultaneously so that they best agree with what we see in the actual defaults/other exits. A full statistical treatment would be too involved here, but the key takeaway is that the parameters are selected to explain the default data as much as possible.

Recall that CRI produces a term structure of PDs from 1 month up to 5 years. To achieve this, we need to estimate not only a single set of parameters $\beta_0, \beta_1, \beta_2, \dots, \beta_{12}$, but an entire term structure of them up to 5 years into the future. Intuitively, it might sound infeasible or impossible to estimate the parameters for each and every time point into the future up to 5 years. But it is actually doable via a decomposability result established in Duan, Sun and Wang (2012). Naturally, we would expect adjacent horizons to share similar parameter values, and hence we can assume that the term structure of the parameters follow a particular class of curves. So using β_5 as an example again, we assume that its values across the various forward-looking horizons can be described as a curve, then instead of estimating each and every β_5, we estimate the shape of the curve which turns out to be computationally more challenging but intuitively more desirable. Apart from smoothing the parameter estimates, an added benefit is that the parameter curve allows for easy extrapolation into longer horizons for which data may not even exist. Exactly what form this curve takes is a matter of choice. For example, we may simply decide to fit a quadratic (polynomial) curve, but this form may perform poorly. Duan, Sun and Wang (2012) deployed a Nelson–Siegel curve and the CRI model followed. While less widely known, the Nelson-Siegel curve is adequately simple and works well. This approach is very similar to how the term structure of bond yields is modelled.

Calibration Groups

Earlier we mentioned that the CRI coverage spans around 65,000 firms. With such an extensive diversity, it is only natural to expect that the parameters $\beta_0, \beta_1, \beta_2, \ldots, \beta_{12}$ differ across different economies of firms. But at the same time, one must strike a balance with the need to have a sufficient number of defaults in the sample to conduct model estimation/calibration. The CRI system thus deploys calibration groups by putting together firms from different economies that are expected to be more similar than not.

Currently, Canada and the US belong to the North America calibration group, and the developed economies of Asia-Pacific (Australia, Hong Kong, Japan, Singapore, South Korea, Taiwan and New Zealand) form another calibration group. China and India, the two major emerging economies of Asia Pacific are each calibrated as individual groups. All the European countries covered by the CRI are in a single calibration group. The other emerging economies of Asia Pacific, Latin America, Middle-East and Africa form the "emerging markets" calibration group.

As a general rule, all economies in the same calibration group share the same coefficients for all variables except for the short-term risk-free interest rate variable[7]. The short-term interest rate variable is entered as the current value minus the historical month-end mean in order to reflect the contemporary change relative to the historical average. Its coefficient is allowed to vary across economies, because different economies with different currencies may have different dependencies on their interest rates, the levels of which can also differ significantly across economies. The Euro zone deserves a special explanation, where all countries after joining the Euro zone begin to use German short-term interest rate.

Prediction Accuracy

The CRI conducts tests to ascertain that the PDs are informative of potential defaults. A popular standard in assessing the discriminatory power of a rating system is to use Accuracy Ratio (AR). The intuition behind AR is that if firms with high PD are indeed those that actually default, then AR is high. In other words, the distressed firms have been properly discriminated or distinguished from the safe firms. On the other hand, if PD levels have little to do with defaults, then AR is low. AR ranges from 0 to 1, with 0 indicating a completely uninformative rating system, and 1 a perfect system. The use of AR has been discussed, for example, in a paper on rating validation prepared by the Basel Committee on Banking Supervision (BCBS, 2005b).

Below, Table 6 reports the ARs for various prediction horizons and economies. For example, procedure for 1-year AR calculation is as follows: first, we calibrate the parameters using the full data set (in the tables below, this is until March 2017). Next, standing at a particular point in time, say 31 December 2000, we extract the PD forecasts 1 year aheadfor all firms based on these parameters as well as the actual defaults that occurred in 2001. We subsequently pool the PD forecasts across all time spots for this economy and compare them against the actual defaults so as to get the 1-year AR.

---

[7] As exceptions, Eurozone uses German interest rate owing to their economic integration. In addition, Indonesia has its own coefficient on relative size.

Table 6: Accuracy Ratios for the CRI-PD Model

	AR			
Economy	1mth	1yr	2yr	5yr
Australia	0.83246	0.67209	0.56095	0.39794
Brazil	0.87197	0.80106	0.71095	0.49877
Canada	0.95304	0.82627	0.70260	0.49930
China	0.70164	0.69218	0.65805	0.53890
Germany	0.86694	0.69298	0.57753	0.46766
Denmark	0.82046	0.76551	0.61542	0.53745
France	0.85554	0.72841	0.63541	0.55865
Hong Kong	0.77080	0.53696	0.42181	0.23816
India	0.72501	0.65636	0.60090	0.47297
Indonesia	0.73299	0.69022	0.60649	0.41965
Italy	0.87879	0.82573	0.66153	0.46295
Japan	0.91234	0.85442	0.79959	0.66773
Malaysia	0.84062	0.78735	0.71994	0.53608
Mexico	0.82134	0.79538	0.73614	0.56559
Netherlands	0.88257	0.83564	0.68133	0.55146
Norway	0.95316	0.82900	0.60700	0.29491
Philippines	0.72890	0.65725	0.64384	0.57070
Poland	0.87899	0.75093	0.59734	0.34276
Russian Federation	0.79778	0.42557	0.19024	0.083178
Singapore	0.81505	0.71364	0.54761	0.32252
South Africa	0.92794	0.85689	0.74132	0.46368
South Korea	0.87184	0.73760	0.66442	0.55799
Sweden	0.90945	0.79508	0.68702	0.41611
Taiwan	0.87882	0.77457	0.70024	0.61224
Thailand	0.81878	0.78254	0.73483	0.61134
United Kingdom	0.88910	0.77209	0.62690	0.43333
United States	0.94433	0.83615	0.72316	0.53831
Developed Asia-Pacific	0.86693	0.74847	0.66370	0.53063
Emerging MKT	0.82510	0.77456	0.70990	0.55807
Europe	0.87762	0.75037	0.61715	0.45191
North America	0.94531	0.83533	0.72152	0.53539

*Figure taken from addendum 2 to the CRI technical report 2016 update 1.

The model is able to achieve strong AR results mostly greater than 0.80 at the one-month horizon, with stronger results for developed economies. At 1-year, ARs are mostly healthy and above 0.70. There is a drop in AR at the 2-year and 5-year horizons, but this is to be expected as we move further on in the term structure.

The ARs in some emerging market economies such as India, Indonesia, and the Philippines are noticeably weaker than the results in the developed economies. This can be due to a number of issues. The quality of data is worse in emerging markets, in terms of availability and data errors. This may be due to lower reporting and auditing standards. Also, variable selection is likely to play a more important role in emerging markets. The variables were selected based on the predictive power in the US. Performing variable selections specific to the calibration group are expected to improve predictive accuracy, especially in emerging market economies. Finally, there could be structural differences in how defaults and bankruptcies occur in emerging market economies. If the judicial system is weak and there are no repercussions for default, firms may be more prone to default.

Previously, China's AR for 1-year PD was 57%, but has improved to 69% with the introduction of structural break estimation. The structural break occurs in December 2004, and we allow the coefficients for DTD level and the intercept to be different after this break. However, we incorporate a modification to standard structural break estimation. Instead of a sudden change in these coefficients, we allow the parameters to change smoothly[8] into the new parameters over time, reflecting gradual changes rather than being brought about by a sudden shift in economic structure.

ARs are good for determining if PDs rank the firms correctly according to their relative default risks. For example, a riskier firm should have a higher PD. There is a slight catch, however. Theoretically speaking, even if we multiplied all our PDs by a factor of say 2 or 3, the accuracy ratio would remain unchanged. Therefore, we also want to find out if the actual PD level, say a PD of 5%, actually reflects a 5% default risk. In other words, we want to assess the goodness of fit of the PDs. This is done by comparing the predicted and actual number of defaults at the aggregate level over time. Standing at a particular time point, we ask how many firms are expected to default in the next year, which we can compute using the forward PDs that have already been estimated at that point. We then compare this number with how many firms actually defaulted in the next year. Figure 4 shows this for the US and Figure 5 for China[9]. The predicted levels tracks the actual levels quite nicely, bearing in mind that the numbers are predicted in advance of the actual defaults. Most of the other economies exhibit a similar pattern.

Figure 4: Predicted and Actual Defaults for United States

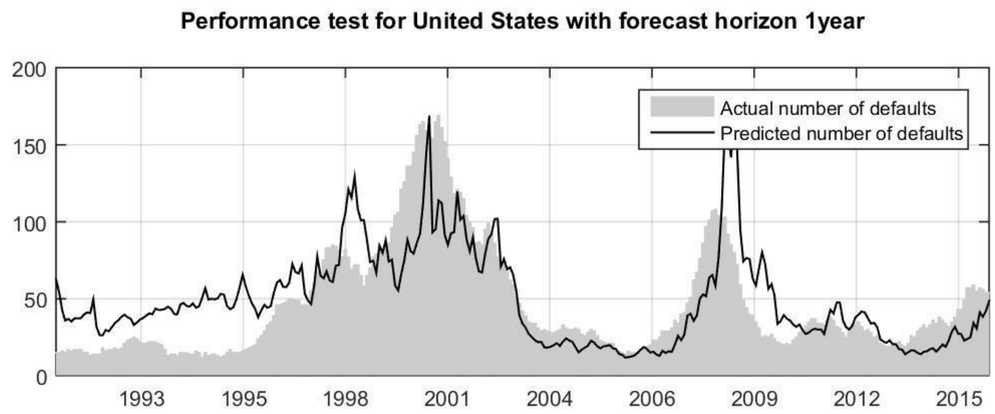
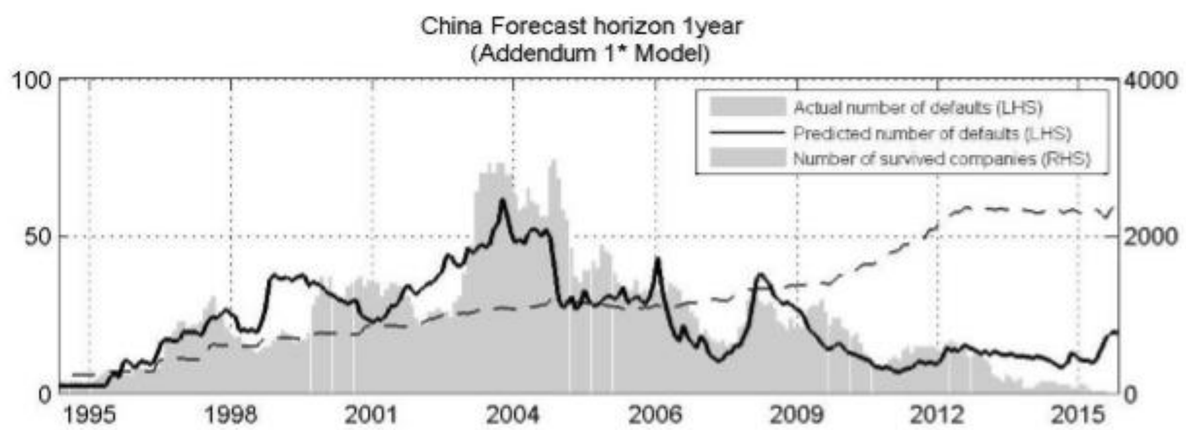


Figure 5: Predicted and Actual Defaults for China



[8] This is implemented using a logistic function which starts with the old set of parameters and tends to the new set of parameters as time goes on.
[9] Figure 4 and Figure 5 are from CRI staff calculations.

Alternatives

We now briefly mention a few alternatives which may be used to assess the credit quality of a company.

Letter credit ratings issued by credit rating agencies (CRAs) such as Standard & Poor's, Moody's Investor Services, and Fitch Ratings are possibly among the most widely used. For example, the AAA/Aaa grade is the highest possible rating and represents companies with extremely low credit risks. The opposite is true for companies rated CCC. While popular and easy to understand, criticisms of such letter credit ratings have arisen in particular after the financial crisis of 2007-2009. One topic is the lack of responsiveness to changes to the company's circumstances. Indeed, Lehman Brothers was rated as an investment grade company with at least an A rating by the big three rating agencies up until mid-September right before its bankruptcy. Another criticism is that CRAs have a profit incentive in the companies which they rate, leading to potential conflicts of interest, as noted in The Financial Crisis Inquiry Report of 2011 prepared by the Financial Crisis Inquiry Commission.

In comparison, CRI's PDs are more responsive, being updated daily and recalibrated monthly. The model incorporates market data such as the company's distance-to-default (DTD), the stock market performance, and prevailing interest rate, among other information. CRI's PDs are provided as a public good and distributed as a free service with the intention of promoting research and development in credit risk.

More closely related to CRI's PDs are Moody's Expected Default Frequency (EDF) database and Kamakura's PDs, but neither of them intends to measure the probability that a firm will default over a specified period of time. According to available information, Moody's EDFs primarily incorporate a distance-to-default measure but without further technical details on its model, whereas the Kamakura's method is unclear to us. Both Moody's EDFs and Kamakura's PDs are paid services. CRI's PDs, on the other hand, incorporates not only a DTD measure, but also a range of macroeconomic and firm-specific attributes as explained earlier, and the methodology is fully disclosed to the public.

Summary of Section 1

We wrap up section 1 with a quick recap of the material covered so far. PDs tell us how likely it is for a firm to default over a given period in future, and convey a more granular sense of the credit quality of a company. We can use PD to study a range of topics, including questions pertaining to financial stability. The PDs produced by CRI have the advantages of global coverage, a term structure of up to 5 years, and they are free with the methodology fully disclosed. Being a non-profit setup, it does not bear an inherent conflict of interest with profit incentives.

The CRI model uses the Poisson process to model survival, and is calibrated against actual defaults and other exits. The default and other-exit intensities for the Poisson processes are linked to 2 economy-wide risk factors and 10 firm-specific attributes, namely, stock index, short-term interest rate, market-to-book ratio, idiosyncratic volatility, DTD, liquidity, profitability, and relative size, with the latter 4 having both level and trend variables. DTD is particularly helpful for predicting defaults, and our modification of incorporating customer deposits and insurance policy obligations (other liabilities) makes it applicable to financial firms.

For the purpose of the BuDA scenario analysis tool, CRI's PDs are already included in the distribution. Users are still free to access the latest PDs via the CRI website (https://rmicri.org) for other purposes. A brief guide has been included in the Appendix.

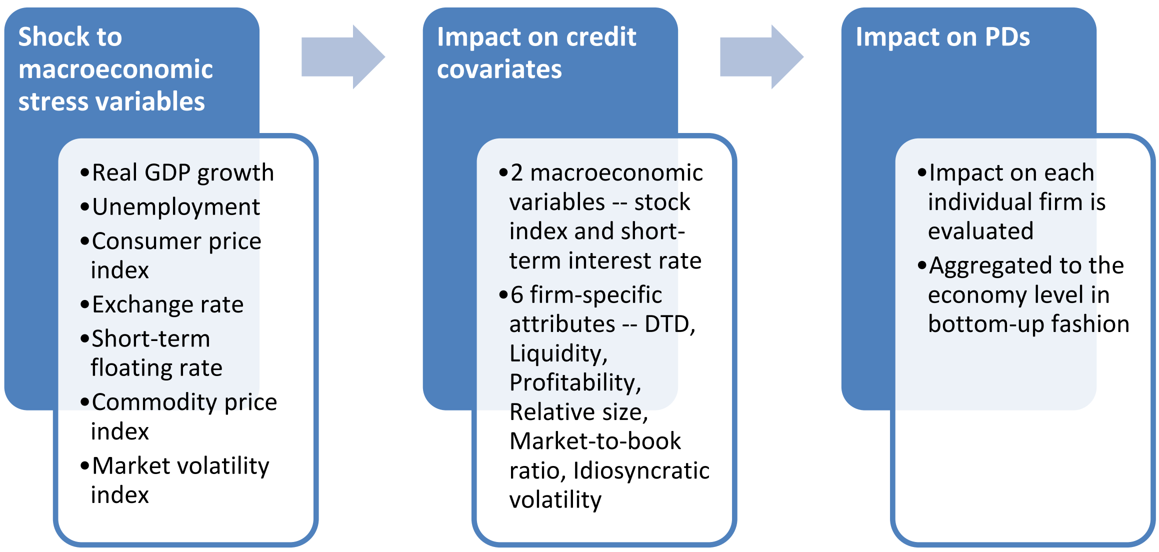
Section 2: Macroeconomic Scenario Analysis

Equipped with a basic but workable understanding of CRI's PDs, we will explore the conceptual underpinnings of the bottom-up default analysis in this section. Implementation details and operation of the toolkit will be left for Section 3.

What is bottom-up default analysis (BuDA)?

BuDA is an approach to credit stress testing and scenario analysis that relates a shock profile to macroeconomic stress variables, such as GDP, unemployment, inflation and others, to the credit quality of the economy. For example, an economist might be interested in analysing the health of the economy when a prescribed severe recession, i.e. consecutive falls in GDP, occurs. BuDA translates the specified fall in GDP into an impact on the default covariates/predictors which we discussed in Section 1, which in turn determines the impact to the PDs. In other words, starting with the macroeconomic dimensions which an economist might be concerned with, BuDA interprets these shocks in terms of their corresponding effect on credit quality. BuDA may be useful to regulators, central banks, and banks for increasing awareness and understanding risks.[10]

Figure 6: Overview of BuDA Approach



As one might expect, the macroeconomic variables interesting an economist may not coincide neatly with the default covariates/predictors – 2 economy-wide risk factors and 10 firm-specific attributes which were useful for PD estimation, covered in Section 1. The idea is therefore to map the impact of the macroeconomic shocks to these default covariates/predictors, and in turn measure the follow-on impact on PDs. Let's use some math, only because it conveys the intuition concisely.

[10] The BuDA tool caters for user-supplied stress variables other than the 7 listed here. It also caters for aggregation to a user-supplied portfolio instead of pre-specified economies.

Equation 1: Stress Testing Regressions

Change in economy-wide risk factors | Macroeconomic stress variables | Lagged economy-wide risk factors | Random error terms

$$\Delta X_{m,t} = \beta^X_{m,0} + \sum_{k=1}^{n} \beta^X_{m,k} Z_{k,t} + \gamma^X_{m,1} X_{m,t-1} + \gamma^X_{m,2} X_{m,t-2} + \varepsilon^X_{m,t},$$

$$\Delta \bar{Y}_{i,j,t} = \beta^Y_{i,j,0} + \sum_{k=1}^{n} \beta^Y_{i,j,k} Z_{k,t} + \gamma^Y_{i,j,1} \bar{Y}_{i,j,t-1} + \gamma^Y_{i,j,2} \bar{Y}_{i,j,t-2} + \varepsilon^Y_{i,j,t},$$

Change in firm-specific attributes | Macroeconomic stress variables | Lagged firm-specific attributes | Random error terms

As illustrated in the equations, changes to the default covariates/predictors are driven by the macroeconomic stress variables. Since some of the covariates/predictors exhibit persistence, we include lag terms to take care of autocorrelation. For those who are familiar with autoregressive systems, the model implemented above is an AR(2) model, i.e. 2 lags. This linkage to macroeconomic stress variables of interest affords flexibility of choice and effectively decouples the variables for stress consideration from the covariates/predictors that work for default prediction.

Now pay attention to the subscript for the macroeconomic stress variables on the right side of the equation, then, look at the subscript for the change in default covariates/predictors on the left side. Notice that both have the subscript *"t"*? This means that information pertaining to the macroeconomic stress variables enter into the default covariates over the same time period with no lag, in other words, the default covariates/predictors respond at the same time. We say that macroeconomic stress variables have a *contemporaneous effect* on the default covariates/predictors.

We take this opportunity to also point out the random error terms at the end of each equation. As much as we attempt to predict the impact which macroeconomic stress variables may have on the default covariates/predictors, there is an element of uncertainty, and it is only wise to recognise it. Consequently, the model is not limited to predicting a single deterministic outcome under the prescribed macro scenario, but can predict a range of possible outcomes. Later on, we will use the above equation to simulate the system over the intended horizon of interest by drawing random numbers for the error terms.

We need to make one clarification with regards to the second regression pertaining to changes in firm-specific attributes. Recall that we have a dataset for around 65,000 firms, and a good proportion of these do not have a long history of data for stable parameter estimation. Also, estimating the regression for each and every one of them might entail additional sampling errors which would be impounded into the final results. Even if we are only focusing on specific regions, the US has over 15,000 firms, while the Eurozone 12 is shy of 12,000 firms and ASEAN 5 spills slightly over 6,000. To accommodate the large number of firms and for feasibility sake, we need some simplification and choose to conduct the stress testing regressions only for the *mean*[11] of each firm-specific attribute, where the mean is taken across firms. That said, we continue to model each firm's individual firm-specific attributes by regressing[12] them against the mean values so that their values are updated after the macroeconomic scenario is applied. The benefit comes from running the stress testing regressions only for one set of (mean) variables.

Scenario Analysis: Macroeconomic Variables

Table 7 below lists and briefly describes each of the standard variables implemented within BuDA. Within the implementation which we discuss in the next section, users may define their own macroeconomic stress variables, but for now let us focus on the standard (default) ones.

Table 7: Macroeconomic Stress Variables

Macro variables	Description
Real GDP growth rate (GDP)	Percentage change of seasonally adjusted real gross domestic product
Unemployment rate (UNEMP)	Difference of seasonally adjusted unemployment rate
Consumer price index (INFL)	Percentage change of consumer price index
Exchange rate (NEER)	Percentage change of the BIS nominal effective exchange rate
Short-term floating rate (IBOR)	Difference of 3-month interbank offer rate
Commodity price index (GSCI)	Percentage change of the S&P GSCI commodity index
Market volatility index (VIX)	Percentage change of the CBOE VIX

Real GDP reflects the state of an economy with its growth rate serving as a proxy for the growth in incomes and earnings of firms. A higher growth would generally lead to higher corporate earnings and lower default risk.

Unemployment rate affects the consumption and spending of households. An increase in the unemployment rate would generally result in lower revenues for firms, particularly those that are consumer-oriented (e.g., restaurants, retailers, etc.), and increase their default likelihoods.

Consumer price index provides a measure of inflation and controlling inflation rates is one of the primary objectives of monetary policy. High inflation is usually considered a signal of macroeconomic mismanagement and a source of uncertainty. Higher inflation leads to increased costs and tends to

[11] More precisely, we use the 20% trimmed mean, in other words, the top and bottom 20% of observations are dropped before taking the mean.
[12] Specifically, we use an AR(3) model with LASSO-OLS.

impair credit quality. However, higher inflation may also reduce debt burden in real terms, and thereby improve creditworthiness.

Exchange rate affects the bottom-line of a wide range of firms directly or indirectly through trade and investment flows; for example, a stronger domestic currency will typically benefit importers and likely hurt exporters. Firms in a small open economy will be particularly sensitive to exchange rate movements. We use the BIS nominal effective exchange rate indices, which are calculated as geometric weighted averages of bilateral exchange rates.

Short-term floating rate is a common benchmark that banks use in determining lending rates for variable-rate loans. It is also the rate that is applicable to short-term corporate funding via debt/commercial paper markets. A higher borrowing rate/cost of funding is expected to increase default risk of firms owing to higher interest expense.

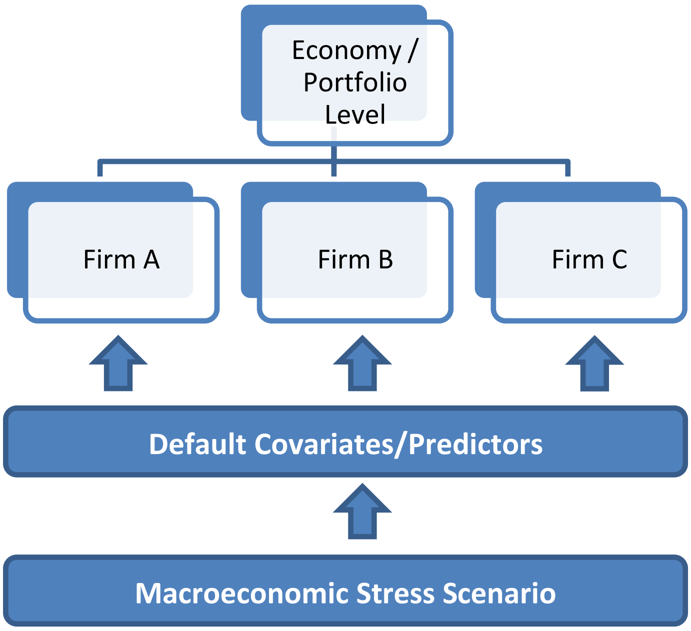
The S&P GSCI commodity index captures an important production factor cost. A higher commodity price typically benefits commodity producers but causes deteriorated creditworthiness for other companies.

The VIX index offered by the Chicago Board Options Exchange reflects the volatility in the S&P 500 index portfolio and is a commonly used proxy variable for gauging the risk level in US stock markets, but can also be indicative of how volatile stock markets in other economies, given the global linkage of financial markets and dominance of the US economy.

A word needs to be said about mixed-frequency data. GDP and unemployment data is typically available quarterly, whereas the other variables are available monthly. Furthermore, the PD data is available monthly (in fact daily) as well. Hence, if we simply treated the data as a monthly panel, GDP and unemployment would exhibit spikes every 3 months and flat at other times. This is not ideal. To address this issue, we incorporate two features. First, instead of jumps every 3 months, the data is interpolated so that it changes gradually rather than suddenly. Second, we apply a trick and rewrite the scenario analysis regressions in Equation 1 so that the variables become 12-month aggregates instead of 1-month. For those in the know, this is done by iteratively substituting the autoregressive formula into itself. The regression of default covariates/predictors against stress variables on a 12-month aggregate basis makes the estimation much less sensitive to how the quarterly data are converted into monthly data.

Top-down vs bottom-up

Figure 7: Bottom-up Aggregation



As the name suggests, "bottom-up" refers to an approach where the impact on each individual firm in the pool of firms can be evaluated, and aggregated up into the portfolio or economy level. The methodology computes PD impacts for the most granular firm level, and builds it up to the overall portfolio level by aggregation or taking a statistic such as the average, median or percentile.

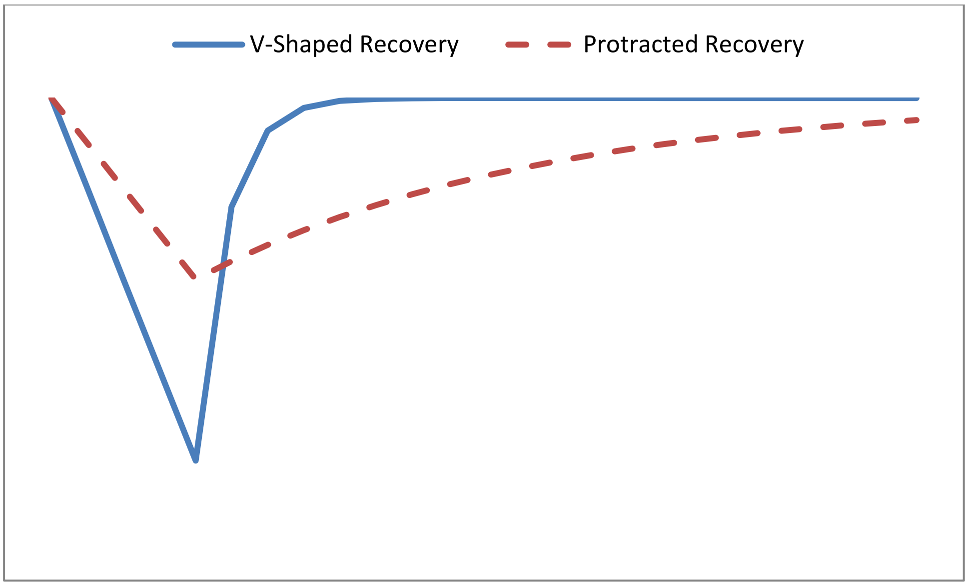
On the other hand, "top-down" scenario analysis involves first consolidating the data, and then performing the scenario analysis on the consolidated data. In a typical setup, a central regulator or financial stability function might not be able to work with granular individual firm data, hence limited to using consolidated data. The BuDA approach which leverages on the vast PD dataset produced by CRI provides a practical bottom-up solution.

Macroeconomic scenarios

BuDA allows a range of macro-economic stress scenarios to be specified without imposing too much limitation. One important point to note is that the stress scenario may be multi-period, occurring over several months or years instead of a single time period – in fact, the worst situations are likely an accumulation of undesirable events instead of a drastic single period event (although that can be accommodated).

Next, stress scenarios need not be stylised as a unidirectional shock. For example, a "V-shaped" recession and recovery used by the IMF may be applied. In this case, both the fall of GDP into recession and subsequent recovery are incorporated. Alternatively, a protracted recovery situation may also be analysed. As can be seen, the flexibility here allows even the actual evolution of historical events to form the stress scenarios.

Figure 8: Examples of GDP Growth Rates under V-Shaped Recovery and Protracted Recovery



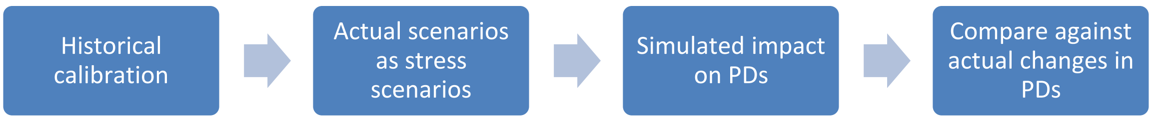
Performance and Backtesting

How well does this scenario analysis methodology work? To answer this question, we can conduct backtesting and assess the suitability along a few dimensions. Does the methodology perform well for different regions or industries? How about different forward-looking horizons?

In the backtesting exercise which we share here, we rewind time back to the past. After a series of events that involve sitting in our DeLorean and a bolt of lightning, we arrive at today, 31 December 2001. Using data from January 1994 to December 2001, the scenario analysis regressions are calibrated, so that the model only relies on data available up to today, 31 December 2001.

We are interested in how the model performs 12 months later. Does it predict the new PD profile well? Since we are future beings, we already know what will happen to the macroeconomic stress variables from 1 January 2002 to 31 December 2002. We also know what will happen to the PDs over the same period. To test the model, we feed the *actual scenario* for the macroeconomic stress variables to the model, and let the model forecast the PD profile for the future date of 31 December 2002. If the PDs generated by the model are good, they should match well with the actual changes in the PD profile.

Specifically, the model simulates 12 months of time periods from January 2002 to December 2002, given the actual macroeconomic stress scenario, and obtains a possible PD outcome as at 31 December 2002. This is done 1,000 times to obtain 1,000 possible realisations of the PD profile. The averages of the simulated and actual PD profiles are then compared.

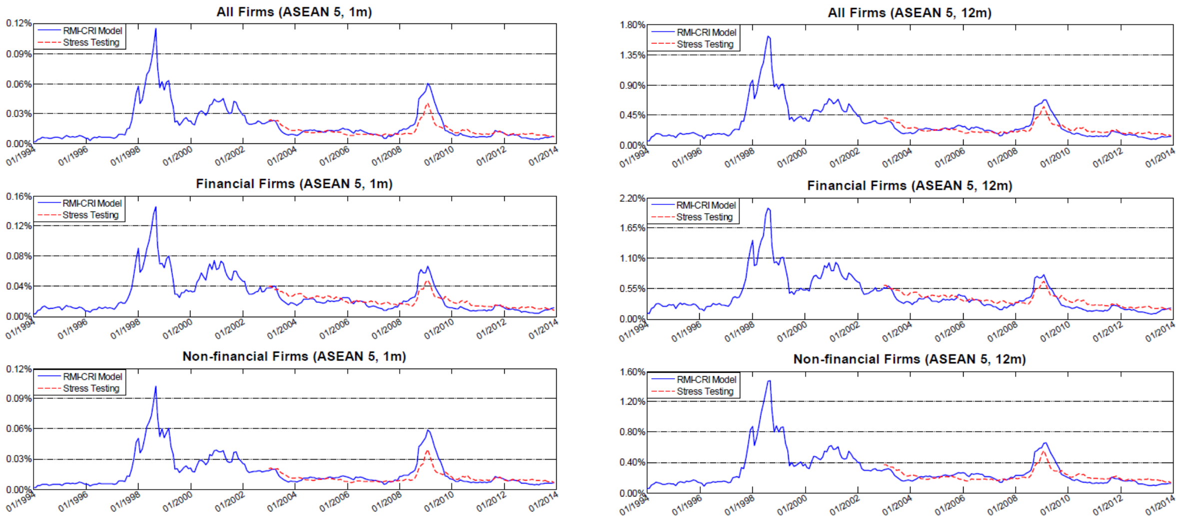


We can take this one step further and do the backtesting for each month starting from January 2002 up until December 2012, so that we can plot a time profile of the simulated PDs against the actual PDs. The charts below[13] show how well the prediction has done over time.

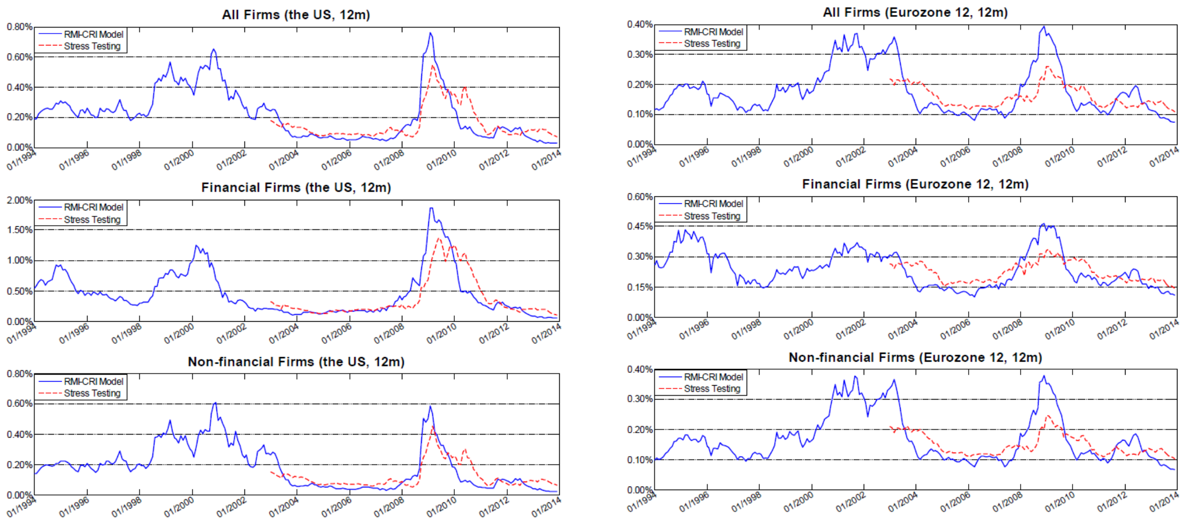
The backtesting was done for 3 different regions – ASEAN 5, US, and Eurozone 12. ASEAN 5 includes Indonesia, Malaysia, the Philippines, Singapore and Thailand and Eurozone 12 includes Austria, Belgium, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, the Netherlands, Portugal and Spain. For each region, we further analyse the performance of the financial and non-financial sector.

For ASEAN 5, data from the Asian Financial Crisis was helpful for calibrating model responses, and performance during the Credit Crisis is encouraging. Both the 1 month and 12 month PDs performed well. For US and Eurozone, both did not have a major crisis to learn from since 1994 (remember we are still in 2001). Hence, performance during the Credit Crisis would not be stellar. Therefore, for the backtesting purpose, the whole sample up until December 2012 was used. In other words, the model was trained with the Credit Crisis data for backtesting purposes. This is not entirely unreasonable, since going forward in, say, 2017, training of the model will include the Credit Crisis.

Figure 9: Backtesting Charts



13 Extracted from Duan, Miao and Wang (2015).



Another aspect which we can assess is the R^2 of the scenario analysis regressions. If the variables on the right-hand side of Equation 1 can predict the default covariates/redictors on the left-hand side well, then the model is more likely to perform well. The following Table 8, Table 9, and Table 10 show the R^2 for each of the 3 regions backtested. For each default covariate/predictor and each sector, the R^2 is reported.

Table 8: Average in-sample 12-month R^2 for stress-testing regressions for ASEAN 5

Sectors	Input Variables of the Default prediction model					
	DTD	LIQ	PROF	RSIZE	MB	SIGMA
Basic material	35.61%	37.41%	36.60%	25.42%	40.82%	36.80%
Communications	39.83%	44.21%	39.44%	26.84%	47.92%	34.22%
Consumer (cyclical)	41.07%	27.60%	37.87%	26.27%	40.28%	37.25%
Consumer (noncyclical)	37.46%	31.83%	42.64%	34.30%	39.57%	40.47%
Diversified	31.07%	37.99%	35.67%	40.79%	45.05%	41.25%
Energy	41.71%	30.27%	43.82%	33.78%	47.14%	38.35%
Financial	37.00%	14.56%	39.30%	32.36%	48.20%	42.45%
Industrial	38.20%	28.44%	36.07%	25.99%	45.13%	38.96%
Technology	35.55%	28.65%	39.75%	24.22%	39.54%	35.90%
Utilities	36.98%	39.60%	40.49%	47.08%	52.11%	39.53%

Table 9: Average in-sample 12-month R^2 for stress-testing regressions for the US

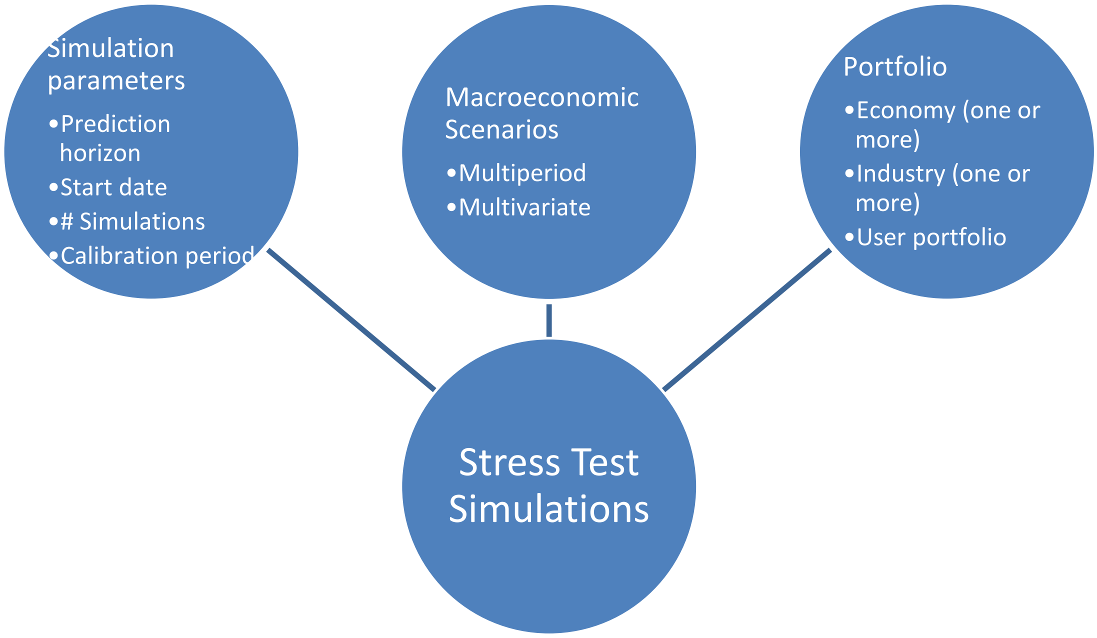
Sectors	Input Variables of the Default prediction model					
	DTD	LIQ	PROF	RSIZE	MB	SIGMA
Basic material	50.18%	2.04%	14.02%	20.53%	54.89%	42.69%
Communications	46.33%	15.26%	17.32%	35.74%	47.46%	20.62%
Consumer (cyclical)	43.38%	2.50%	43.65%	11.02%	52.07%	39.62%
Consumer (noncyclical)	38.11%	19.24%	18.29%	16.37%	53.18%	24.51%
Diversified	24.22%	25.32%	53.28%	35.73%	47.38%	15.34%
Energy	53.35%	26.08%	39.11%	36.94%	59.26%	40.13%
Financial	33.44%	46.27%	38.99%	15.23%	42.65%	50.18%
Industrial	52.67%	3.87%	13.32%	15.86%	57.34%	32.71%
Technology	44.31%	12.46%	21.65%	23.76%	52.72%	15.66%
Utilities	37.00%	1.92%	10.11%	18.66%	38.74%	41.15%

Table 10: Average in-sample 12-month R^2 for stress-testing regressions for Eurozone 12

Sectors	Input Variables of the Default prediction model					
	DTD	LIQ	PROF	RSIZE	MB	SIGMA
Basic material	38.66%	36.68%	34.32%	30.57%	40.00%	41.00%
Communications	36.12%	34.31%	26.77%	28.64%	33.65%	28.45%
Consumer (cyclical)	33.03%	30.65%	27.07%	34.95%	36.81%	34.98%
Consumer (noncyclical)	36.85%	28.35%	25.85%	31.06%	34.77%	34.57%
Diversified	39.05%	37.82%	34.85%	43.57%	39.61%	42.21%
Energy	43.00%	36.40%	36.40%	24.49%	43.08%	46.04%
Financial	31.28%	38.43%	33.15%	35.02%	35.71%	38.80%
Industrial	31.33%	32.59%	34.22%	28.78%	38.97%	39.09%
Technology	41.46%	36.39%	35.19%	26.05%	35.28%	37.61%
Utilities	40.08%	27.95%	25.59%	27.68%	33.41%	43.29%

The goodness of fit, measures by the R^2, ranges from over 20 percent to 50 percent, which is adequate when dealing with econometric time series data. The macroeconomic stress variables are able to explain the changes in the default covariates/predictors well.

Section 3: Operating the BuDA Matlab Toolkit



BuDA comes complete with a Matlab toolkit for users to easily conduct the scenario analysis described herein. It incorporates a user friendly frontend GUI with clickable options. This section describes how to setup and utilize the toolkit. For the best experience, it is advisable to use the latest version of Matlab (at least 2014 and above).

Setup

The toolkit is available as a package within a self-contained folder which can be conveniently copied. There is no need to run any installation program; all that needs to be done is to copy the entire folder with its contents.

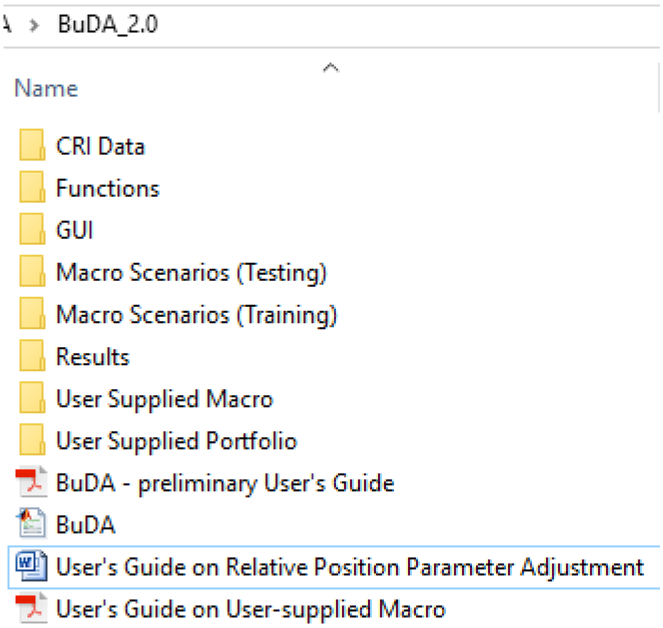


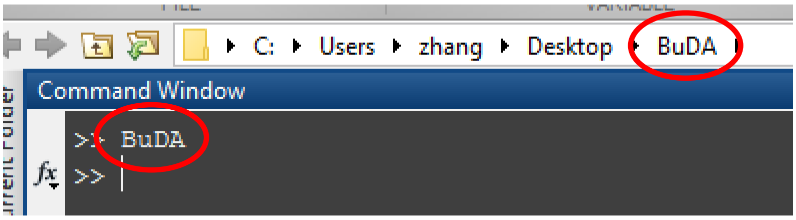
Table 11 below lists the key components of the BuDA tool.

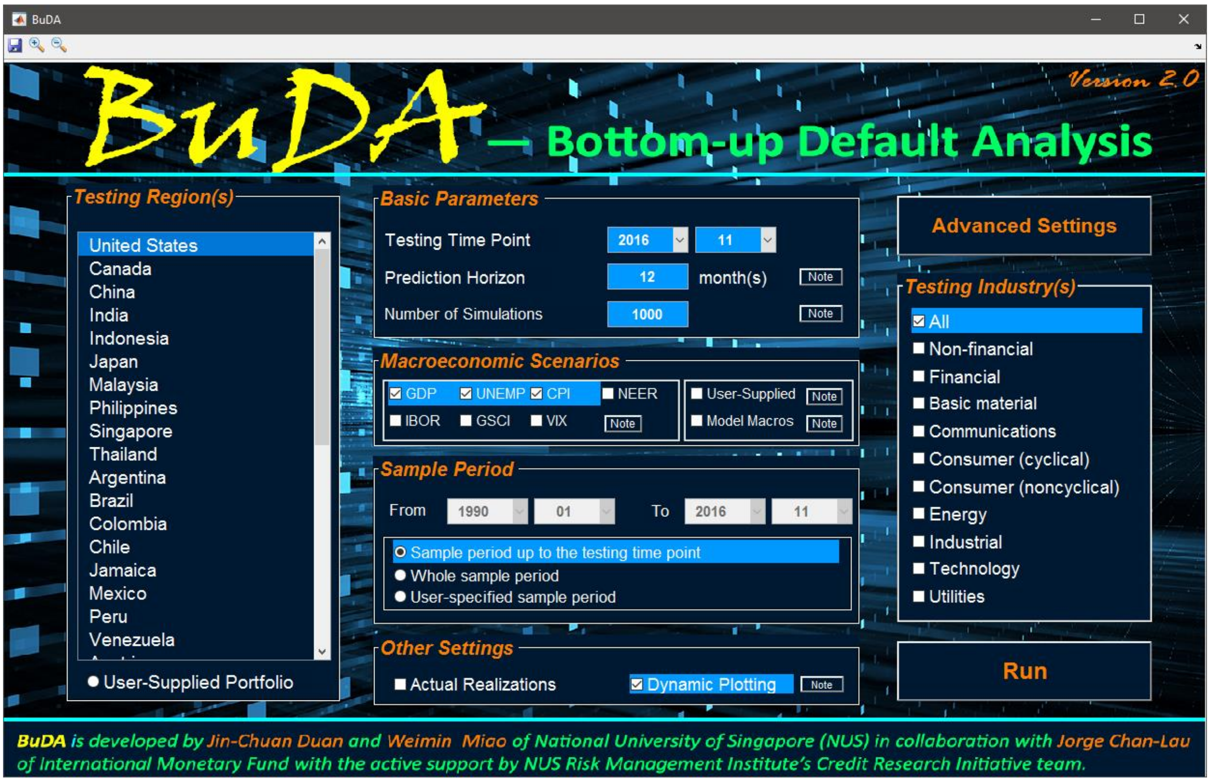
Table 11: Key Components of BuDA Toolkit

Components	**Description**	**Relevant Folders / Files**
Required		
Matlab codes	Codes needed to run the program. User only needs to interact with the GUI.	BuDA.mat, Functions, and GUI.
Granular PD data from RMI-CRI	Dataset of PDs and risk factors (e.g. DTD) from CRI database. They are updated upon request.	CRI Data
Macroeconomic historical data for training / calibration	Historical time series of macroeconomic data used by the program to calibrate the model.	Macro Scenarios (Training)
Macroeconomic stress scenarios	User-specified stress scenarios which the program will use for stress testing or scenario analysis.	Macro Scenarios (Testing)
Optional		
User-supplied macroeconomic variables	The program allows users to specify their desired macroeconomic stress variables (e.g. crude oil price) if the pre-set ones are not adequate (i.e. GDP, unemployment, inflation, etc, as described above).	User Supplied Macro
Model macros	User can also use the common risk factors in the PD model, i.e. stock index return and 3-momth interest, as the stress variables.	
User-supplied portfolios	The program allows users to specify their desired portfolio, which can be a single or a group of firms.	User Supplied Portfolio

Setting up a BuDA Session

To initiate the tool, simply start Matlab in the BuDA folder and type "BuDA" at the Matlab prompt. The user interface will then popup promptly.





It is quite intuitive, but let's discuss how to use this tool.

Select a Portfolio

The first order of business is to select the portfolio which we will simulate in the scenario analysis. On the left, firms in one or more economies may be selected for inclusion in the analysis. Hold down the shift or control buttons to make multiple selections. Also note that the right side allows the user to pick the industries to be included. Again, multiple selections are supported. The option "User-Supplied Portfolio" can instead be selected if the desired portfolio does not fit nicely into built-in options, in which case, the user needs to specify the file "User_Supplied_Portfolio_IDBB.xlsx" in the "User Supplied Portfolio" folder. The said file contains a list of company codes to be included.

Company codes are maintained by CRI and may be referenced off a readily available master list “Company_Information_Full.csv” in the “CRI Data” folder.

Set up the Basic Specifications

Next, the “Basic Parameters” are the simulation parameters which need to be defined. In the example here, the starting data or “Testing Time Point” is November 2016. Data from this point in time is used to initialize the simulations. The “Prediction Horizon” of the PD series can be from 1 month to 60 months (5 years). The example shows the use of PDs with a 12 months (1 year) forward-looking prediction horizon. “Number of Simulations” is defaulted to 1000. A greater number requires more time to run, but is expected to provide less statistical error. The “Sample Period” defines the historical period to be used for training the model and estimation of parameters. The default is to use all available historical data, but users may have reasons to calibrate to a particular period of interest and we afford that flexibility.

If “Dynamic Plotting” in the “Other Settings” panel is selected, a plot of the results will be shown on the screen after every batch of simulation is done, instead of showing it all at the end. Exactly what is plotted can be selected under “Advanced Settings”, which allows for mean, median, or a specific quantile. To be more precise, first and always, the median PD of the firms in the selected portfolio is computed after each simulation run. Then, according to the setting, the statistic – mean, median or quantile as specified in the settings – is taken across the 1000 simulations (or how many were specified) and plotted.

To assess the validity of the model, it is often useful to conduct backtesting. Examples of how backtesting results look like can be found in Figure 9. By selecting “Actual Realisations”, the program will run the simulation by assuming the stress scenario to be the one characterized by the actual realization of the macroeconomic variables. . Besides the PD computed under such a stress scenario, the output figure also includes an additional plot which shows how the PDs would have evolved under the actual (single) realisation of the default covariates. Naturally, this option will only work for historical analysis where the forecasting period is contained within the historical dataset.

Specify Macroeconomic Stress Variables

Users should define their “Macroeconomic Scenarios” by selecting or inputting their stress variables. Users can choose one of the following three ways to do so.

Seven Default Macro Variables
One can choose one or multiple of the 7 default macros (e.g. GDP growth rate, CPI growth rate, etc.) given in BuDA. In this case, the user needs to provide an Excel file in the “Macro Scenarios (Testing)” folder, specifying the prescribed future paths of the selected variables. The picture below shows one example of a presumed stressed macroeconomic scenario from 2015 onward defined only by GDP, unemployment and CPI.

.

United States								
Unit of measure: %								
All the macroeconomic scenarios for testing are assumed to be quarter-on-quarter, rather than year-on-year.								
year	quarter	GDP (growth rate)	UNEMP (difference)	CPI (growth rate)	NEER (growth rate)	IBOR (difference)	GSCI (growth rate)	VIX (growth rate)
2015	1	0.774700991	-0.1705	0.025989866				
2015	2	0.774700991	-0.1705	0.025989866				
2015	3	0.774700991	-0.1705	0.025989866				
2015	4	0.774700991	-0.1705	0.025989866				
2016	1	0.756375096	-0.0795	0.36944758				
2016	2	0.756375096	-0.0795	0.36944758				
2016	3	0.756375096	-0.0795	0.36944758				
2016	4	0.756375096	-0.0795	0.36944758				
2017	1	0.658957955	-0.03	0.588042722				
2017	2	0.658957955	-0.03	0.588042722				
2017	3	0.658957955	-0.03	0.588042722				
2017	4	0.658957955	-0.03	0.588042722				
2018	1	0.583620892	-0.0305	0.629285046				
2018	2	0.583620892	-0.0305	0.629285046				
2018	3	0.583620892	-0.0305	0.629285046				
2018	4	0.583620892	-0.0305	0.629285046				
2019	1	0.505898044	-0.01	0.577724211				
2019	2	0.505898044	-0.01	0.577724211				
2019	3	0.505898044	-0.01	0.577724211				
2019	4	0.505898044	-0.01	0.577724211				
2020	1	0.502204182	-0.00625	0.571580732				
2020	2	0.502204182	-0.00625	0.571580732				
2020	3	0.502204182	-0.00625	0.571580732				
2020	4	0.502204182	-0.00625	0.571580732				

In addition to the standard 7 scenario analysis variables, users have the option to include other variables.

Common Macros in the Underlying PD Model

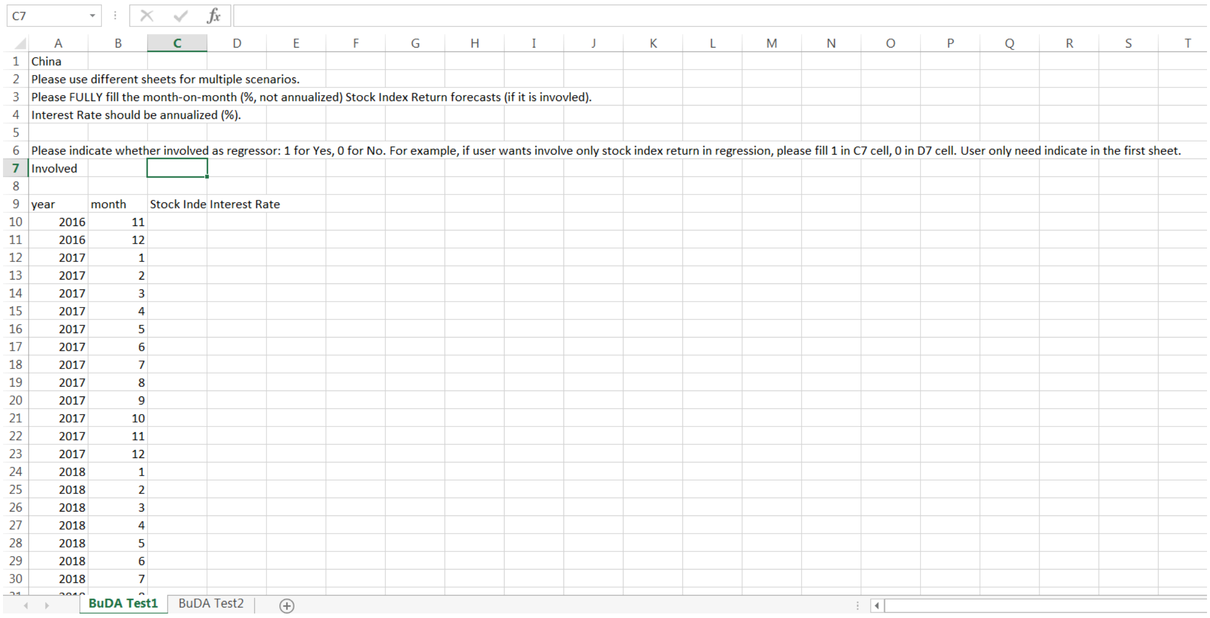
By checking “Model Macros”, users may include the stock index return and/or the 3-month interest rate as variables. This option was added as stock index return and 3-month interest rate are macroeconomic variables in the underlying PD model, and users may be interested in scenarios where they are in distress.

After the user clicks “Run”, the program will generate a new folder “\Internal Macros\Test_<Testing Time Point>_<Econ List>\”, which in turn contains 2 subfolders – one containing the data used for training, and another for the scenario testing. The training data is automatically populated and the user typically does not need to make any amendments.

It is necessary, however, for the user to supply the scenario which is to be tested. The program will pause to allow the user to specify the scenario, with the following prompt:

```
Please provide and save the testing scenarios of internal macros in:
C:\BuDA2.1\Internal Macros\Test_201611_2_9\Testing
then press any key but 'q' to continue:
```

User should then provide the scenario in "\Testing\macro_X_test.xlsx". Scenarios may be provided for either one or both variables (optional). To include a variable, indicate "1" in the "Involved" field. To exclude a variable, indicate "0". In the picture below, these fields correspond to cells C7 and D7 for the stock index return and interest rate respectively.



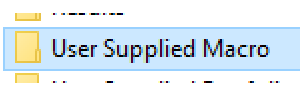
	A	B	C	D
1	China			
2	Please use different sheets for multiple scenarios.			
3	Please FULLY fill the month-on-month (%, not annualized) Stock Index Return forecasts (if it is invovled).			
4	Interest Rate should be annualized (%).			
5				
6	Please indicate whether involved as regressor: 1 for Yes, 0 for No. For example, if user wants involve only stock index return in regression, please fill 1 in C7 cell, 0 in D7 cell. User only need indicate in the first sheet.			
7	Involved			
8				
9	year	month	Stock Inde	Interest Rate
10	2016	11		
11	2016	12		
12	2017	1		
13	2017	2		
14	2017	3		
15	2017	4		
16	2017	5		
17	2017	6		
18	2017	7		
19	2017	8		
20	2017	9		
21	2017	10		
22	2017	11		
23	2017	12		
24	2018	1		
25	2018	2		
26	2018	3		
27	2018	4		
28	2018	5		
29	2018	6		
30	2018	7		

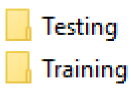
To define the scenario, stock index returns should be provided as month-on-month simple returns in column C. Interest rates should be provided as annualized 3-month rate (in %). Multiple scenarios can be provided by including more Excel sheets. As before, if more than 1 economy is selected, each economy should have the same number of scenarios.

Once the scenarios have been defined, the user may continue the program.

Custom Macroeconomic Stress Variables

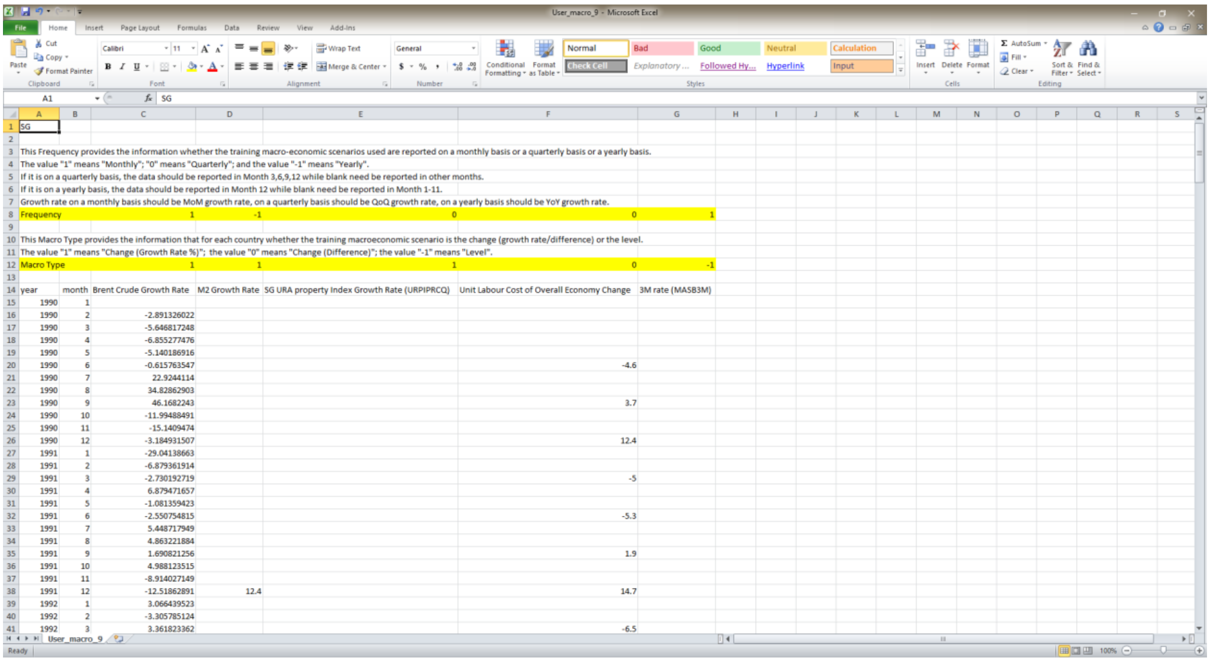
In addition, users may also specify custom macroeconomic stress variables. To do so, firstly check the "User-Supplied" checkbox. To setup the variables, access the "User Supplied Macro" folder. Within the folder are two subfolders "Training" and "Testing" which are used for training (calibrating) the model, and the stress testing / scenario analysis respectively.





Users specifying their own macroeconomic variables should note that one set of training and testing files should be prepared for *each* economy. For example, if both US and Canada are being analysed, then there should be four files in total, with each of the two countries having one training and one testing file each.

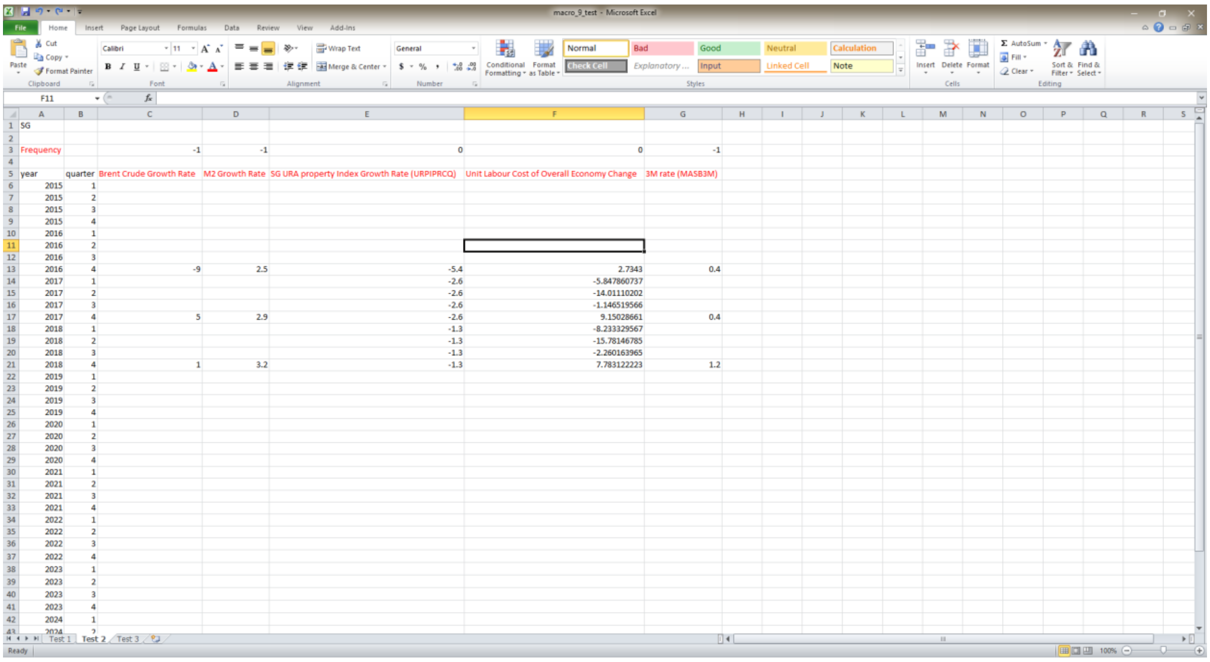
Instructions for setting up the training files are contained within the sample files. The program accommodates variables of different frequencies. This feature is implemented as a setting where users specify the value "1" for monthly, "0" for quarterly, and "-1" for yearly frequencies. In addition, the program allows the users to implement different specifications on the stress testing model by indicating the nature of the macro variables. An indicator of "1" means the macro series is recorded in "growth rate" (e.g. GDP growth rate), "0" refers to "change in difference" (e.g. change in unemployment rate), and "-1" refers to "the value itself".



	A	B	C	D	E	F	G
1	SG						
2							
3	This Frequency provides the information whether the training macro-economic scenarios used are reported on a monthly basis or a quarterly basis or a yearly basis.						
4	The value "1" means "Monthly"; "0" means "Quarterly"; and the value "-1" means "Yearly".						
5	If it is on a quarterly basis, the data should be reported in Month 3,6,9,12 while blank need be reported in other months.						
6	If it is on a yearly basis, the data should be reported in Month 12 while blank need be reported in Month 1-11.						
7	Growth rate on a monthly basis should be MoM growth rate, on a quarterly basis should be QoQ growth rate, on a yearly basis should be YoY growth rate.						
8	Frequency		1	-1	0	0	1
9							
10	This Macro Type provides the information that for each country whether the training macroeconomic scenario is the change (growth rate/difference) or the level.						
11	The value "1" means "Change (Growth Rate %)"; the value "0" means "Change (Difference)"; the value "-1" means "Level".						
12	Macro Type		1	1	1	0	-1
13							
14	year	month	Brent Crude Growth Rate	M2 Growth Rate	SG URA property Index Growth Rate (URPIPRCQ)	Unit Labour Cost of Overall Economy Change	3M rate (MASB3M)
15	1990	1					
16	1990	2	-2.891326022				
17	1990	3	-5.646817248				
18	1990	4	-6.855277476				
19	1990	5	-5.140186916				
20	1990	6	-0.615763547			-4.6	
21	1990	7	22.9244114				
22	1990	8	34.82862903				
23	1990	9	46.1682243			3.7	
24	1990	10	-11.99488491				
25	1990	11	-15.1409474				
26	1990	12	-3.184931507			12.4	
27	1991	1	-29.04138663				
28	1991	2	-6.879361914				
29	1991	3	-2.730192719			-5	
30	1991	4	6.879471657				
31	1991	5	-1.081359423				
32	1991	6	-2.550754815			-5.3	
33	1991	7	5.448717949				
34	1991	8	4.863221884				
35	1991	9	1.690821256			1.9	
36	1991	10	4.988123515				
37	1991	11	-8.914027149				
38	1991	12	-12.51862891	12.4		14.7	
39	1992	1	3.066439523				
40	1992	2	-3.305785124				
41	1992	3	3.361823362			-6.5	

Likewise, the testing files allow stress variables in quarterly or yearly frequencies, and the body contains the scenarios.

One thing to note when specifying the testing file: if firms from more than one country are selected, BuDA will conduct the scenario analysis separately for each country. For example, if 5 countries are selected, then the stress testing regressions are run for each of those 5 countries, and the PDs are predicted accordingly. This requires one testing file for each country involved in the selection of firms for a total of 5 files. While this may seem slightly tedious, it affords more flexibility than picking a single set of macroeconomic variables. The PDs across all countries are then pooled together to plot the final graph.



year	quarter	Brent Crude Growth Rate	M2 Growth Rate	SG URA property Index Growth Rate (URPIPRCQ)	Unit Labour Cost of Overall Economy Change	3M rate (MASB3M)
SG						
Frequency		-1	-1	0	0	-1
2015	1					
2015	2					
2015	3					
2015	4					
2016	1					
2016	2					
2016	3					
2016	4	-9	2.5	-5.4	2.7343	0.4
2017	1			-2.6	-5.847860737	
2017	2			-2.6	-14.01110202	
2017	3			-2.6	-1.146519566	
2017	4	5	2.9	-2.6	9.15028661	0.4
2018	1			-1.3	-8.233329567	
2018	2			-1.3	-15.78146785	
2018	3			-1.3	-2.260163965	
2018	4	1	3.2	-1.3	7.783122223	1.2
2019	1					
2019	2					
2019	3					
2019	4					
2020	1					
2020	2					
2020	3					
2020	4					
2021	1					
2021	2					
2021	3					
2021	4					
2022	1					
2022	2					
2022	3					
2022	4					
2023	1					
2023	2					
2023	3					
2023	4					
2024	1					

Advanced Setting

You might remember that we specified the scenario analysis regressions as AR(2) models, that is, the default covariates have 2 lags, and the BuDA tool defaults to this. More or fewer lags can be accommodated by changing the “Advanced Settings”, under the setting “Regression AR Lags”. Another option here is “Regression Aggregation Months” which specifies how much aggregation is done prior to estimating the scenario analysis regression parameters, which as mentioned above, alleviates the challenge with mixed-frequency data.

Recall that we use the trimmed mean firm-specific variables for the stress testing regressions, and track each firm’s response by regressing its individual firm-specific variables against the trimmed mean values[14]. This is automatic and no user input is required. However, if users wish to make adjustments to the regression parameters, this can be done by checking “Adjust Parameters after Regression” under “Advanced Settings”. If this is enabled, the simulation will run twice. During the first run, there will be no parameter adjustment. Before entering the second run, the program will pause and display “Please adjust and save the parameters in: Results\Test_201611_9\relative_position_regression then press any key but 'q' to continue: ”. Users may then manually edit the “FirmParainEcon**.xlsx” file in the said folder before proceeding. This is useful in the event that the user may have a better specification for the regression parameters[15].

By selecting “Yes” in the drop-down menu in the “Sensitivity Analysis” panel, the user can assess the contribution of each of the stress variables to the stressed PD if they are included in the stress testing equation separately. The user can also see the difference between the sum of the individual effects and the gross effect when all variables are accounted for simultaneously. We call this

[14] Recall that this is based on an AR(3) model.

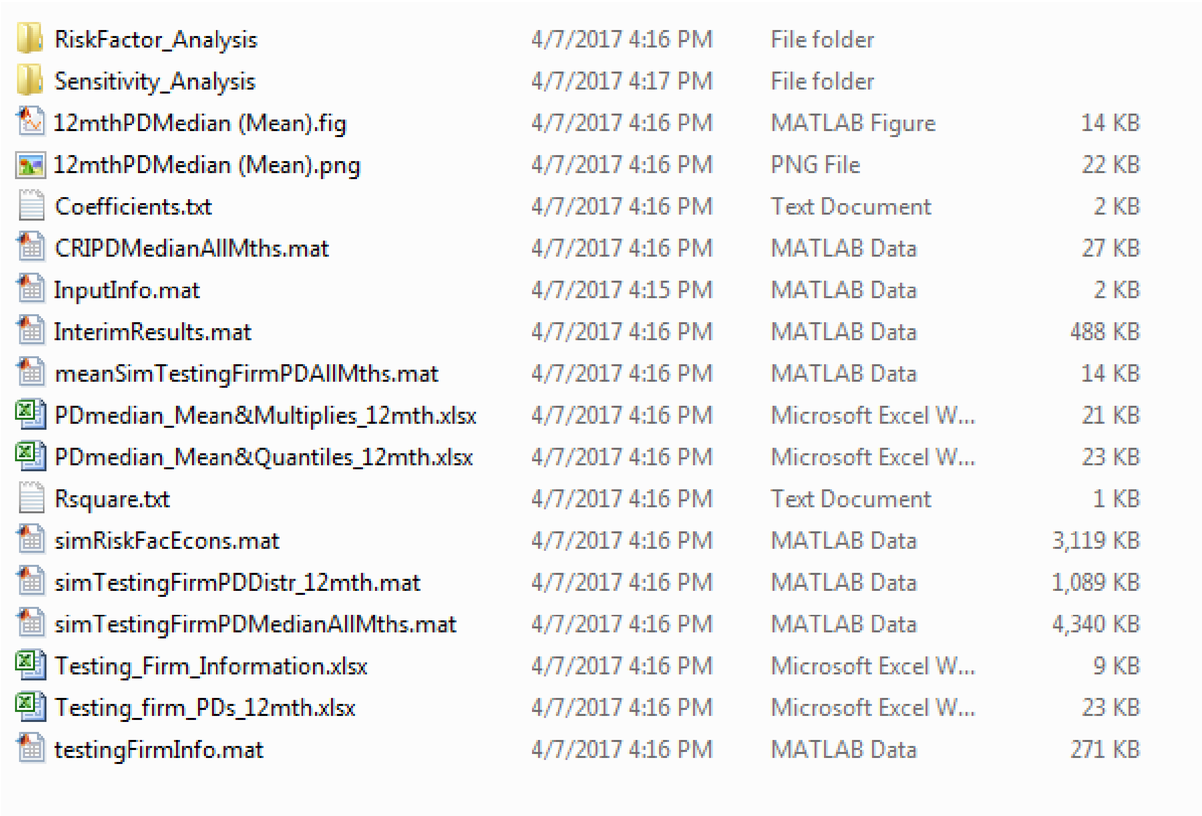
[15] For example, in some cases, individual level risk factors may decay too quickly in the estimation and users may like them to be more persistent.

difference the cross effect. The relevant output files and their interpretation are given in the “Results” section below.

Results

Once all parameters have been set, click the “Run” button at the bottom right of the GUI. Depending on the number of firms and economies being analysed, and number of simulations, running the program may take a few minutes to hours on a normal desktop.

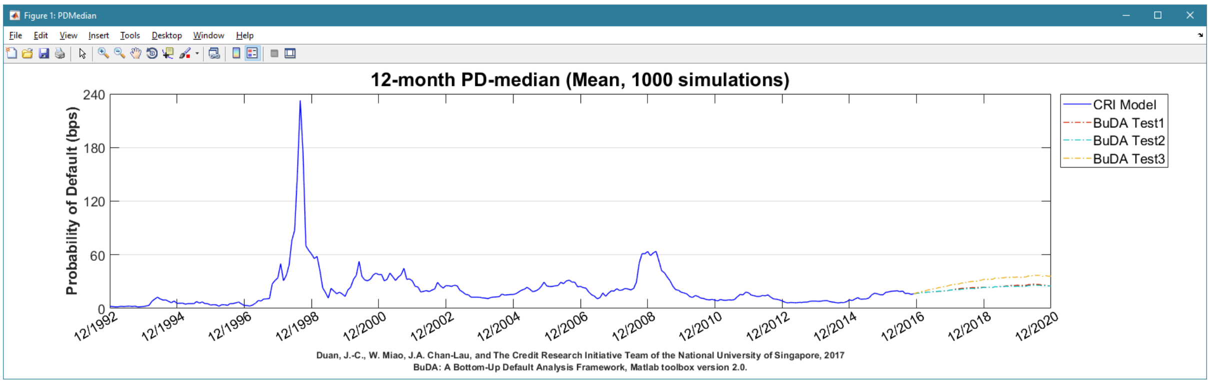
All results are stored in the “Results” folder, which has the structure as follows.



Among others, the most important output files for the user are the four Excel files and a plot:
a. **monthPDMedian (Mean).png
b. Testing_Firm_Information.xlsx
c. Testing_firm_PDs_**mth.xlsx
d. PDmedian_Mean&Multiplies_**mth.xlsx
e. PDMedian_Mean&Quantiles_**mth.xlsx

In **monthPDMedian (Mean).png, as Figure 10 shows, the "CRI Model" plot shows the PDs derived from the actual default covariates, while "BuDA Test" shows the result under the stress scenario. If more than one scenario has been specified, then additional plots "BuDA Test 2", "BuDA Test 3", etc.

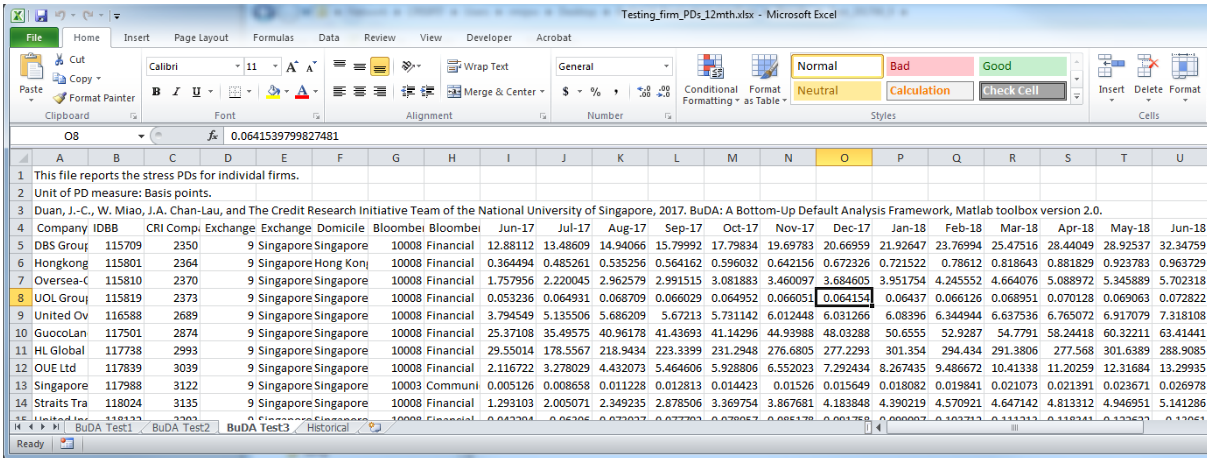
Figure 10: BuDA Scenario Analysis Results



Testing_Firm_Information contains for all firms used for testing the information on company name, the IDBB (Bloomberg ID), the CRI company ID, the firm's exchange country code, the firm's domicile, the Bloomberg industry code of the firm and the name of that industry.

	A	B	C	D	E	F	G	H
1	Company Name	IDBB	CRI Company ID	Exchange Country Code	Exchange Country	Domicile Country	Bloomber Industry Code	Bloomberg Industry
2	Creative Technology Ltd/Singapore	103117	18	9	Singapore	Singapore	10013	Technology
3	GL Ltd	113174	426	9	Singapore	Singapore	10004	Consumer Cyclical
4	DBS Group Holdings Ltd	115709	2350	9	Singapore	Singapore	10008	Financial
5	Dairy Farm International Holdings Ltd	115780	2354	9	Singapore	Hong Kong	10005	Consumer Non-cyclical
6	Mandarin Oriental International Ltd	115796	2359	9	Singapore	Hong Kong	10004	Consumer Cyclical
7	Jardine Strategic Holdings Ltd	115798	2361	9	Singapore	Hong Kong	10006	Diversified
8	Hongkong Land Holdings Ltd	115801	2364	9	Singapore	Hong Kong	10008	Financial
9	Singapore Airlines Ltd	115808	2368	9	Singapore	Singapore	10004	Consumer Cyclical
10	Oversea-Chinese Banking Corp Ltd	115810	2370	9	Singapore	Singapore	10008	Financial
11	UOL Group Ltd	115819	2373	9	Singapore	Singapore	10008	Financial
12	Haw Par Corp Ltd	116382	2491	9	Singapore	Singapore	10006	Diversified
13	Jardine Matheson Holdings Ltd	116423	2532	9	Singapore	Hong Kong	10006	Diversified
14	United Overseas Bank Ltd	116588	2689	9	Singapore	Singapore	10008	Financial
15	Keppel Corp Ltd	116942	2735	9	Singapore	Singapore	10006	Diversified

Testing_firm_PDs_**mth contains the historical month-end PDs for each of the abovementioned firms, which are used in the simulation, in the "Historical" tab. It also contains the stressed PDs in to the future in the "BuDA Test*" tab(s).



This file reports the stress PDs for individual firms.

Unit of PD measure: Basis points.

Duan, J.-C., W. Miao, J.A. Chan-Lau, and The Credit Research Initiative Team of the National University of Singapore, 2017. BuDA: A Bottom-Up Default Analysis Framework, Matlab toolbox version 2.0.

Company	IDBB	CRI Comp	Exchange	Exchange	Domicile	Bloomber	Bloomber	Jun-17	Jul-17	Aug-17	Sep-17	Oct-17	Nov-17	Dec-17	Jan-18	Feb-18	Mar-18	Apr-18	May-18	Jun-18
DBS Group	115709	2350	9	Singapore	Singapore	10008	Financial	12.88112	13.48609	14.94066	15.79992	17.79834	19.69783	20.66959	21.92647	23.76994	25.47516	28.44049	28.92537	32.34759
Hongkong	115801	2364	9	Singapore	Hong Kong	10008	Financial	0.364494	0.485261	0.535256	0.564162	0.596032	0.642156	0.672326	0.721522	0.78612	0.818643	0.881829	0.923783	0.963729
Oversea-C	115810	2370	9	Singapore	Singapore	10008	Financial	1.757956	2.220045	2.962579	2.991515	3.081883	3.460097	3.684605	3.951754	4.245552	4.664076	5.088972	5.345889	5.702318
UOL Group	115819	2373	9	Singapore	Singapore	10008	Financial	0.053236	0.064931	0.068709	0.066029	0.064952	0.066051	0.064154	0.06437	0.066126	0.068951	0.070128	0.069063	0.072822
United Ov	116588	2689	9	Singapore	Singapore	10008	Financial	3.794549	5.135506	5.686209	5.67213	5.731142	6.012448	6.031266	6.08396	6.344944	6.637536	6.765072	6.917079	7.318108
GuocoLan	117501	2874	9	Singapore	Singapore	10008	Financial	25.37108	35.49575	40.96178	41.43693	41.14296	44.93988	48.03288	50.6555	52.9287	54.7791	58.24418	60.32211	63.41441
HL Global	117738	2993	9	Singapore	Singapore	10008	Financial	29.55014	178.5567	218.9434	223.3399	231.2948	276.6805	277.2293	301.354	294.434	291.3806	277.568	301.6389	288.9085
OUE Ltd	117839	3039	9	Singapore	Singapore	10008	Financial	2.116722	3.278029	4.432073	5.464606	5.928806	6.552023	7.292434	8.267435	9.486672	10.41338	11.20259	12.31684	13.29935
Singapore	117988	3122	9	Singapore	Singapore	10003	Communi	0.005126	0.008658	0.011228	0.012813	0.014423	0.01526	0.015649	0.018082	0.019841	0.021073	0.021391	0.023671	0.026978
Straits Tra	118024	3135	9	Singapore	Singapore	10008	Financial	1.293103	2.005071	2.349235	2.878506	3.369754	3.867681	4.183848	4.390219	4.570921	4.647142	4.813312	4.946951	5.141286

PDmedian_Mean&Multiplies_12mth

- The reported are raw numbers
- column C shows 12 month CRI PD-medians for the dates given in columns A and B *(also available in Matlab file CRIPDMedianAllMths.mat as last time series)*
- column D shows the mean of the PD medians of the BuDA simulations *(also available in InterimResults.mat as last time series of meanSimTestingFirmPDMedianAllMths)*
- columns E to I show the probabilities that the BuDA PD simulations exceed different multiples of the CRI PD-medians *(also available in InterimResults.mat as last time series of multipleSimTestingFirmPDMedianAllMths)*
- e.g. In the example below, in November 2015, about 76 percent of the PD forecasts exceed the median PD value of 0.001989 in October 2015. But no simulated value exceeded two times the median PD value.

4	Year	Month	CRI	Mean	1-time prc	2-time prc	3-time prc	4-time prc	5-time probability
300	2015	8	0.001827						
301	2015	9	0.002015						
302	2015	10	0.001989						
303	2015	11		0.002208	0.758	0	0	0	0
304	2015	12		0.002542	0.868	0.011	0	0	0

PDmedian_Mean&Quantiles_12mth

- for this file the PDs are given in bps
- columns A to D as in PDmedian_Mean&Multiplies_12mth
- column E shows the median PD medians of the BuDA simulations *(this data is not directly available, but can be easily calculated from the variable simTestingFirmPDMedianAllMths in simTestingFirmPDMedianAllMths.mat)*
- columns F to I show different quantiles of the BuDA PD-median simulations *(also available in InterimResults.mat as last time series of quantileSimTestingFirmPDMedianAllMths)*
- e.g. in the example below, in December 2016, the 75% quantile of the simulated PDs is 0.0002 or 2bps.

4	Year	Month	CRI	Mean	Median	95%-quan	75%-quan	25%-quan	5%-quantile
295	2016	8	0.000121						
296	2016	9	0.00015						
297	2016	10	0.000121						
298	2016	11	0.000168						
299	2016	12		0.000177	0.000172	0.000258	0.000201	0.000146	0.000117
300	2017	1		0.000226	0.000216	0.000354	0.000267	0.00018	0.000133
301	2017	2		0.000245	0.000229	0.000404	0.000287	0.000181	0.000137
302	2017	3		0.000286	0.000268	0.000479	0.000346	0.000207	0.000146
303	2017	4		0.000308	0.000284	0.000535	0.000367	0.000227	0.000157

Some other useful files include:
InputInfo contains the parameters which the user selected on the user interface, e.g. number of simulation paths, testing time point, etc.

testingFirmInfo

Contains three tables:

- TestingFirmPD contains for each testing firm for each month in the sample period historical PDs with 1-60 month prediction horizons
- testingFirmList contains ID, Industry and economy ID for testing firms
- dateTestingFirmPD contains the dates for testing

CRIPDMedianAllMths contains two tables:

- CRIPDMedianAllMths contains the median CRI PDs at all time points of the sample period for the entire prediction horizon in one month increments
- dateCRIPDs contains the dates for the sample period

InterimResults contains a lot of data:

- comFacEcons: the common factor values for the economy for the entire history
- comFacForRegEcons: the common factor values used in the regression
- dateFacEcons: contains the date series
- dateTestingMacroEcons: contains the dates used for testing
- dateTrainingMacroEcons: contains the dates used for training
- meanSimTestingFirmPDMedianAllMths: contains BuDA means for all firms and predicting horizons and scenarios
- quantileSimTestingFirmPDMedianAllMths: contains BuDA quantiles for all firms and predicting horizons and scenarios
- regResComFacEcons: contains the regression results for the common factors
- regResRelPosEcons: contains the regression results for the relative positions
- regResComSpecEcons: contains the regression results for the firm-specific factors
- specFacEcons: the firm-specific factor values for the economy
- specFacForRegEcons: the firm-specific factor values used in the regression
- specFacMapEcons: contains ones for mapping

- testingFirmListEcons: Like the firm list in the monthly calibration, contains ID, Industry and economy
- testingFirmSpecRelPosRecentEcons contains the firm specific information for all testing firms in the most recent 2 years.
- testingInduCodesEcons: contains the industry codes used for testing
- testingMacroEcons: contains macro scenario testing data
- trainingMacroEcons: contains macro scenario training data for the entire history
- trainingMacroForRegEcons: contains macro scenario training data used for regression

<u>Coefficients</u> is a text file containing the estimation of the stress testing regression, i.e. the loadings of the default covariates (DTD, CASH/TA, NI/TA, etc.) on the stress variables (solely GDP in this example).

```
Coefficients in Stress Testing Regressions:
------------------------------------------------------------------------------------------------------------------------
----------------------------------------
SINGAPORE
Common variables
                   Stock Index Return  Interest Rate
Intercept          -0.0043             -0.0253
GDP                0.0277              0.0271
Lag 1              0.1043              -0.0314
Lag 2              -0.2331             -0.0386

Sectors
Basic Materials
                   DTD        CASH/TA    NI/TA      SIZE       M/B        SIGMA
Intercept          0.1203     0.0039     0.0001     -0.0124    0.0666     0.0092
GDP                0.0671     0.0013     0.0001     0.0150     0.0184     -0.0037
Lag 1              0.0782     -0.1538    0.0313     -0.2167    -0.1023    0.1693
Lag 2              -0.1498    0.1175     -0.0825    0.1949     0.0305     -0.2060

Communications
                   DTD        CASH/TA    NI/TA      SIZE       M/B        SIGMA
Intercept          0.1675     0.0131     0.0002     -0.0024    0.1051     0.0108
GDP                0.0408     0.0011     0.0001     0.0155     0.0278     -0.0022
Lag 1              0.0558     -0.1223    0.0116     -0.1216    0.0942     0.2267
Lag 2              -0.1224    0.0528     -0.0832    0.0802     -0.1791    -0.2804

Consumer, Cyclical
                   DTD        CASH/TA    NI/TA      SIZE       M/B        SIGMA
Intercept          0.0762     0.0052     0.0001     -0.0077    0.0675     0.0077
```

<u>Rsquare</u>, as the name suggests, contains the R-squared values for the same set of regressions. Users may wish to view these TXT files in Notepad++ or a landscape view in word document with Courier New font for proper alignment.

```
R^2 in Stress Testing Regressions:
------------------------------------------------------------------------------------------------------------------------
SINGAPORE
Common variables
Stock Index Return      0.5937
Interest Rate           0.3273

Sectors                 DTD        CASH/TA    NI/TA      SIZE       M/B        SIGMA
Basic Materials         0.5175     0.2655     0.1880     0.1893     0.3627     0.2758
Communications          0.3836     0.3384     0.1707     0.2742     0.4139     0.3446
Consumer, Cyclical      0.2920     0.2830     0.3410     0.1001     0.3868     0.2710
Consumer, Non-Cyclical  0.3014     0.3657     0.3106     0.2180     0.3989     0.2772
Diversified             0.2796     0.2433     0.4039     0.3270     0.2370     0.2384
Energy                  0.2997     0.2466     0.2848     0.1738     0.2788     0.2785
Financial               0.2339     0.1577     0.5347     0.2202     0.2438     0.3310
Industrial              0.2909     0.5254     0.3666     0.0806     0.2101     0.3530
Technology              0.3460     0.1773     0.3031     0.1115     0.3875     0.2855
Utilities               0.2978     0.6036     0.4256     0.0767     0.3498     0.3511

Duan, J.-C., W. Miao, J.A. Chan-Lau, and The Credit Research Initiative Team of the National University of Singapore, 2017
BuDA: A Bottom-Up Default Analysis Framework, Matlab toolbox version 2.0.
```

<u>simRiskFacEcons</u>

Contains four tables:

- simComFacEcons contains for each simulation the results for the common factors
- simSpecFacEcons contains for each simulation the results for the firm-specific factors

- specFacMapEcons: whether observations for a risk factor, say DTD, are sufficient (i.e. at least 5 firms in this industry with 60 observations) for calculating its average value for a particular industry, say utility. 1 means data are sufficient, while 0 means otherwise. In the case of 0, the average DTD for the utility sector is obtained by averaging the mean DTD across other sectors.
- testingEconCodes: contains econ number

<u>meanSimTestingFirmPDAllMths</u>

Contains two tables:

- meanSimTestingFirmPDAllMths contains the mean PDs at all time points of the forecast horizon in one month increments for all testing firms
- dateSimPDs contains the dates for the forecast horizon

<u>simTestingFirmPDMedianAllMths</u>

Contains two tables:

- simTestingFirmPDMedianAllMths contains the median of all firm PDs for all months until the prediction horizon
- dateSimPDs contains the dates for the forecast horizon

<u>simTestingFirmPDDistr_**mth</u>

Contains three tables:

- simTestingFirmPDDistrPredictMth contains for each simulation and scenario the PDs results for the prediction horizon
- testingFirmList contains ID, Industry and economy for testing firms
- dateSimPDs contains the dates for the forecast horizon

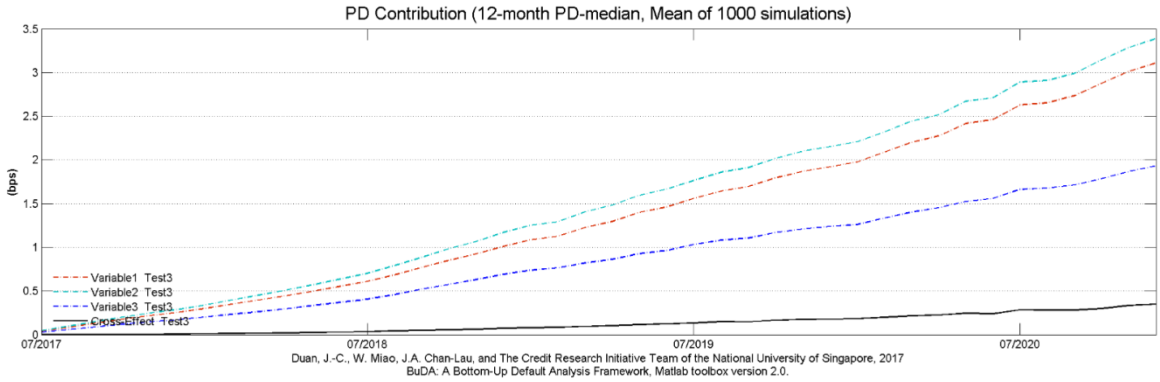
<u>12mthPDMedian (Mean)</u> contains the summary plot as per Figure 10. The PNG file can be used readily as a picture when the FIG file can be opened in Matlab for manipulation.

In the <u>"RiskFactor_Analysis"</u> folder, "RiskFactorAnalysis [econ] [industry].png" displays the simulated paths of the average risk factors under the specified scenarios. "RiskFactorAnalysis [econ] [industry].xlsx stores both the historical and simulated values for those risk factors under various scenarios.

In the <u>"Sensitivity_Analysis"</u> folder, "Testing_firm_PDs_**mth.xlsx" stores simulated PDs when only one stress variable moves as the scenario defines, while others stay constant. In "SensitivityAnalysis_Test*_**mthPDMedian (Mean).png", the difference between those projected values and the values under the "zero-change" scenario are displayed. Here, "zero-change" scenario means all state variables defining the scenario stay constant in the years to come, for example the GDP is fixed for the next five years. Notation wise, PD_{flat} is the PD in the zero-change scenario, and PD_1, PD_2 and PD_3 are the stressed PDs when only one of them moves as prescribed. PD_{All} is the PD when all variables are considered simultaneously.

In the sample figure below, the difference between the red curve and the X-axis is the contribution of variable 1 to the PD, i.e. $PD_1 - PD_{flat}$. The difference between the green curve and the X-axis is

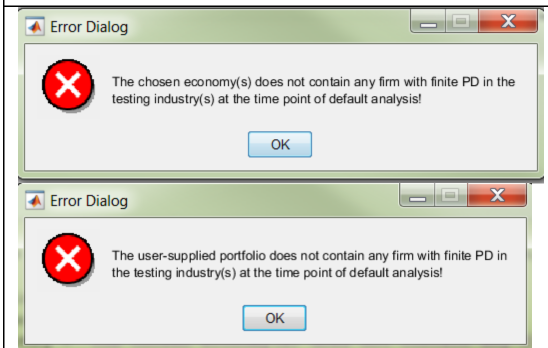
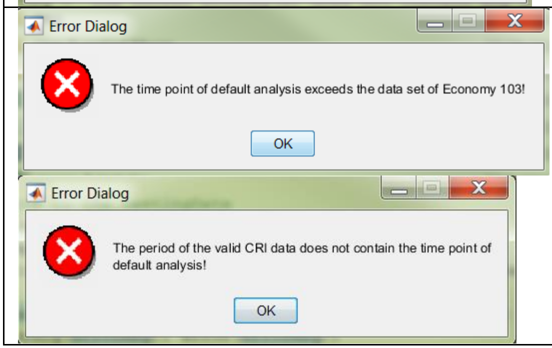
$PD_1 + PD_2 - 2PD_{flat}$, which is the combined contribution of variable 1 and 2, when they are included into the stress testing separately. Likewise, the difference between the blue curve and the X-axis is $PD_1 + PD_2 + PD_3 - 3PD_{flat}$. The blue curve falls below the green and red in this case, because the contribution of variable 3 to the PD is actually negative. The cross-effect, represented by the distance between the black curve and the X-axis, is the difference between the PD when stress variables are considered separately and that when all variables are accounted for simultaneously. Mathematically, it is $PD_1 + PD_2 + PD_3 - 2PD_{flat} - PD_{All}$.

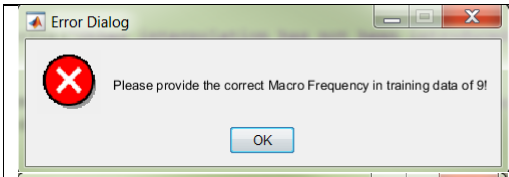
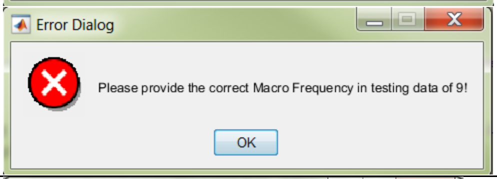
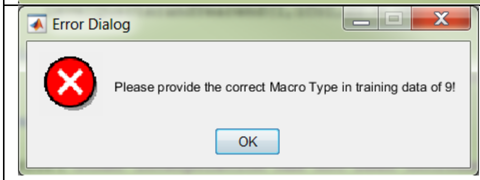
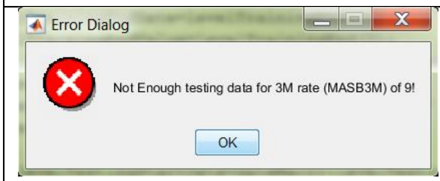
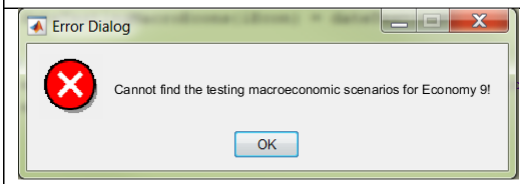
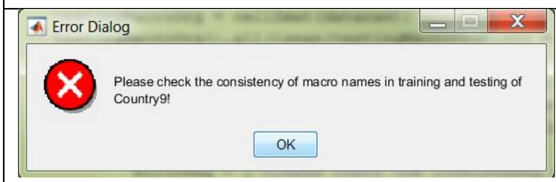
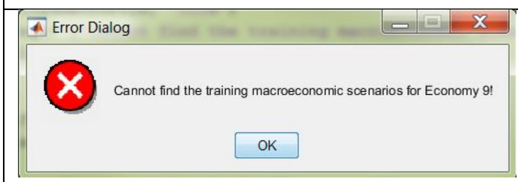


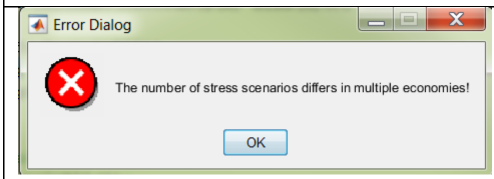
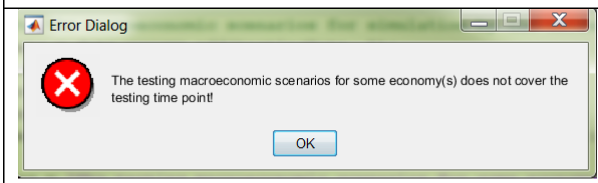
Troubleshooting

The program contains a number of checks to ensure that the input is sensible. Table 12 below describes the common errors and explanation of possible reasons that should be addressed.

Table 12: Common Errors and Explanations

Error	Explanation
	These errors occur when the choice of firms to be analysed to not have valid PDs. This should not happen if the CRI data is included correctly. To fix this, ensure that the PD data is in the correct folder and that the files have not been corrupted. Examples: Specified firm identities wrongly when defining user-supplied portfolio, did not download and extract “CRI Data” files correctly.
	These errors appear when the “Testing Time Point” in the “Basic Parameters” in the GUI is not in the time range of the CRI data set. Example: If data for US is only available up to September 2016, but the user selects a future date, say December 2016.

	BuDA uses "1" for "Monthly", "0" for "Quarterly", "-1" for "Yearly" to represent frequencies in the macroeconomic data. The error appears when the frequency is given incorrectly. Example: User inputs "2", which is not a valid input.
	BuDA uses "1" for "Change (Growth Rate %)", "0" for "Change (Difference)", "-1" for "Level" to represent macro type in User-Supplied Macro data. The error appears when the type is given incorrectly in training data. Example: User inputs "2", which is not a valid input.
	The error appears when the testing data is inadequate for the macroeconomic variable. This only applies for user-supplied macroeconomic variables. Example: User specifies that a particular macroeconmic variable is to be used in the training file, but there are no valid values in the time series.
	This error appears when the testing file in "Macro Scenarios (Testing)" folder (for default seven macros) or "User Supplied Macro \Testing" folder (for user-supplied macros) is not correctly provided or named. Example: "macro_test_9.xlsx" was accidentally deleted or renamed and user tries to run the simulation for Singapore.
	BuDA requires that the name of the macros be same in respective training and testing file for user-supplied macros. The error appears when they are inconsistent. Example: User wishes to run the simulation for Singapore. In "\User Supplied Macro\Training \User_macro_9", the macrovariable was named "Brent Crude Growth Rate", but in "\User Supplied Macro\Testing\macro_9_test" it has a different name, say "Brent Crude Growth".
	This error appears when the training file in "Macro Scenarios (Training)" folder (for default seven macros) or "User Supplied Macro \Training" folder (for user-supplied macros) is not correctly provided or named.

	Example: User selects “United States” in “Testing Region(s)”, but the file “User_macro_15” does not exist or has been accidentally renamed.
	The error appears when the user-supplied portfolio is not correctly provided or named. Example: “User_Supplied_Portfolio_IDBB” in “\User Supplied Portfolio” has been accidentally renamed or deleted.
	For stress testing of multiple economies, BuDA requires that the number of stress scenarios should be same. The error appears when they are different. Example: User selects both “United States” and “Canada” in “Testing Regions(s)”, but in “\Macro Scenarios (Testing)”, “macro_test_15” contains 3 scenarios (Excel tabs) while “macro_test_16” contains only 2 scenarios.
	In testing file, the date should cover the quarter containing the next month of Testing Time Point. The error appears if it is not satisfied. Example: User wishes to run the simulation for Singapore. “Testing Time Point” in the GUI is set to 2016 November, but “\Macro Scenarios (Testing)\ macro_test_9” does not contain data for 2016/Q4. The error message can also appear if one fails to prescribe the future paths for all stress variables that she selects on the user interface. For example, the person chooses all of the seven stress variables on the GUI but only provides projections for three of them in the testing file.

Section 4: IMF Case Studies

References

1. Basel Committee on Banking Supervision, 2005b, Studies on the Validation of Internal Rating Systems (revised), Working Paper No. 14.

2. Bharath, S. T., & Shumway, T., 2008, Forecasting default with the Merton distance to default model, *Review of Financial Studies*, 21(3), 1339-1369.

3. Crosbie, Peter J., and Jeffrey R. Bohn, 2001, Modeling Default Risk, KMV, LLC, San Francisco, CA.

4. Duan, J.-C., Sun, J., and Wang, T., 2012, Multiperiod corporate default prediction – a forward intensity approach*, Journal of Econometrics* 170(1), 191-209.

5. Duan, J.-C., Miao, W., & Wang, T., 2015, Stress Testing with a Bottom-Up Corporate Default Prediction Model, RMI Working Paper.

6. Duan, J.-C., Wang, T., 2012, Measuring Distance-to-Default for Financial and Non-Financial Firms, *Global Credit Review* 2, 95-108.

7. Duffie, D., Saita, L., & Wang, K., 2007, Multi-period corporate default prediction with stochastic covariates, *Journal of Financial Economics*, 83(3), 635-665.

8. Financial Crisis Inquiry Commission, 2011, The Financial Crisis Inquiry Report: Final Report of the National Commission on the Causes of the Financial and Economic Crisis in the United States, Washington, D. C.: Government Printing Office.

9. Merton, R. C., 1974, On the pricing of corporate debt: The risk structure of interest rates, *The Journal of Finance*, 29(2), 449-470.

10. NUS-RMI Staff, 2016, NUS-RMI Credit Research Initiative Technical Report Version: 2016 Update 1, *Global Credit Revi*ew 6, 49-132.

11. NUS-RMI Staff, 2016, Addendum 1* to the CRI Technical Report (Version: 2016, Update 1), CRI website.

12. NUS-RMI Staff, 2016, Addendum 2 to the CRI Technical Report (Version: 2016, Update 1), CRI website.

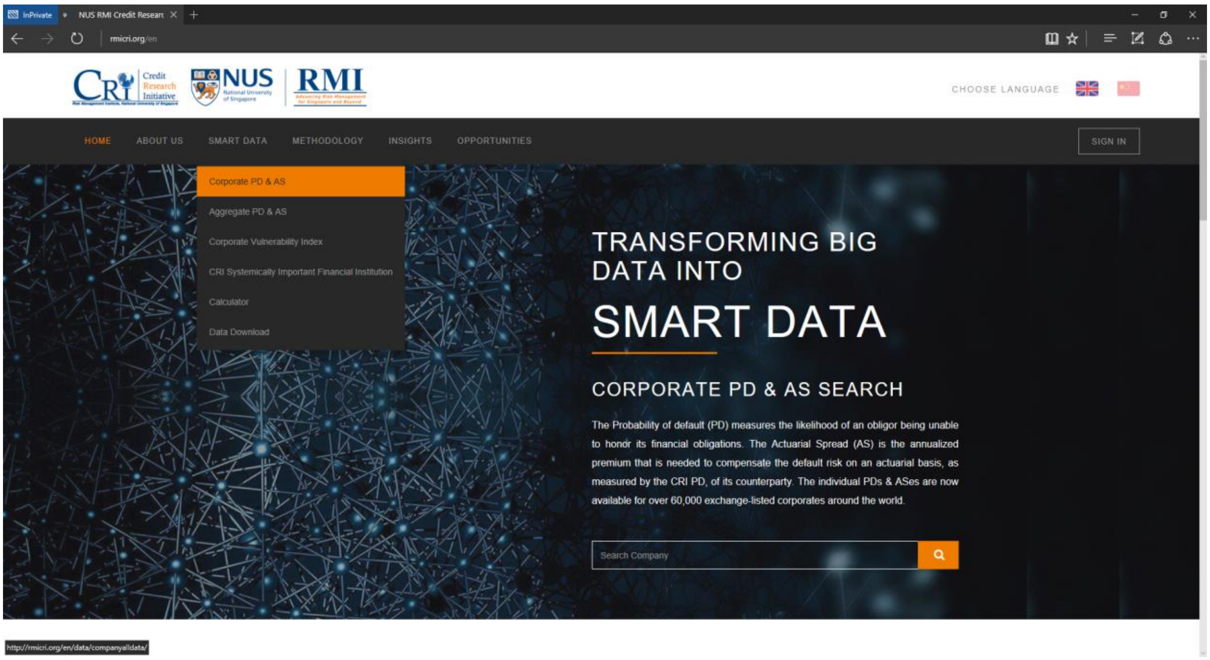
13. Vassalou, M., & Xing, Y., 2004, Default risk in equity returns, *Journal of Finance*, 59(2), 831-868.

14. Zou, Q., 2016, Essays on Distress Risk (Essay 2), National University of Singapore Department of Finance PhD Thesis.

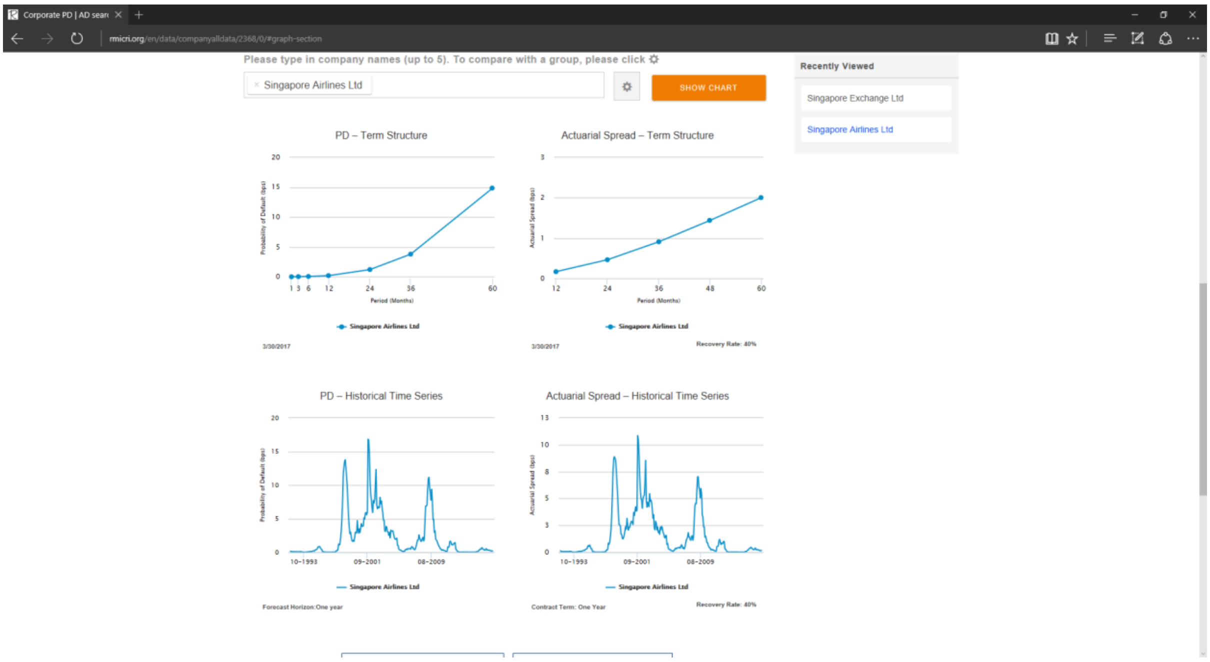
Appendices

Accessing PDs on CRI Website

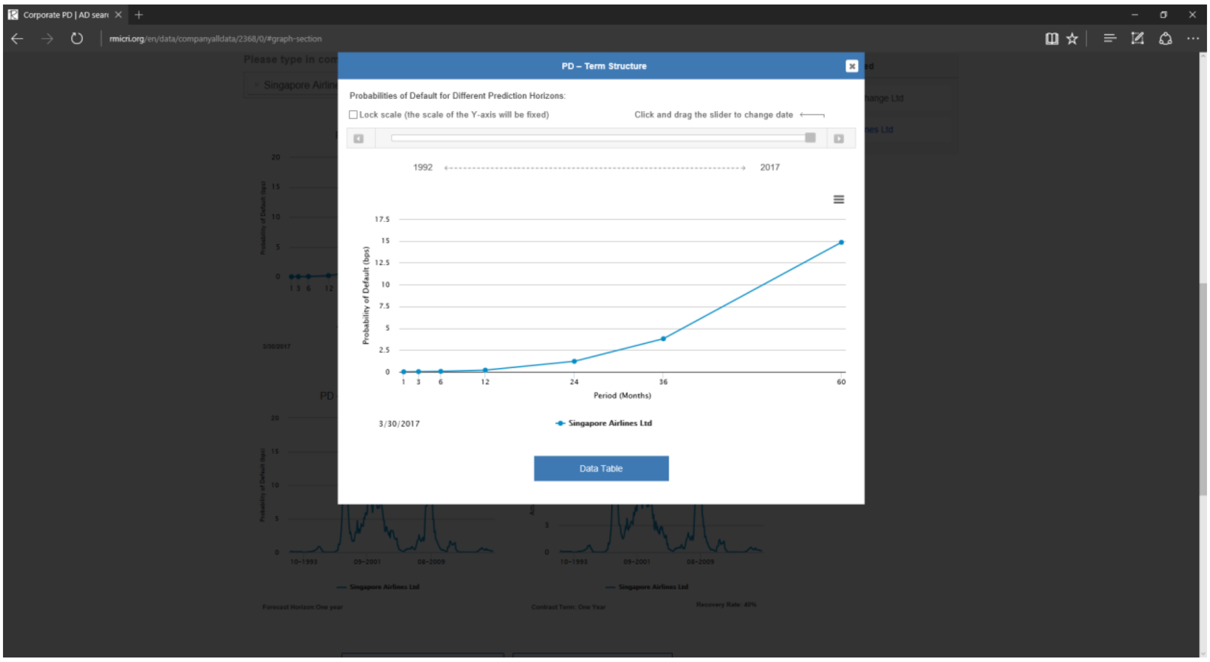
The CRI website provides a front facing interface for accessing the PDs in a user-friendly manner. To begin, setup a free account at http://rmicri.org. This account will provide the user with access to not only CRI's PDs, but also to our other research outputs including actuarial spreads, corporate vulnerability indices and CriSIFI.



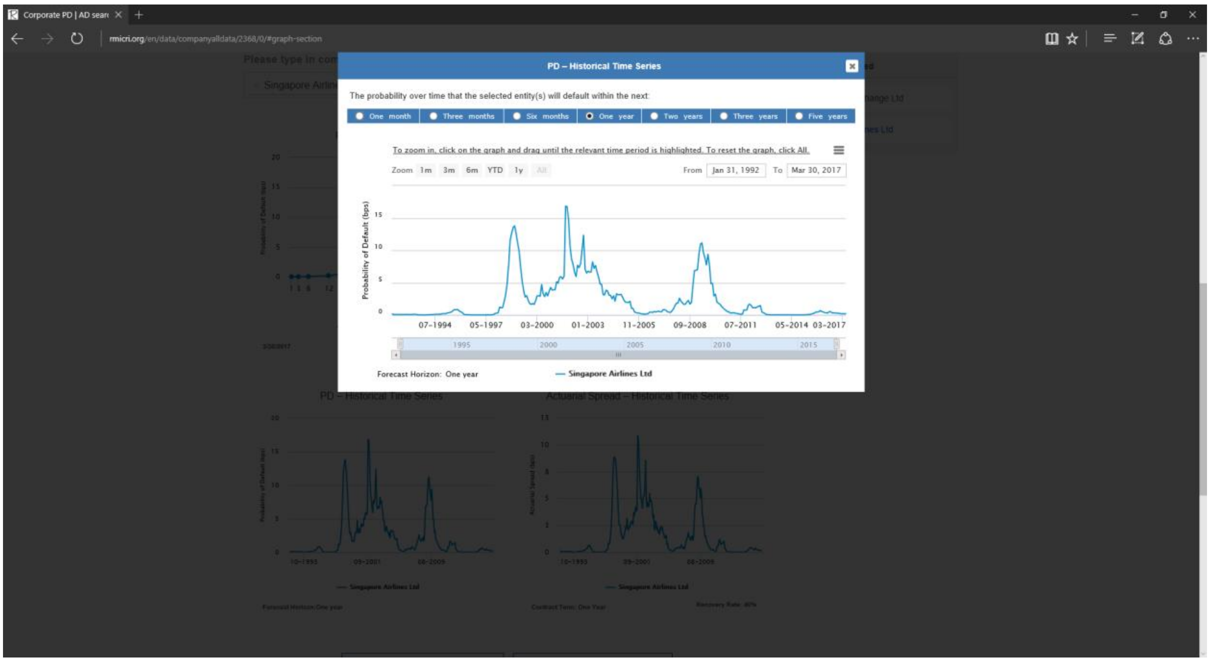
Once the account has been setup, select one of the options under the "Smart Data" tab. "Company PD and AS" provides forecasts for individual firms while "Aggregate PD and AS" provides the same set of charts for the aggregate of an economy, region, or sector.



Four charts show an overview of the firm selected. We shall focus on the two on the left – PD Term Structure and PD Historical Time Series. Each of these charts may be clicked for further options.



Let's start with the PD Term Structure chart. This chart shows the PDs for a range of forward-looking horizons starting from 1 month up to 5 years. The question is, "How likely is it for the firm to default in the next 1 month, 3 month, 6 months, 1 year, 2 years 3 years and 5 years?" Each of these is represented as a point on the chart. Moving the top slider allows the user to visualize the curve at different historical points in time. As can be seen from the screenshot, we are currently looking at the latest point in time since the slider is all the way to the right. To obtain the precise numerical values for each plotted point, the user may click on the "Data Table" button. Printing and downloading options are available through the menu on the top right indicated by the hamburger icon (three stacked lines).



The PD Historical Time Series chart shows the time profile for a selected forecast horizon. In the example, historical values of the 1-year ahead PDs are plotted. The desired horizon can be changed with the row of tabs located at the top and the user can also zoom in for higher resolution on selected segments of the chart.

PDs are available for bulk download via the “Smart Data” → “Data Download” tab. The historical panel of individual firms and their term-structures may be downloaded. For even lower level data such as input covariates and intermediate variables, direct access to the database is possible. Please feel free to approach one of our CRI staff for assistance.