# NUS-RMI Credit Rating Initiative

# Technical Report Version: 2012 update 1 *(02-07-2012)*

Global Credit Review
Vol. 2 (2012) 109–138

DOI: S2010493612500079

# NUS-RMI Credit Research Initiative Technical Report

# Version: 2012 update 1

**RMI staff article**
For any questions or comments on this article, please contact Oliver Chen at
ochen@nus.edu.sg

## INTRODUCTION

This document describes the implementation of the system which the NUS Risk Management Institute's Credit Research Initiative uses to produce probability of default (PDs). As of this version of the Technical Report, these PDs cover exchange listed firms in 37 economies in Asia, Asia-Pacific, North America, Western Europe and Latin America. Currently, RMI covers 35,000 listed companies and individual company PDs are computed daily. 2,300 of these firms' default forecasts are freely available to all users at http://www.rmi.nus.edu.sg/cri, along with aggregate PDs at the economy and sector level for all the firms.

The primary goal of this initiative is to drive research and development in the critical area of credit rating systems. As such, a transparent methodology is essential to this initiative. Having the details of the methodology available to everybody means that there is a base from which suggestions and improvements can be made. The objective of this Technical Report is to provide a full exposition of the CRI system. Readers of this document who have access to the necessary data and who have a sufficient level of technical expertise will be able to implement a similar system on their own. For a full exposition of the conceptual framework of the CRI, see Duan and Van Laere (2012).

The system used by the CRI will evolve as new innovations and enhancements are applied. The most substantial changes to the 2011 technical report and operational implementation of our model are (1) the default definition which now excludes covenant breaches and some default corporate actions that are specific to Taiwan (e.g., bounced checks); (2) priority of financial statements and treatment of net income, with the latter now being included on a quarterly basis when available; (3) treatment of stale market capitalization prices; (4) regrouping of economies for calibration purposes; and (5) increased coverage to include Latin America. This version of the technical report provides an update on the operational implementation of the CRI and includes all changes to the system that had been implemented by May 2012. The latest version of the Technical Report is available via the web portal and will include any changes to the system that have been implemented since the printing of this version.

The remainder of this Technical Report is organized as follows. The next section describes the quantitative model that is currently used to compute PDs from the CRI. The model was first described in Duan *et al*. (2012). The description includes calibration procedures, which are performed on a monthly basis, and individual firm PD computations, which are performed on a daily basis.

Section 2 describes the input variables of the model as well as the data used to produce the variables for input into the model. This model uses both input variables that are common to all firms in an economy and input variables that are firm-specific. Another critical component when calibrating a probability of default estimation system is the default data, and this is also described in this section.

While Section 1 provides a broader description of the model, Section 3 describes the implementation details that are necessary to apply given real world issues of, for example, bad or missing data. The specific technical details needed to develop an operational system are also given, including details on the monthly calibration, daily computation of individual firm PDs and aggregation of the individual firm PDs. Distance-to-default (DTD) in a Merton-type model is one of the firm-specific variables. The calculation for DTD is not the standard one, and has been modified to allow a meaningful computation of the DTD for financial firms. While most academic studies on default prediction exclude financial firms from consideration, it is important to include them given that the financial sector is a critical component in every economy. The calculation for DTD is detailed in this section.

Section 4 shows an empirical analysis for those economies that are currently covered. While the analysis shows excellent results in several economies, there is room for improvement in a few others. This is because, at the CRI's current stage of development, the economies all use the variables used in the academic study of US firms in Duan *et al*. (2012). Future development within the CRI will deal with variable selection specific to different economies, and the performance is then expected to improve. Variable selection and other planned developments are discussed in Section 5.

## I. MODEL DESCRIPTION

The quantitative model that is currently being used by the CRI is a forward intensity model that was introduced in Duan *et al*. (2012). This model allows probability of default forecasts to be made at a range of horizons. In the current CRI implementation of this model, PDs are made from a horizon of one month up to a horizon of two years. In other words, for every firm, the probability of that firm defaulting within one month, three months, six months, one year, eighteen months and two years is given. The ability to assess credit quality for different horizons is a useful tool for risk management, credit portfolio management, policy setting and regulatory purposes, since short- and long-term credit risk profiles can differ greatly depending on a firm's liquidity, debt structures and other factors.

The forward intensity model is a reduced form model in which the probability of default is computed as a function of different input variables. These can be firm-specific or common to all firms within an economy. The other category of default prediction model is the structural model, whereby the corporate structure of a firm is modeled in order to assess the firm's probability of default.

A similar reduced form model by Duffie *et al*. (2007) relied on modeling the time series dynamics of the input variables in order to make PD forecasts for different horizons. However, there is little consensus on assumptions for the dynamics of variables such as accounting ratios, and the model output will be highly dependent on these assumptions. In addition, the time series dynamics will be of very high dimension. For example, with the two common variables and two firm-specific variables that Duffie *et al*. (2007) use, a sample of 10,000 firms gives a dimension of the state variables of 20,002.

Given the complexity in modeling the dynamics of variables such as accounting ratios, this model will be diffcult to implement if different forecast horizons are required. The key innovation of the forward intensity model is that PD for different horizons can be consistently and effciently computed based only on the value of the input variables at the time the prediction is made. Thus, the model specification becomes far more tractable.

Fully specifying a reduced form model includes the specification of the function that computes a PD from the input variables. This function is parameterized, and finding appropriate parameter values is called calibrating the model. The forward intensity model can be calibrated by maximizing a pseudo-likelihood function. The calibration is carried out by economy and all firms within an economy will use the same parameter values along with each firm's variables in order to compute the firm's PD.

Subsection 1.1 will describe the modeling framework, including the way PDs are computed based on a set of parameter values for the economy and a set of input variables for a firm. Subsection 1.2 explains how the model can be calibrated.

## 1.1. Modeling Framework

While the model can be formulated in a continuous time framework, as done in Duan *et al*. (2012), an operational implementation will require discretization in time. Since the model is more easily understood in discrete time, the following exposition of the model will begin in a discrete time framework.

Variables for default prediction can have vastly different update frequencies. Financial statement data is updated only once a quarter or even once a year, while market data like stock prices are available at frequencies of seconds. A way of compromising between these two extremes is to have a fundamental time period $\Delta t$ of one month in the modeling framework. As will be seen later, this does not preclude updating the PD forecasts on a daily basis. This is important since, for example, large daily changes in a firm's stock price can signal changes in credit quality even when there is no change in financial statement data.

Thus, for the purposes of calibration and subsequently for computing time series of PD, the input variables at the end of each month will be kept for each firm. The input variables associated with the $i^{\text{th}}$ firm at the end of the $n^{\text{th}}$ month (at time $t = n\Delta t$) is denoted by $X_i(n)$. This is a vector consisting of two parts: $X_i(n) = (W(n), U_i(n))$. Here, $W(n)$ is a vector of variables at the end of month $n$ that is common to all firms in the economy and $U_i(n)$ is a vector of variables specific to firm $i$.

In the forward intensity model, a firm's default is signaled by a jump in a Poisson process. The probability of a jump in the Poisson process is determined by the intensity of the Poisson process. The forward intensity model draws an explicit dependence of intensities at time periods in the future (that is, forward intensities) to the value of input variables at the time of prediction. With forward intensities, PDs for any forecast horizon can be computed knowing only the value of the input variable at the time of prediction, without needing to simulate future values of the input variables.

There is a direct analogy in interest rate modeling. In spot rate models where dynamics on a short-term spot rate are specified, bond pricing requires expectations on realizations of the short rate. Alternatively, bond prices can be computed directly if the forward rate curve is known.

One issue in default prediction is that firms can exit public exchanges for reasons other than default. For example, in mergers and acquisitions involving two public companies, there will be one company that delists from its stock exchange. This is important in predicting defaults because a default cannot happen if a firm has been previously delisted. An exception is if the exit is a distressed exit and is followed soon after by a credit event. See Subsection 2.4 for details on how this case is handled in the CRI system.

In order to take these other exits into account, defaults and other exits are modeled as two independent Poisson processes, each with their own intensity. While defaults and exits classified as non-defaults are mutually exclusive by definition, the assumption of independent Poisson processes does not pose a problem since the probability of a simultaneous jump in the two Poisson processes is negligible. In the discrete time framework, the probability of simultaneous jumps in the same time interval is non-zero. As a modeling assumption, a simultaneous jump in the same time interval by both the default Poisson process and the non-default type exit Poisson process is considered as a default. In this way, there are three mutually exclusive possibilities during each time interval: survival, default and non-default exit. As with defaults, the forward intensity of the Poisson process for other exits is a function of the input variables. The parameters of this function can also be calibrated.

To further illustrate the discrete framework, the three possibilities for a firm at each time point are diagrammed. Either the firm survives for the next time period $\Delta t$, or it defaults within $\Delta t$, or it has a non-default exit within $\Delta t$. This setup is pictured in Figure 1. Information about firm $i$ is known up until time $t = m\Delta t$ and the figure illustrates possibilities in the future between $t = (n - 1)\Delta t$ and $(n + 1)\Delta t$. Here, $m$ and $n$ are integers with $m < n$.

The probabilities of each branch are, for example: $p_i(m, n)$ the conditional probability viewed from $t = m\Delta t$ that firm $i$ will default before $(n + 1)\Delta t$, conditioned on firm $i$ surviving up until $n\Delta t$. Likewise, $\bar{p}_i(m, n)$ is the conditional probability viewed from $t = m\Delta t$ that firm $i$ will have a non-default exit before $(n + 1)\Delta t$, conditioned on firm $i$ surviving up until $n\Delta t$. It is the modeler's objective to determine $p_i(m, n)$ and $\bar{p}_i(m, n)$, but for now it is assumed that these quantities are known. With the conditional default and other exit probabilities known, the corresponding conditional survival probability of firm $i$ is $1 - p_i(m, n) - \bar{p}_i(m,n)$.

With this diagram in mind, the probability that a particular path will be followed is the product of the conditional probabilities along the path. For example, the probability at time $t = m\Delta t$ of firm $i$ surviving until $(n - 1)\Delta t$ and then defaulting between $(n - 1)\Delta t$ and $n\Delta t$ is:

$$Prob_{t=m\Delta t}[\tau_i = n, \tau_i < \bar{\tau}_i] = p_i(m, n-1) \prod_{j=m}^{n-2} [1 - p_i(m, j) - \bar{p}_i(m, j)]. \quad (1)$$

Here, $\tau_i$ is the default time for firm $i$ measured in units of months, $\bar{\tau}_i$ is the other exit time measured in units of months, and the product is equal to one if there are no terms in the product. The condition $\tau_i < \bar{\tau}_i$ is the requirement that the firm defaults before it has a non-default type of exit. Note that by measuring exits in units of months, if, for example, a default occurs at any time in the interval $((n - 1)\Delta t, n\Delta t]$ then $\tau_i = n$.

Using equation (1), cumulative default probabilities can be computed. At $m\Delta t$ the probability of firm $i$ defaulting at or before $n\Delta t$ and not having

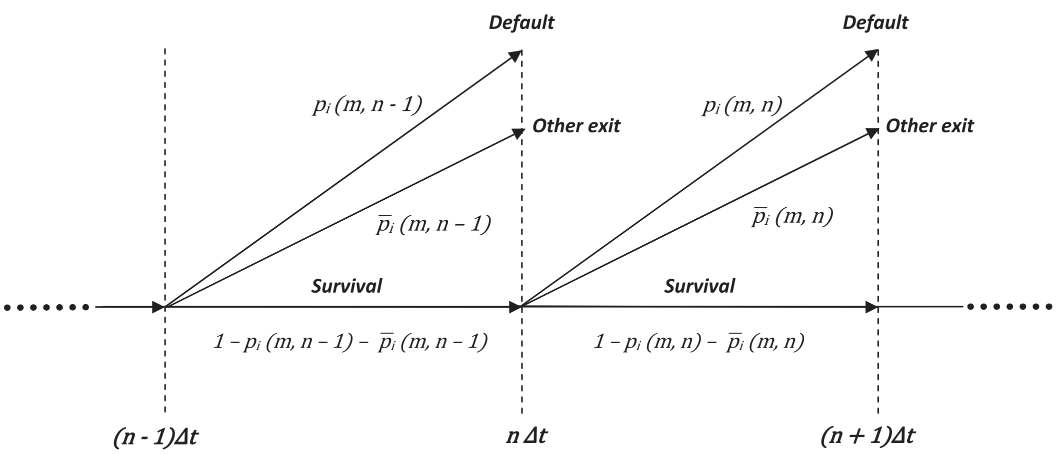


**Figure 1.** Default-other exit-survival tree for firm $i$, viewed from time $t = m\Delta t$.

another exit before $t = n\Delta t$ is obtained by taking the sum of all of the paths that lead to default at or before $n\Delta t$:

$$Prob_{t=m\Delta t}[m < \tau_i \le n, \tau_i < \bar{\tau}_i] = \sum_{k=m}^{n-1} \left\{ p_i(m,k) \prod_{j=m}^{k-1} [1 - p_i(m,j) - \bar{p}_i(m,j)] \right\}. \quad (2)$$

While it is convenient to derive the probabilities given in equations (1) and (2) in terms of the conditional probabilities, expressions for these in terms of the forward intensities need to be found, since the forward intensities will be functions of the input variable $X_i(m)$. The forward intensity for the default of firm $i$ that is observed at time $t = m\Delta t$ for the forward time interval from $t = n\Delta t$ to $(n + 1)\Delta t$, is denoted by $h_i(m, n)$ where $m \le n$. The corresponding forward intensity for a non-default exit is denoted in $\bar{h}_i(m, n)$. Because default is signaled by a jump by a Poisson process, its conditional probability is a simple function of its forward intensity:

$$p_i(m,n) = 1 - \exp[-\Delta t\, h_i(m,n)]. \quad (3)$$

Since joint jumps in the same time interval are assigned as defaults, the conditional other exit probability needs to take this into account:

$$\bar{p}_i(m,n) = \exp[-\Delta t\, h_i(m,n)]\{1 - \exp[-\Delta t\, \bar{h}_i(m,n)]\}. \quad (4)$$

The conditional survival probabilities in equations (1) and (2) are computed as the conditional probability that the firm does not default in the period and the firm does not have a non-default exit either:

$$Prob_{t=m\Delta t}[\tau_i, \bar{\tau}_i > n+1 \,|\, \tau_i, \bar{\tau}_i > n] = \exp\{-\Delta t[h_i(m,n) + \bar{h}_i(m,n)]\}. \quad (5)$$

It remains to specify the dependence of the forward intensities on the input variable $X_i(m)$. The forward intensities need to be positive so that the conditional probabilities are non-negative. A standard way to impose this constraint is to specify the forward intensities as exponentials of a linear combination of the input variables:

$$\begin{aligned} h_i(m,n) &= \exp[\beta(n-m) \cdot Y_i(m)], \\ \bar{h}_i(m,n) &= \exp[\bar{\beta}(n-m) \cdot Y_i(m)]. \end{aligned} \quad (6)$$

Here, $\beta$ and $\bar{\beta}$ are coefficient vectors that are functions of the number of months between the observation date and the beginning of the forward period $(n - m)$, and $Y_i(m)$ is simply the vector $X_i(m)$ augmented by a preceding unit element: $Y_i(m) = (1, X_i(m))$. The unit element allows the linear combination in the argument of the exponentials in equation (6) to have a non-zero intercept.

In the current implementation of the forward intensity model in the CRI, the maximum forecast horizon is 24 months and there are 12 input variables plus the intercept. So there are 24 sets of each of the coefficient vectors denoted by $\beta(0), \ldots, \beta(23)$ and $\bar{\beta}(0), \ldots, \bar{\beta}(23)$ and each of these coefficient vectors has 13 elements. While this is a large set of parameters, as will be seen in the next part, the calibration is tractable because the parameters for each horizon can be done independently from each other, and the default parameters can be calibrated separately from the other exit parameters.

Before giving the probabilities in (1) and (2) in terms of the forward intensities, a notation is introduced for the forward intensities that makes clear which parameters are needed for the forward intensity in question:

$$H(\beta(n-m), X_i(m)) := \exp[\beta(n-m) \cdot Y_i(m)]. \quad (7)$$

This is the forward default intensity. The corresponding notation for other exit forward intensities is then just $H(\bar{\beta}(n-m), X_i(m))$. So, the probability in (1) is expressed in terms of the forward intensities, using

(3) for the conditional default probability and (5) for the conditional survival probability:

$$
\begin{aligned}
&Prob_{t=m\Delta t}[\tau_i = n, \tau_i < \bar{\tau}_i] \\
&\quad = \{1-\exp[-\Delta t\, H(\beta(n-1-m), X_i(m))]\} \\
&\qquad \times \prod_{j=m}^{n-2} \exp\{-\Delta t\,[H(\beta(j-m), X_i(m)) \\
&\qquad + H(\bar{\beta}(j-m), X_i(m))]\} \\
&\quad = \{1-\exp[-\Delta t\, H(\beta(n-m-1), X_i(m))]\} \\
&\qquad \times \exp\left\{-\Delta t \sum_{j=m}^{n-2} [H(\beta(j-m), X_i(m)) \right. \\
&\qquad\qquad \left. + H(\bar{\beta}(j-m), X_i(m))]\right\}. \quad (8)
\end{aligned}
$$

This probability will be relevant in the next part during the calibration. The cumulative default probability given in equation (2) in terms of the forward intensities is then:

$$
\begin{aligned}
&Prob_{t=m\Delta t}[m < \tau_i \le n, \tau_i < \bar{\tau}_i] \\
&\quad = \sum_{k=m}^{n-1} \left\{ \{1-\exp[-\Delta t\, H(\beta(k-m), X_i(m))]\} \right. \\
&\quad \times \exp\left\{-\Delta t \sum_{j=m}^{k-1} [H(\beta(j-m), X_i(m)) \right. \\
&\qquad \left.\left. + H(\bar{\beta}(j-m), X_i(m))] \right\}\right\}. \quad (9)
\end{aligned}
$$

This formula is used to compute the main output of the CRI: an individual firm's PD within various time horizons. The $\beta$ and $\bar{\beta}$ parameters are obtained when the firm's economy is calibrated, and using those together with the firm's input variables yields the firm's PD.

## 1.2. Model Calibration

The empirical dataset used for calibration can be described as follows. For the economy as a whole, there are $N$ end of month observations, indexed as $n = 1, \ldots, N$. Of course, not all firms will have observations for each of the $N$ months as they may start later than the start of the economy's dataset or they may exit before the end of the economy's dataset. There are a total of $I$ firms in the economy, and they are indexed as $i = 1, \ldots, I$. As before, the input variables for the $i^{th}$ firm in the $n^{th}$ month is $X_i(n)$. The set of all observations for all firms is denoted by $X$.

In addition, the default times $\tau_i$ and non-default exit times $\bar{\tau}_i$ for the $i^{th}$ firm are known if the default or other exit occurs after time $t = \Delta t$ and at or before $t = N\Delta t$. The possible values for $\tau_i$ and $\bar{\tau}_i$ are integers between 2 and $N$, inclusive. If a firm exits before the month end, then the exit time is recorded as the first month end after the exit. If the firm does not exit before $t = N\Delta t$, then the convention can be used such that both of these values are infinite. If the firm has a default type of exit within the dataset, then $\bar{\tau}_i$ can be considered as infinite. If instead the firm has a non-default type of exit within the dataset, then $\tau_i$ can be considered as infinite. The set of all default times and non-default exit times for all firms is denoted by $\tau$ and $\bar{\tau}$, respectively. The first month in which firm $i$ has an observation is denoted by $t_{0i}$. Except for cases of missing data, these observations continue until the end of the dataset if the firm never exits. If the firm does exit, the last needed input variable $X_i(n)$ is for $n = \min(\tau_i, \bar{\tau}_i) - 1$.

The calibration of the $\beta$ and $\bar{\beta}$ parameters is done by maximizing a pseudo-likelihood function. The function to be maximized violates the standard assumptions of likelihood functions, but Appendix A in Duan *et al*. (2012) derives the large sample properties of the pseudo-likelihood function.

In formulating the pseudo-likelihood function, the assumption is made that the firms are conditionally independent from each other. In other words, correlations arise naturally from sharing common factors $W(n)$ and any correlations there are between different firms' firm-specific variables. With this assumption, the pseudo-likelihood function for a horizon of $\ell$ months, a set of parameters $\beta$ and $\bar{\beta}$ and the dataset $(\tau, \bar{\tau}, X)$ is:

$$\mathcal{L}_\ell(\beta, \bar{\beta}; \tau, \bar{\tau}, X) = \prod_{m=1}^{N-\ell} \prod_{i=1}^{I} P_\ell(\beta, \bar{\beta}; \tau_i, \bar{\tau}_i, X_i(m)). \quad (10)$$

Here, $P_\ell(\beta, \bar{\beta}; \tau_i, \bar{\tau}_i, X_i(m))$ is a probability for firm $i$, with the nature of the probability depending on what happens to the firm during the period from month $m$ to month $m + \ell$. This is defined as:

$$
\begin{aligned}
&P_\ell(\beta,\bar{\beta};\tau_i,\bar{\tau}_i,X_i(m))\\
&=1_{\{t_{0i}\le m,\min(\tau,\bar{\tau})>m+\ell\}}\\
&\times\exp\left\{-\Delta t\sum_{j=0}^{\ell-1}[H(\beta(j),X_i(m))+H(\bar{\beta}(j),X_i(m))]\right\}\\
&+1_{\{t_{0i}\le m,\tau_i\le\bar{\tau}_i,\tau_i\le m+\ell\}}\{1-\exp[-\Delta t\, H(\beta(\tau_i-m-1),X_i(m))]\}\\
&\times\exp\left\{-\Delta t\sum_{j=0}^{\tau_i-m-2}[H(\beta(j),X_i(m))+H(\bar{\beta}(j),X_i(m))]\right\}\\
&+1_{\{t_{0i}\le m,\bar{\tau}_i\le\tau_i,\bar{\tau}_i\le m+\ell\}}\{1-\exp[-\Delta t\, H(\bar{\beta}(\bar{\tau}_i-m-1),X_i(m))]\}\\
&\times\exp[-\Delta t\, H(\beta(\tau_i-m-1),X_i(m))]\\
&\times\exp\left\{-\Delta t\sum_{j=0}^{\bar{\tau}_i-m-2}[H(\beta(j),X_i(m))+H(\bar{\beta}(j),X_i(m))]\right\}\\
&+1_{\{t_{0i}>m\}}+1_{\{\min(\tau_i,\bar{\tau}_i)\le m\}}. \qquad (11)
\end{aligned}
$$

In words, if firm $i$ survives from the observation time at month $m$ for the full horizon $\ell$ until at least $m + \ell$, then the probability is the model-based survival probability for this period. This is the first term in (11). The second term handles the cases where the firm has a default within the horizon, in which case the probability is the model-based probability of the firm defaulting at the month that it ends up defaulting, as given in equation (8). The third term handles the cases where the firm has a non-default exit within the horizon, in which case the probability is the model-based probability of the firm having a non-default type exit at the month that the exit actually does occur. The expression for this probability uses the conditional non-default type exit probability given in equation (4). The final two terms handle the cases where the firm is not in the data set at month $m$ — either the first observation for the firm is after $m$ or the firm has already exited. A constant value is assigned in this case so that this firm will not affect the maximization at this time point.

The pseudo likelihood function given in (10) can be numerically maximized to give estimates for the coefficients $\beta$ and $\bar{\beta}$. Notice though that the sample observations for the pseudo-likelihood function are overlapping if the horizon is longer than one month. For example, when $\ell = 2$, default over the next two periods from month $m$ is correlated to default over the next two periods from month $m + 1$ due to the common month in the two sample observations. However, in Appendix A of Duan *et al*. (2012), the maximum pseudo-likelihood estimator is shown to be consistent, in the sense that the estimators converge to the "true" parameter value in the large sample limit.

It would not be feasible to numerically maximize the pseudo-likelihood function using the expression given in (11), due to the large dimension of the $\beta$ and $\bar{\beta}$ parameters. Notice though that each of the terms in (11) can be written as a product of terms containing only $\beta$ and terms containing only $\bar{\beta}$. This will allow separate maximizations with respect to $\beta$ and with respect to $\bar{\beta}$.

The $\beta$ and $\bar{\beta}$ specific versions of (11) are:

$$
\begin{aligned}
&P_\ell^\beta(\beta;\tau_i,\bar{\tau}_i,X_i(m))\\
&=1_{\{t_{0i}\le m,\min(\tau,\bar{\tau}_i)>m+\ell\}}\exp\left\{-\Delta t\sum_{j=0}^{\ell-1}H(\beta(j),X_i(m))\right\}\\
&\quad+1_{\{t_{0i}\le m,\tau_i\le\bar{\tau}_i,\tau_i\le m+\ell\}}\exp\left\{-\Delta t\sum_{j=0}^{\tau_i-m-2}H(\beta(j),X_i(m))\right\}\\
&\quad\times\{1-\exp[-\Delta t\, H(\beta(\tau_i-m-1),X_i(m))]\}\\
&\quad+1_{\{t_{0i}\le m,\bar{\tau}_i\le\tau_i,\bar{\tau}_i\le m+\ell\}}\exp\left\{-\Delta t\sum_{j=0}^{\bar{\tau}_i-m-2}H(\beta(j),X_i(m))\right\}\\
&\quad\times\exp[-\Delta t\, H(\beta(\tau_i-m-1),X_i(m))]\\
&\quad+1_{\{t_{0i}>m\}}+1_{\{\min(\tau_i,\bar{\tau}_i)\le m\}},\\
&P_\ell^{\bar{\beta}}(\bar{\beta};\tau_i,\bar{\tau}_i,X_i(m))\\
&=1_{\{t_{0i}\le m,\min(\tau,\bar{\tau})>m+\ell\}}\exp\left\{-\Delta t\sum_{j=0}^{\ell-1}H(\bar{\beta}(j),X_i(m))\right\}\\
&\quad+1_{\{t_{0i}\le m,\tau_i\le\bar{\tau}_i,\tau_i\le m+\ell\}}\exp\left\{-\Delta t\sum_{j=0}^{\tau_i-m-2}H(\bar{\beta}(j),X_i(m))\right\}\\
&\quad+1_{\{t_{0i}\le m,\bar{\tau}_i\le\tau_i,\bar{\tau}_i\le m+\ell\}}\exp\left\{-\Delta t\sum_{j=0}^{\bar{\tau}_i-m-2}H(\bar{\beta}(j),X_i(m))\right\}\\
&\quad\times\{1-\exp[-\Delta t\, H(\bar{\beta}(\bar{\tau}_i-m-1),X_i(m))]\}\\
&\quad+1_{\{t_{0i}>m\}}+1_{\{\min(\tau_i,\bar{\tau}_i)\le m\}}. \qquad (12)
\end{aligned}
$$

Then, the $\beta$ and $\bar{\beta}$ specific versions of the pseudo-likelihood function are given by:

$$\begin{aligned}\mathcal{L}_\ell^{\beta}\left(\beta;\tau,\overline{\tau},X\right) &= \prod_{m=1}^{N-\ell}\prod_{i=1}^{I} P_\ell^{\beta}\left(\beta;\tau_i,\overline{\tau}_i,X_i(m)\right)\\ \mathcal{L}_\ell^{\overline{\beta}}\left(\overline{\beta};\tau,\overline{\tau},X\right) &= \prod_{m=1}^{N-\ell}\prod_{i=1}^{I} P_\ell^{\overline{\beta}}\left(\overline{\beta};\tau_i,\overline{\tau}_i,X_i(m)\right).\end{aligned} \tag{13}$$

With the definitions given in (12) and (13), it can be seen that:

$$\mathcal{L}_\ell(\beta,\overline{\beta};\tau,\overline{\tau},X) = \mathcal{L}_\ell^{\beta}(\beta;\tau,\overline{\tau},X)\,\mathcal{L}_\ell^{\overline{\beta}}(\overline{\beta};\tau,\overline{\tau},X). \tag{14}$$

Thus, $\mathcal{L}_\ell^{\bar{\beta}}$ and $\mathcal{L}_\ell^{\beta}$ can be separately maximized to find their respective parameters. A further important separation is a separation by horizons. Notice that we can decompose $P_\ell^{\beta}$ and $P_\ell^{\bar{\beta}}$ as:

$$\begin{aligned}P_\ell^{\beta}\left(\beta;\tau_i,\overline{\tau}_i,X_i(m)\right) &= \prod_{\ell'=0}^{\ell-1} P^{\beta(\ell')}\left(\beta(\ell');\tau_i,\overline{\tau}_i,X_i(m)\right),\\ P_\ell^{\overline{\beta}}\left(\overline{\beta};\tau_i,\overline{\tau}_i,X_i(m)\right) &= \prod_{\ell'=0}^{\ell-1} P^{\overline{\beta}(\ell')}\left(\overline{\beta}(\ell');\tau_i,\overline{\tau}_i,X_i(m)\right),\end{aligned} \tag{15}$$

where

$$\begin{aligned}&P^{\beta(\ell')}(\beta(\ell');\tau_i,\overline{\tau}_i,X_i(m))\\ &= 1_{\{t_{0i}\le m,\min(\tau_i,\overline{\tau}_i)>m+\ell'+1\}}\exp\left[-\Delta t\, H(\beta(\ell'),X_i(m))\right]\\ &\quad + 1_{\{t_{0i}\le m,\tau_i\le\overline{\tau}_i,\tau_i=m+\ell'+1\}}\left\{1-\exp\left[-\Delta t\, H(\beta(\ell'),X_i(m))\right]\right\}\\ &\quad + 1_{\{t_{0i}\le m,\overline{\tau}_i<\tau_i,\overline{\tau}_i=m+\ell'+1\}}\exp\left[-\Delta t\, H(\beta(\ell'),X_i(m))\right]\\ &\quad + 1_{\{t_{0i}>m\}} + 1_{\{\min(\tau_i,\overline{\tau}_i)<m+\ell'+1\}},\\ &P^{\overline{\beta}(\ell')}(\overline{\beta}(\ell');\tau_i,\overline{\tau}_i,X_i(m))\\ &= 1_{\{t_{0i}\le m,\min(\tau_i,\overline{\tau}_i)>m+\ell'+1\}}\exp\left[-\Delta t\, H(\overline{\beta}(\ell'),X_i(m))\right]\\ &\quad + 1_{\{t_{0i}\le m,\tau_i\le\overline{\tau}_i,\tau_i=m+\ell'+1\}}\\ &\quad + 1_{\{t_{0i}\le m,\overline{\tau}_i\le\tau_i,\overline{\tau}_i=m+\ell'+1\}}\{1-\exp[-\Delta t\, H(\overline{\beta}(\ell'),X_i(m))]\}\\ &\quad + 1_{\{t_{0i}>m\}} + 1_{\{\min(\tau_i,\overline{\tau}_i)<m+\ell'+1\}}.\end{aligned} \tag{16}$$

Thus, the $\beta$ and $\bar{\beta}$ specific pseudo-likelihood functions can be decomposed as:

$$\begin{aligned}\mathcal{L}_\ell^{\beta}\left(\beta;\tau,\overline{\tau},X\right) &= \prod_{\ell'=0}^{\ell-1} L^{\beta(\ell')}\left(\beta(\ell');\tau,\overline{\tau},X\right)\\ \mathcal{L}_\ell^{\overline{\beta}}\left(\overline{\beta};\tau,\overline{\tau},X\right) &= \prod_{\ell'=0}^{\ell-1} L^{\overline{\beta}(\ell')}\left(\overline{\beta}(\ell');\tau,\overline{\tau},X\right).\end{aligned} \tag{17}$$

Where

$$\begin{aligned}\mathcal{L}^{\beta(\ell')}\left(\beta(\ell');\tau,\overline{\tau},X\right) &= \prod_{m=1}^{N-\ell}\prod_{i=1}^{\ell} P^{\beta(\ell')}\left(\beta(\ell');\tau_i,\overline{\tau}_i,X_i(m)\right)\\ \mathcal{L}^{\overline{\beta}(\ell')}\left(\overline{\beta}(\ell');\tau,\overline{\tau},X\right) &= \prod_{m=1}^{N-\ell}\prod_{i=1}^{\ell} P^{\overline{\beta}(\ell')}\left(\overline{\beta}(\ell');\tau_i,\overline{\tau}_i,X_i(m)\right).\end{aligned} \tag{18}$$

Thus, for every horizon $\ell'$, $\mathcal{L}^{\bar{\beta}(\ell')}(\beta(\ell');\ \tau,\ \bar{\tau},\ X)$ and $\mathcal{L}^{\bar{\beta}(\ell')}(\bar{\beta}(\ell');\ \tau,\ \bar{\tau},\ X)$ can be separately maximized. In summary, for the current CRI implementation where the horizons are from one month to 24 months, and where there are 13 variables, a $2 \times 24 \times 13$ dimensional maximization is turned into a 13 dimensional maximization done $2 \times 24$ times. This makes the calibration problem tractable. Additional implementation details on the calibration are given in Section 3.

## II. INPUT VARIABLES AND DATA

Subsection 2.1 describes the input variables used in the quantitative model. Currently, the same set of input variables is common to all of the economies under the CRI's coverage. Future enhancements to the CRI system will allow different input variables for different economies. The effect of each of the variables on the PD output is discussed in the empirical analysis of Section 4.

Subsection 2.2 gives the data sources and relevant details of the data sources. There are two categories of data sources: current and historical. Data sources used for current data need to be updated in a timely manner so that daily updates of PD forecasts are meaningful. They also need to be comprehensive in their current coverage of firms. Data sources that are comprehensive for current data may not necessarily have comprehensive historical coverage for different economies. Other data sources are thus merged in order to obtain comprehensive coverage for historical and current data.

Subsection 2.3 indicates the fields from the data sources that are used to construct the input variables. For some of the fields, proxies need to be used for a firm if the preferred field is not available for that firm.

Subsection 2.4 discusses the definition and sources of defaults and of other exits used in the CRI.

## 2.1. Input Variables

Following the notation that was introduced in Section 1, firm $i$'s input variables at time $t = n\Delta t$ are represented by the vector $X_i(n) = (W(n), U_i(n))$ consisting of a vector $W(n)$ that is common to all firms in the same economy, and a firm-specific vector $U_i(n)$ which is observable from the date the firm's first financial statement is released, until the month end before the month in which the firm exits, if it does exit.

In Duan *et al*. (2011), different variables that are commonly used in the literature were tested as candidates for the elements of $W(n)$ and $U_i(n)$. Two common variables and ten firm-specific variables, as described below, were selected as having the greatest predictive power for corporate defaults in the United States. In the current stage of development, this same set of twelve input variables is used for all economies. Future development will include variable selection for firms in different economies.

- Common variables

  The vector $W(n)$ contains two elements, consisting of:

  1. Stock index return: the trailing one-year simple return on a major stock index of the economy.
  2. Interest rate: a representative three-month short-term interest rate with the historical mean subtracted to obtain a de-meaned time series.

- Firm-specific variables

  The ten firm-specific input variables are transformations of measures of six different firm characteristics. The six firm characteristics are: (i) volatility-adjusted leverage; (ii) liquidity; (iii) profitability; (iv) relative size; (v) market misvaluation/future growth opportunities; and (vi) idiosyncratic volatility.

Volatility-adjusted leverage is measured as the distance-to-default (DTD) in a Merton-type model. The calculation of DTD used by the CRI allows a meaningful DTD for financial firms, a critical group that must be excluded from most DTD computations. This calculation is detailed in Section 3.

Liquidity is measured as a ratio of cash and short-term investments to total assets, profitability is measured as a ratio of net income to total assets, and relative size is measured as the logarithm of the ratio of market capitalization to the economy's median market capitalization.

Duan *et al*. (2012) transformed these first four characteristics into level and trend versions of the measures. For each of these, the level is 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. The level and trend of a measure has seldom been used in the academic or industry literature for default prediction, and Duan *et al*. (2012) found that using the level and trend significantly improves the predictive power of the model for short-term horizons.

To understand the intuition behind using level and trend of a measure as opposed to using just the current value, consider the case of two firms with the same current value for all measures. If the level and trend transformations were not performed, then only the current values would be used and the two firms would have identical PD. Suppose that for the first firm the DTD had reached its current level from a high level, and for the second firm the DTD had reached its current level from a lower level (see Figure 2). The first firm's leverage is increasing (worsening) and the second firm's leverage is decreasing (improving). If there is a momentum effect in DTD, then firm 1 should have a higher PD than firm 2.

Duan *et al*. (2012) found evidence of the momentum effect in DTD, liquidity, profitability and size. For the other two firm characteristics, applying the level and trend transformation did not improve the predictive power of the model.

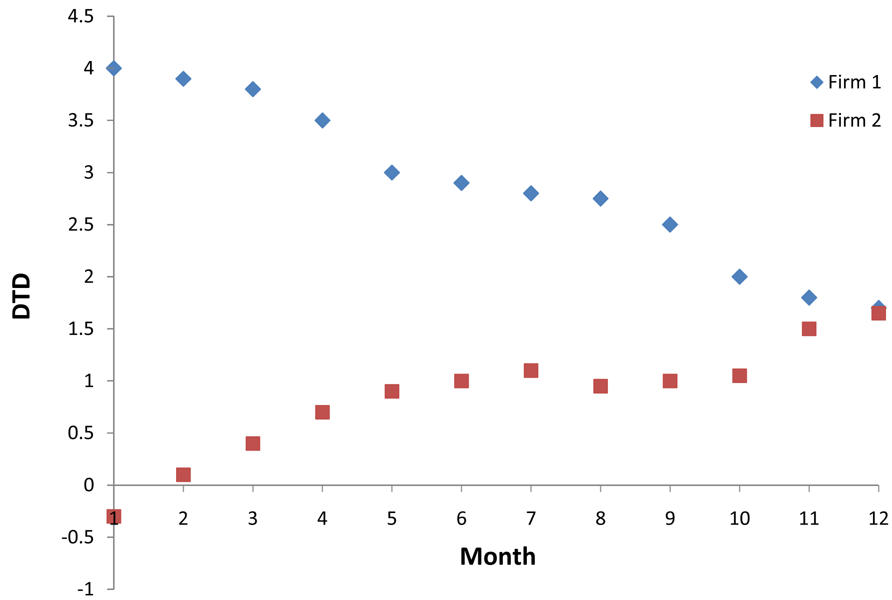


**Figure 2.** Two firms with all current values equal to each other, but DTD trending in the opposite direction.

One of the remaining two firm characteristics is the market mis-valuation/future growth opportunities characteristic, which is taken as the market-to-book asset ratio and measured as a ratio of market capitalization and total liabilities to total assets. One can see whether the market mis-valuation effect or the future growth opportunities effect dominates this measure by looking at whether the parameter for this variable is positive or negative. This is further discussed in the empirical analysis of Section 4.

The final firm characteristic is the idiosyncratic volatility which is taken as sigma, following Shumway (2001). Sigma is computed by regressing the monthly returns of the firm's market capitalization on the monthly returns of the economy's stock index, for the previous 12 months. Sigma is defined to be the standard deviation of the residuals of this regression. Shumway (2001) reasons that sigma should be logically related to bankruptcy since firms with more variable cash flows and therefore more variable stock returns relative to a market index are likely to have a higher probability of bankruptcy.

Finally, the vector $U_i(n)$ contains ten elements, consisting of:

1. Level of DTD.
2. Trend of DTD.
3. Level of (Cash + Short-term investments)/Total assets, abbreviated as CASH/TA.
4. Trend of CASH/TA.
5. Level of Net income / Total Assets, abbreviated as NI/TA.
6. Trend of NI/TA.
7. Level of log (Firm market capitalization/ Economy's median market capitalization), abbreviated as SIZE.
8. Trend of SIZE.
9. Current value of (Market capitalization + total liabilities)/Total asset, abbreviated as M/B.
10. Current value of SIGMA.

The data fields that are needed to compute DTD and short-term investments are described in Subsection 2.3. The remaining data fields required are straightforward and standard. The computation for DTD is explained in Section 3.

## 2.2. Data Sources

There are two data sources that are used for the daily PD forecast updates: Thomson Reuters Datastream and the Bloomberg Data License Back Office Product. Many of the common factors such as stock index prices and short-term interest rates are retrieved from Datastream.

Firm-specific data comes from Bloomberg's Back Office Product which delivers daily update files by region via FTP after respective market closes. All relevant data is extracted from the FTP files and uploaded into the CRI database for storage. From this, the necessary fields are extracted and joined with previous months of data.

The Back Office Product includes daily market capitalization data based on closing share prices and also includes new financial statements as companies release them. Firms will often have multiple versions of financial statements within the same period, with different accounting standards, filing statuses (most recent, preliminary, original, reclassified or restated), currencies or consolidated/unconsolidated indicators. A major challenge lies in prioritizing these financial statements to decide which data should be used. The priority rules are described in Section 3.

The firm coverage of the Back Office Product is of sufficient quality that over 35,000 firms can be updated on a daily basis in the 37 economies under the CRI's coverage. While the current coverage is quite comprehensive, historical data from the Back Office Product can be sparse for certain economies. For this reason, various other databases are merged in order to fill out the historical data. The other databases used for historical data are: a database from the Taiwan Economics Journal (TEJ) for Taiwanese firms; a database provided by Korea University for South Korean firms; and data from Prowess for Indian firms.

With all of the databases merged together and for the 37 economies under CRI's coverage, over 62,000 exchange listed firms are in the CRI database. This includes over 20,000 delisted firms. The historical coverage of the firm data goes back to the early 1990's.

## 2.3. Constructing Input Variables

The chosen stock indices and short-term interest rates for the 37 economies under the CRI's current coverage are listed in Tables A.2 and A.3, respectively. All economies are listed by their three letter ISO code given in Table A.1.

Most of the firm-specific variables can be readily constructed from standard fields within firms' financial statements in addition to daily market capitalization values. The only two exceptions are the DTD and the liquidity measure.

The calculation for DTD is explained in Section 3. In the calculation, several variables are required. One variable is a proxy for a one-year risk-free interest rate, and the choices for each of the 37 economies are listed in Table A.4. Total assets, long-term borrowing and total liabilities are also required, but are standard financial statement fields and present no difficulties.

Total current liabilities are also required, and due to the relatively large numbers of firms that are missing this value, proxies had to be found. The preferred Bloomberg field for this is BS_CUR_LIAB. If this is missing, then the sum of BS_ST_BORROW, BS_OTHER_ST_LIAB and BS_CUST_ACCPT_LIAB_CUSTDY_SEC (customers' acceptance and liabilities/custody securities) is used. If one or two of these are missing, zero is inserted for those fields, but at least one field is required.

The liquidity measure requires different fields between financial and non-financial firms. For non-financial firms, the numerator of the ratio (Cash + Short-term investments) is taken as the sum of BS_CASH_NEAR_CASH_ITEM and BS_MKT_SEC_OTHER_ST_INVEST (marketable securities and other short-term investments). If BS_MKT_SEC_OTHER_ST_INVEST is missing, we substitute with zero but the field BS_CASH_NEAR_CASH_ITEM is required.

It was found that this sum frequently overstated the liquidity for financial firms. In place of BS_MKT_SEC_OTHER_ST_INVEST, financial firms use the sum of ARD_SEC_PURC_UNDER_AGR_TO_RESELL (securities purchased under agreement to re-sell), ARD_ST_INVEST and BS_INTERBANK_ASSET. If one or two of these are missing, zero is inserted for those fields, but at least one field is required. The "ARD" prefix indicates that these are "as reported" numbers directly from the financial statements. As such, for some firms these fields may need to be adjusted to the same units before adding them to other fields.

Summary statistics of the firm-specific variables: DTD, CASH/TA, NI/TA, SIZE, M/B, and Sigma, with the summary statistics provided for firms grouped by economy are available in the section on the technical report at the CRI web portal.

## 2.4. Data for Defaults

The Credit Research Initiative database contains credit events of over 4,000 firms from 1990 to the present. The default events come from numerous sources, including Bloomberg, Compustat, CRSP, Moody's reports, TEJ, exchange web sites and news sources.

The default events that are recognized by the CRI can be classified under one of the following events:

1. Bankruptcy filing, receivership, administration, liquidation or any other legal impasse to the timely settlement of interest and/or principal payments;
2. A missed or delayed payment of interest and/or principal, excluding delayed payments made within a grace period;
3. Debt restructuring/distressed exchange, in which debt holders are offered a new security or package of securities that result in a diminished financial obligation (e.g., a conversion of debt to equity, debt with lower coupon or par amount, debt with lower seniority, debt with longer maturity).

The more precise sub-categories of default corporate actions are listed in Table A.5.

Delisting due to other reasons such as failure to meet listing requirements, inactive stock prices or M&A are counted as "other exits" and are not considered as default. However, firms that are delisted from an exchange and which experience a default event within 365 calendar days of the delisting will have an exit event reclassified as credit default. Technical defaults such as covenant violations are not included in our definition of default. The exit events that are not considered as defaults in the CRI system are listed in Table A.6.

In addition to the aforementioned events, there are still cases that require special attention and will be assessed on a case-by-case basis, e.g., subsidiary default. As a general rule, the CRI does not consider related party-default (e.g., subsidiary bankruptcy) as a default event. However, when a non-operating holding parent company relies heavily on its subsidiary, bankruptcy by the subsidiary will cause a considerable economic impact on the parent company. Such cases are reviewed and final classifications made.

The total number of firms, number of defaults and number of other exits in each of the 37 economies each year from 1992 to 2012 are listed in the section on the technical report at the CRI web portal. Note that the total number of firms here includes all firms where the primary listing of the shares are on an exchange in that economy and may include firms where there are too many missing data values for a PD estimate to be made. However, the number of firms listed on the CRI web portal under the tab Aggregate forecast includes firms that are domiciled in that economy and excludes firms where a PD cannot be produced due to missing data.

# III. IMPLEMENTATION DETAILS

Section 1 describes the modeling framework underlying the current implementation of the CRI system. It focuses on theory rather than the details encountered in an operational implementation. The present section describes how the CRI system handles these more specific issues.

Subsection 3.1 describes implementation details related to data, mainly dealing with data cleaning and missing data. Subsection 3.2 describes the specific computation of distance-to-default (DTD) used by the CRI system that leads to meaningful DTD for financial firms. Subsection 3.3 explains how the calibration previously described in Subsection 2.2 can be implemented. Subsection 3.4 gives the implementation details relevant to the daily output. This includes an explanation of the various modifications needed to compute daily PD so that the daily PD is consistent with the usual month end PD, and a description of the computation of the aggregate PDs provided by the CRI.

## 3.1. Data Treatment

**Fitting data to monthly frequency:** Historical end of month data for every firm in an economy is required to calibrate the model. For daily data such as market capitalization, interest rates and stock index values, the last day of the month for which there is valid data is used.

For financial statement variables, data is used starting from the period end of the statement lagged by three months. This is to ensure (insofar as is possible) that predictions are made based on information that was available at the time the prediction was made. Of course, for more recent data where the CRI database contains the financial statement but the period end lagged by three months is after the current day, the financial statement is used in making PD forecasts. The CRI considers financial statement variables to be valid for one year without restriction after they are first used.

Currency conversions are required if the market capitalization or any of the financial statement variables are reported in a currency different than the currency of the economy. If a currency conversion is required, the foreign exchange rate used is that reported at the relevant market close. For firms traded in Asia and Asia-Pacific, the Tokyo closing rate is used; for firms traded in Western Europe, the London closing rate is used; and for firms traded in North America and Latin America, the New York closing rate is used. For market capitalizations, the FX rate used is for the date that the market capitalization is reported. For financial statement variables, the FX rate used is for the date of the period end of the statement.

**Priority of financial statements:** As described in Subsection 2.2, data provided in Bloomberg's Back Office Product can include numerous versions of financial statements within the same period. If there are multiple financial statements with the same period end, priority rules must be followed in order to determine which to use. The formulation and implementation of these rules is a major challenge and an area of continuing development.

The first rule prioritizes by consolidated/unconsolidated status. This status is relevant only to firms in India, Japan, South Korea and Taiwan, so this rule is only relevant in those economies. Most firms in these economies issue unconsolidated financial statements more frequently than consolidated ones, so these are given higher priority. This simple prioritization can, however, lead to cases where the financial statements used switch from consolidated statements to unconsolidated statements and back again. A more complex prioritization rule is currently under development, with the intention of avoiding this situation.

If, after the first prioritization rule has been applied, there are still multiple financial statements, the second rule is applied. This is prioritization by fiscal period. In most economies, annual statements are required to be audited, whereas other fiscal periods are not necessarily audited. The order of priority from highest to lowest is, therefore: annual, semi-annual, quarterly, cumulative, and finally other fiscal periods.

The third prioritization rule is based on filing status. The "Most Recent" statement is used before the "Original" statement, which is used before the "Preliminary" statement.

The final prioritization rule is based on the accounting standard. Here, financial statements that are reported using Generally Accepted Accounting Principles (GAAP) are given higher priority than financial statements that are reported using International Financial Reporting Standards (IFRS). If an accounting standard is not indicated at all, the financial statement is not used.

Financial statement entries with all other descriptors being the same but with different filing statuses will be grouped together. For each variable separately, the variable value is taken from the highest priority financial statement within the group where the value is non-null.

For example, suppose two financial statement entries have the same period end, are both annual statements, are both consolidated statements, and both use the same accounting standard, but the first entry is classified as the "Most Recent" and the second is classified as the "Original" entry. Suppose the total assets and total liabilities are reported in the "Original" entry, and in the "Most Recent" entry only the total liabilities have been updated with a null value for the "Original" entry. Then, the total liabilities will be taken from the "Most Recent" entry while the total assets will be taken from the "Original" entry.

This allows for the grouping of, for example, "Most Recent" and "Original" entries together because Bloomberg occasionally only updates values that change without updating other values. If the entries are not grouped, then most of the variables would have null values.

One variable that needs special attention is net income. Net income is a flow variable and needs to be adjusted based on the period of the financial statement. More specifically we transform the net income into a monthly net income by dividing the net income by the number of months that the financial statement covers. Due to the different coverage periods, several sources for the net income may be available. For example, the monthly net income can be computed from the annual net income divided by 12, the semi-annual net income divided by six and the quarterly net income divided by three. When the monthly net income can be obtained from different sources simultaneously, the quarterly net income will have higher priority than any other because it covers a more recent period.

**Treatment of stale market capitalization prices:** The market capitalization of a firm is required in a few input variables: DTD, SIZE, M/B and SIGMA. For most firms, the market capitalization is available from Bloomberg on a daily basis.

A check on the trading volume of shares is used to remove stale prices. Specifically, if there are more than two consecutive days of identical market capitalization prices, subsequent identical prices are removed only if the trading volume is equal to zero. This is to avoid, for example, cases where the shares of a company are under a trading suspension but the market capitalization data is incorrectly carried forward.

An exception is for Indian companies, where it is common for some companies to have market capitalizations reported only once a month with several consecutive months having identical prices and positive trading volume. These prices are very likely not to be accurate reflections of the firms' value. So, the trading volume is not checked for Indian firms and market capitalizations are excluded after more than two repeated prices.

For some firms, there are gaps in the market capitalization data provided by Bloomberg. Previously, the first recourse was to use the share price multiplied by the shares outstanding listed in the balance sheet and multiplied by an adjustment factor that Bloomberg provides to account for splits, dividends, etc. However, this data is frequently in error and using the shares outstanding as the previous available market capitalization divided by the price on that day was found to be more reliable.

If the gap in market capitalization data is more than a year, then the previous computation using the shares outstanding from the balance sheet is again used. If there are still remaining gaps in the data, then shares outstanding from Compustat data is used.

**Provisions for missing values and outliers:** Missing values and outliers are dealt with by a three step procedure. In the first step, the ten firm-specific input variables are computed for all firms and all months. In the second step, outliers are eliminated by winsorization. In the final step, missing values are replaced under certain conditions.

The first step is to compute the input variables and determine which are missing. As mentioned previously, financial statement variables are carried forward for one year after the date that they are first used. This is generally three months after the period end of the statement. If no financial statement is available for the company within this year, then the financial statement variable will be missing. For market capitalization, if there is no valid market capitalization value within the calendar month, then the value is set to missing.

For illiquid stocks, if there has been no valid market capitalization value for a firm within the last 90 calendar days, then the market capitalization is deemed to not properly reflect the value of the firm. The firm is considered to have exited with a non-default event. Once the firm starts trading again and a new financial statement is released, the firm can enter back into the calibration. With regard to historical PD, the PD can be reported again once there are enough valid variables.

With regard to the level variables, the current month and the last eleven months are averaged to compute the level. There is no lower limit on the number of valid observations. Only if all of the values are missing is the level variable considered to be missing.

For the trend variable, the level is subtracted from the current month. If the current month is missing, then the trend variable is set to missing.

The value of M/B is set to missing if any of the following values are missing: market capitalization, total liabilities or total assets of the firm. For the computation of SIGMA, seven valid returns over the last twelve months of possible returns are required for the regression. If there are less than seven valid returns, SIGMA is set to missing.

In this way, the eight trend and level variables plus M/B and SIGMA are computed and evaluated as missing or present. Winsorization can then be performed as a second step to eliminate outliers. The volume of outliers is too large to be able to determine whether each one is valid or not, so winsorization applies a floor and a cap on each of the variables. The historical 0.1 percentile and 99.9 percentile for all firms in the economy are recorded for each of the ten variables. Any values that exceed these levels are set to equal these boundary values.

With a winsorization level and 0.1 percentile and 99.9 percentile, the boundary values still may not be reasonable. For example, NI/TA levels of nearly −25 have been observed at this stage. In these cases, a more aggressive winsorization level is applied, until the boundary values are reasonable. Thus, the winsorization level is economy and variable specific, and will depend on the data quality for that economy and variable. The applied winsorization levels different from the default of 0.1 percentile and 99.9 percentile are indicated in the tables on our web portal.

A third and final step can be taken to deal with missing values. If, during a particular month, no variables for a firm are missing, then the PD can be computed. If six or more of these ten variables are missing, there is deemed to be too many missing observations and no replacements are made.

If between one and five variables are missing out of the ten, the first step is to trace back for at most twelve months to use previous values of these variables instead. If this does not succeed in replacing all of the variables, a replacement by sector medians is done. The median is for the financial or non-financial firms (as indicated by their Bloomberg industry sector code) within the economy during that month. Replacement by the sector median should have a neutral effect on the PD of the firm; the firm is assessed by the other variables that it does have values for. This sector median is always performed in calibration. However, when reporting historical PD, the sector replacement is not done if it results in a relative change in PD of 10% or more where the initial PD was at or above 100bps, or an absolute change in PD of 10bps or more where the initial PD was below 100bps.

**Inclusion/exclusion of companies for calibration:** Firms are included within an economy for calibration when the primary listing of the firm is on an exchange in the economy. This ensures that all firms within the economy are subject to the same disclosure and accounting rules.

There are a relatively small number of firms that are dual listed, in which two corporations listed in different exchanges operate as a single entity but retain separate legal status. In the CRI system, a combined company will be assigned to the single economy it is most associated with. An example is the Rio Tinto Group. This consists of Rio Tinto plc, listed in the UK; and Rio Tinto Limited, listed in Australia. Most of Rio Tinto's operations are in Australia rather than the UK, so Rio Tinto is assigned to Australia.

In the US, firms traded on the OTC markets or the Pink Sheets are not considered as exchange listed so are not included in calibration or in the reporting of PD forecasts. Many of these firms are small or start-up firms. Including this large group of companies would skew the calibration and the aggregate results. The TSX Venture Exchange in Canada also contains only small and start-up firms, so firms listed on that exchange are also excluded.

Other examples include Taiwan's GreTai Securities Market and Singapore's Catalist. The challenge for markets outside of the US or Canada is that the data on whether firms are listed on the smaller markets rather than the main board is diffcult to obtain. For all economies besides the US and Canada, there is continuing work being done in the CRI system to exclude firms that are not listed on major exchanges within a country.

Firms that record an exit (other than due to no trading for 90 calendar days) are not entered back into the calibration even if the firm continues to trade and issue financial statements, as can happen after firms declare bankruptcy. There are two exceptions to this exclusion. The first, determined on a case by case basis, is if the firm should be deemed to have re-emerged from bankruptcy. The second exception is for all firms in China, where two situations are prevalent. The first situation is that the firm experiences few repercussions from the default and continues operating normally. The other situation is for one firm to take over a defaulted firm's listing. This happens due to the limited supply of exchange listings. Both of these situations can be considered as emerging from default, so the CRI system enters all of these companies back into the calibration as new companies.

## 3.2. Distance-to-Default Computation

The distance-to-default (DTD) computation used in the CRI system is not a standard one. Standard computations exclude financial firms, but excluding the financial sector means neglecting a critical part of any economy. So the standard DTD computation must be extended to give meaningful estimates for financial firms as well. Duan and Wang (2012) provide a review of different DTD calculations with several examples for financial and non-financial firms.

The description of the specialized DTD computation starts with a brief description of the Merton (1974) model. Merton's model makes the simplifying assumption that firms are financed by equity and a single zero-coupon bond with maturity date $T$ and principal $L$. The asset value of the firm $V_t$ follows a geometric Brownian motion:

$$dV_t = \mu V_t dt + \sigma V_t dB_t. \tag{19}$$

Here, $B_t$ is standard Brownian motion, $\mu$ is the drift of the asset value in the physical measure and $\sigma$ is the volatility of the asset value. Equity holders receive the excess value of the firm above the principal of the zero-coupon bond and have limited liability, so the equity value at maturity is: $E_T = \max(V_T - L, 0)$. This is just a call option payoff on the asset value with a strike value of $L$. Thus, the Black–Scholes option pricing formula can be used for the equity value at times $t$ before $T$:

$$E_t = V_t N(d_1) - e^{-r(T-t)} L N(d_2)\cdot \tag{20}$$

where $r$ is the risk-free rate, $N(\cdot)$ is the standard normal cumulative distribution function, and:

$$d_{1,2} = \frac{\log\left(\frac{V_t}{L}\right) + \left(r \pm \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}} \tag{21}$$

In Merton's model, DTD is defined as volatility scaled distance of the expected asset value under the physical measure at maturity $T$ from the default point $L$:

$$DTD_t = \frac{\log\left(\frac{V_t}{L}\right) + \left(\mu - \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}}. \tag{22}$$

The standard KMV assumptions given in Crosbie and Bohn (2003) are to set the time to maturity $T - t$ at a value of one year and the principal of the zero-coupon bond $L$ to a value equal to the firms current liabilities plus one half of its long-term debt. Here, the current liabilities and long-term debt are taken from the firm's financial statements. If the firm is missing the current liabilities field, then various substitutes for this field can be used, as described in Subsection 2.3.

This is a poor assumption of the debt level for financial firms, since they typically have large liabilities, such as deposit accounts, that are neither classified as current liabilities nor long-term debt. Thus, using these standard assumptions means ignoring a large part of the debt of financial firms.

To properly account for the debt of financial firms, Duan (2010) includes a fraction $\delta$ of a firm's other liabilities. The other liabilities are defined as the firm's total liabilities minus both the short and long-term debt. The debt level $L$ then becomes the current liabilities plus half of the long-term debt plus the fraction $\delta$ multiplied by the other liabilities, so that the debt level is a function of $\delta$. The standard KMV assumptions are then a special case where $\delta = 0$.

The fraction $\delta$ can be optimized along with and in the maximum likelihood estimation method developed in Duan (1994, 2000). Following Duan *et al*. (2012), the firm's market value of assets is standardized by its book value $A_t$ so that the scaling effect from a major investment or financing by the firm will not distort the time series from which the parameter values are estimated. Thus, the log-likelihood function is:

$$\begin{aligned}
&\mathcal{L}(\mu,\sigma,\delta) \\
&= -\frac{n-1}{2}\log(2\pi) - \frac{1}{2}\sum_{t=2}^{n}\log(\sigma^2 h_t) - \sum_{t=2}^{n}\log\left(\frac{\hat{V}_t(\sigma,\delta)}{A_t}\right) \\
&\quad - \sum_{t=2}^{n}\log\left[N\left(\hat{d}_1\left(\hat{V}_t(\sigma,\delta),\sigma,\delta\right)\right)\right] \\
&\quad - \sum_{t=2}^{n}\frac{1}{2\sigma^2 h_t}\left[\log\left(\frac{\hat{V}_t(\sigma,\delta)}{A_t}\times\frac{A_{t-1}}{\hat{V}_{t-1}(\sigma,\delta)}\right) - \left(\mu - \frac{\sigma^2}{2}\right)h_t\right]^2.
\end{aligned} \tag{23}$$

Here, $n$ is the number of days with observations of the equity value in the sample, $\hat{V}_t$ is the implied asset value found by solving equation (20), $\hat{d}_1$, is computed with equation (21) using the implied asset value, and $h_t$ is the number of trading days as a fraction of the year between observations $t - 1$ and $t$. Notice that the implied asset value and $\hat{d}_1$ are dependent on $\delta$ by virtue of the dependence of $L$ on $\delta$.

**Implementation of DTD computation:** The DTD at the end of each month is needed for every firm in order to calibrate the forward intensity model. A moving window, consisting of the last one year of data before each month end is used to compute the month end DTD. Daily market capitalization data based on closing prices is used for the equity value in the implied asset value computation of equation (20). If there are fewer than 50 days of valid observations for the market capitalization, then the DTD value is set to missing. An observation is valid if there is positive trading volume that day. If the trading volume is not available, the observation is assumed to be valid if the value for the market capitalization changes often enough. The precise criterion is as follows: if the market capitalization does not change for three days or more in a row, the first day is taken as a valid observation and the remaining days with the same value are set to missing.

The log-likelihood function given in (23) can be maximized as a three dimensional maximization problem over $\mu$, $\sigma$ and $\delta$. After estimates for these three variables are made, the DTD can be computed from equation (22).

However, with quarterly financial statements there will never be more than three changes in the corporate structure (defined in this model by $L$ and $A_t$) throughout the year, leading to possibly unstable estimates of $\delta$. This problem is mitigated by performing a two stage optimization for $\mu$, $\sigma$ and $\delta$.

In the first stage, the optimization for each firm is performed over all three variables. For each firm, in the first month in which DTD can be computed the optimization is unconstrained in $\mu$ and $\sigma$, while $\delta$ is constrained to being in the unit interval $[0, 1]$. Thereafter, at month $n$, the optimization is still unconstrained in $\mu$ and $\sigma$ while $\delta$ is constrained to the interval $[\max(0,\hat{\delta}_{n-1} - 0.05),\min(1,\hat{\delta}_{n-1} + 0.05)]$, where $\hat{\delta}_{(n-1)}$ is the estimate of $\delta$ made in the previous month. In other words, a 10% band around the previous estimate of $\delta$ (where that band is floored with 0 and capped with 1) is applied so that the estimates do not fluctuate too much from month to month.

It was found that this was not enough to obtain stable estimates of $\delta$. For many firms, the estimate of $\delta$ would frequently lie on the boundary of the constraining interval. To impose greater stability, a second stage is added. At each month end, the average estimate for $\delta$ in all financial sector firms in the economy is used for every financial sector firm in the economy, meaning the optimization is only over $\mu$ and $\sigma$. The same is done for non-financial firms. In fact, the optimization can be reduced to be only over $\sigma$ by using the sample mean of the log returns of the implied asset values in place of $\mu$.

Since the first stage is done to obtain a stable sector average estimate of $\delta$, the criteria used to include a firm-month is more strict. In the first stage, a two-year window is used instead of one year, and a minimum of 250 days of valid observations of the market capitalization are required instead of 50. If a firm has less than 250 days of valid observations within the last two years of a particular month end, $\delta$ will not be estimated for that firm and that month end.

It was found that the estimate of $\mu$ was frequently unstable and could lower the explanatory power of DTD. For example, suppose a firm has a large drop in its implied asset value in January 2011, so that the estimated $\mu$ is negative for the DTD calculation at the end of December 2011. If there is little change in the company in January 2012, then the drop in implied asset value in January 2011 is no longer within the observation window for the DTD calculation at the end of January 2012. There will be a large increase in the estimated $\mu$, resulting in a substantial improvement of the DTD just because of the moving observation window.

To avoid this problem, we now set $\mu$ to be equal to $\sigma^2/2$. So in calculating DTD, the second term in the numerator of Equation (22) is eliminated.

In summary, the DTD for each firm is computed using the economy and sector (financial or non-financial) average for $\delta$ in that month, and the estimate of $\sigma$ is based on the last year of data for the firm.

Carrying out this two-stage procedure would take several months of computation time on a single PC, given the millions of firm months that are required. However, each of the stages is parallelizable. In the first stage the DTD can be computed independently between firms. In the second stage, once the sector averages of the $\delta$ have been computed for each month, the DTD can again be computed independently between firms. In the CRI system, a grid of several hundred computers administered by the NUS Computer Center is used. With this, the DTD computation can be performed for all firms over the full history of twenty years in less than one day.

### 3.3. Calibration

**Implementation:** As shown in Section 1, the calibration of the forward intensity model involves multiple maximum pseudo-likelihood estimations, where the pseudo-likelihood functions are given in equation (18). The maximizations are of the logarithm of these expressions, and they are performed independently between the default parameters and the exit parameters, and between parameters for different horizons. In the notation of Section 1, the vectors of parameters $\beta(0), \ldots, \beta$ (23) and $\bar{\beta}(0), \ldots, \bar{\beta}(23)$ are independently estimated.

A few input variables have an unambiguous effect on a firm's probability of default. Increasing values of both the level and trend of DTD, CASH/TA, and NI/TA all indicate that a firm is becoming more credit worthy and should lead to a decreased PD. For large and relatively clean datasets such as the US, an unconstrained optimization leads to parameter values which largely have the expected sign. For each of DTD level and trend, CASH/TA level and trend, and NI/TA level, the default parameters at all horizons are negative. A negative default parameter at a horizon means that if the variable increases, the forward intensity will decrease (by equation (6)), so that the conditional default probability at that horizon will decrease. The one exception is the NI/TA trend variable.

For some of the smaller economies and economies with lower quality datasets, an unconstrained optimization leads to the default parameters for some of these variables to be positive at several horizons. This leads to counter-intuitive results. For example, if the default parameters for CASH/TA are positive, a firm that increases its cash reserves, all other factors being equal, will have a PD that increases. To prevent such situations, the CRI system performs a constrained optimization with only non-positive values allowed for the default parameters associated with the level and trend of DTD, CASH/TA, and NI/TA.

For this, the Matlab function "fmincon" from the Optimization Toolbox is used. The analytic gradient and Hessian are supplied and the algorithm used by "fmincon" is the trust-region-reflective optimization. If "fmincon" fails to converge, "fminsearch" is used. This uses a simplex search method which takes more time but is generally more likely to converge.

Each evaluation of the pseudo-log-likelihood function can be done in a fraction of second on a standard CPU, even for the largest economies. But since the optimization is over 13 dimensions, thousands of evaluations are required. It is therefore important to make each function evaluation as fast as possible.

Notice that at each time point and at any horizon, there are in orders of magnitude more surviving firms than exiting firms. Thus, from equations (16) and (18), it can be seen that the most time-consuming part of evaluating the pseudo-log-likelihood function is the term for the surviving firms. Evaluating the forward intensity function of equation (7) can be formulated as a matrix-vector multiplication, where the rows of the matrix are the different surviving firms variables, and the vector is the vector of parameters. The matrix will typically have several hundreds of thousands of rows and does not change during the optimization (though it will change for different optimizations at different horizons). This type of problem is well-suited for a programmable graphics processing unit (GPU). The CRI system runs the calibrations on an NVIDIA Tesla C2050 card. For each economy, the calibrations for the default and other exit parameters for horizons up to 24 months typically require five minutes or less.

**Grouping for economies:** There are not enough defaults in some small economies and calibrations of these individual economies are not statistically meaningful. In order to ensure that there are enough defaults for calibration, the 37 economies are categorized into groups according to similarities in their stage of development and their geographic locations. Within these groups the economies are combined and calibrated together.

Starting from the May 2011 calibration, all sixteen of the European countries covered by the CRI are in a single calibration group, Canada and the US remain in the same calibration group, and the developed economies of Asia-Pacific (Australia, Hong Kong, Japan, Singapore, South Korea and Taiwan) form another calibration group. China and India, the two major emerging economies of Asia Pacific are each calibrated as an individual group. Starting from June 2012 the other emerging economies of Asia Pacific (Indonesia, Malaysia, Philippines and Thailand) are grouped together with the 7 Latin American countries (Argentina, Brazil, Colombia, Chile, Mexico, Peru and Venezuela) to form the calibration group "emerging markets".

All economies in these new calibrations groups share the same coefficients for all variables except for the benchmark risk-free interest rate variable. The benchmark interest rate's coefficients will be allowed to vary, because different economies based in different currencies naturally have different dependencies on their interest rates, and the interest rate levels can differ significantly across economies. After adopting the euro, all eurozone countries use Germany's three-month Bubill rate as this is more reflective of monetary rather than sovereign credit conditions in each economy, which is the intent of this variable. For the period before joining the eurozone, their own interest rates are used.

In addition, the benchmark interest rate is entered as the current value minus the historical month-end mean. This allows the variable to reflect its value relative to the historical average. When an economy does not have enough default events to identify a separate interest rate coefficient, the interest rate variable will be disabled for that economy by inputting a zero value for the whole time series. In fact, that is also why we de-mean all interest rate series so that setting the interest rate series of a particular economy to zero, when necessary, does not induce a bias by the base economy in the same group.

Since all eurozone countries except Germany do not have enough default events prior to joining the eurozone, their benchmark interest rate is entered as zero for that period. Among the non-eurozone members of the European group, Denmark, Norway, Sweden and the UK each have separate coefficients for the benchmark interest rate. Switzerland and Iceland do not use this variable for their whole history.

In the Developed Asia-Pacific group, all economies have their own coefficient for the benchmark interest rate. For the North American group, both Canada and the US have their own coefficient for the benchmark interest rate. In the Emerging Markets group, there are insufficient defaults in the Latin American economies to calibrate individual economy benchmark interest rate coefficients in a statistically significant way, so all Latin American economies share the same benchmark interest rate coefficient. Each of the Asian economies in the Emerging Markets group, namely Indonesia, Malaysia, Philippines and Thailand, have their own coefficient for the benchmark interest rate.

### 3.4. Daily Output

**Individual firms' PD:** In computing the pseudo-log-likelihood functions in equation (18), only end of month data is needed. The data needs to be extended to daily values in order to produce daily PDs.

For the level variables, the last twelve end of month observations (before averaging) are combined with the current value. The current value is scaled by a fraction equal to the current day of the month divided by the number of calendar days in the month. The earliest monthly value is scaled by one minus this fraction. The sum is then divided by the number of valid monthly observations, with the current value and the earliest monthly value counting as a single observation if either or both are not missing. Not performing this scaling can lead to an artificial jump in PD at the beginning of the month. When performing the scaling, the change in level is more gradual throughout the month.

A similar procedure is done for SIGMA. Here the earliest month is not scaled, but the return from the current day to the previous month end is scaled by the square root of the fraction equal to the current day of the month divided by the number of calendar days in the month.

Computing the DTD for all firms on a daily basis using the two stage process described in Subsection 3.2 would be time consuming, even on the grid. Since there should be little change to $\sigma$ and $\delta$ on a day to day basis, for the daily computation of DTD these are assumed to have the same value as in the previous month's DTD calculation. In other words, the previous month's values for $\sigma$ and $\delta$ together with the new day's equity value are used in equation (20) to obtain the implied asset value. This implied asset value with the previous month's values for $\sigma$ and $\delta$ is used in equation (22) to obtain the new day's DTD, with $\mu$ set to equal to $\sigma^2/2$.

**Aggregating PD:** The CRI provides term structures of the probability distributions for the number of defaults as well as the expected number of defaults for different groups of firms. The companies are grouped by economy (using each firm's country of domicile), by sector (using the firm's Bloomberg industrial sector code) and sectors within economies. With the individual firms' PD, the expected number of defaults is trivial to compute. The algorithm used to compute the probability distribution of the number of defaults was originally reported in Anderson *et al*. (2003). It assumes conditional independence and uses a fast recursive scheme to compute the necessary probability distribution.

Note that while this algorithm is currently used to produce the probability distribution of the number of defaults within an economy or sector, it can easily be generalized to compute loss distributions for a portfolio manager, where the exposure of the portfolio to each firm needs to be input.

**Inclusion of firms in aggregation:** As explained in Subsection 3.1, firms are included in an economy for calibration if the firms' primary listing is on an exchange in that economy. This is to ensure that all firms in an economy are subject to the same disclosure and accounting requirements. In contrast, a firm is included in an economy's aggregate results if the firm is domiciled in that economy. This is because users typically associate firms with their economy of domicile rather than the economy where their primary listing is, if they are different. For example, the Bank of China has its primary listing in Hong Kong, but its economy of domicile is China so the Bank of China is included in the aggregation forecasts for China, and is included under China when searching for the individual PDs.

# IV. EMPIRICAL ANALYSIS

This section presents an empirical analysis of the CRI outputs for the 37 economies that are currently being covered. In Subsection 4.1, an overview is given of the default parameter estimates. Subsection 4.2 explains and provides the accuracy ratios for the different countries under the CRI cover.

## 4.1. Parameter Estimates

With 24 months of forecast horizons, 13 variables and 6 different groups of economies, tables of the parameter estimates occupy over 20 pages and are not included in this Technical Report. They are available in the section on the technical report at the CRI web portal. In the Annex, the parameter estimates are from calibrations performed in June 2012 using data up until the end of May 2012. In this part, a brief overview is given of the general traits and patterns seen in the default parameter estimations of the economies covered by the CRI.

Recall that if a default parameter for a variable at a particular horizon is estimated to be positive (resp. negative) from the maximum pseudo-likelihood estimate, then an increasing value in the associated variable will lead to an increasing (decreasing) value of the forward intensity at that horizon, which in turn means an increasing (decreasing) value for the conditional default probability at that horizon.

For the stock index one-year trailing return variable, most groups have default parameters that are slightly negative in the shorter horizons and then become positive in the longer horizons. When the equity market performs well, this is only a short-term positive for firms and in the longer term, firms are actually more likely to default. This seemingly counter-intuitive result could be due to correlation between the market index and other firm-specific variables. For example, Duffie *et al*. (2009) suggested that a firm's distance-to-default (DTD) can overstate its credit-worthiness after a strong bull market. If this is the case, then the stock index return serves as a correction to the DTD levels at these points in time.

The default parameters for the short-term interest rate variable are significantly positive at one- to two-year horizons for most of the economies. This is consistent with the intuition that increasing short-term interest rates typically signal increased funding costs for companies in the future, increasing the probability of default. The values at shorter horizons are varied between economies from slightly negative to significantly positive, possibly indicating different lead-lag relationships between credit conditions and the raising and cutting of short-term interest rates.

DTD is a measure of the volatility-adjusted leverage of a firm. Low or negative DTD indicates high leverage and high DTD indicates low leverage. Therefore, PD would be expected to increase with decreasing DTD. Indeed, almost all of the calibrations for the different groups lead to negative default parameters for the DTD level, with only China's default parameter estimations hitting the constraint at zero for longer horizons.

The ratio of the sum of cash and short-term investments to total assets (CASH/TA) measures liquidity of a firm. This indicates the availability of a firm's funds and its ability to make interest and principal payments. As expected, for almost all economies (Indonesia being the only exception) the default parameters for CASH/TA level in shorter horizons are significantly negative. The magnitude of the default parameters decreases for longer horizons, indicating that CASH/TA level is a better indicator of a firm's ability to make payments in the short term than the long term.

The ratio of net income to total assets (NI/TA) measures profitability of a firm. The relationship between PD and NI/TA is as expected: the default parameters for NI/TA level is significantly negative for most economies and most horizons.

The logarithm of the market capitalization of a firm over the median market capitalization of firms within the economy (SIZE) does not have a consistent effect on PD across different economies. For example, in the US the default parameters for SIZE level are negative for shorter horizons and positive for longer horizons, suggesting that the advantages enjoyed by larger firms, such as diversified business lines and funding sources, are a benefit in the shorter term but not in the longer term. On the other hand, in Japan the default parameters for SIZE level are negative across all horizons. These differences may reflect differences in the business environments in the respective economies.

The default parameters associated with DTD Trend, CASH/TA Trend and SIZE Trend, are negative across almost all economies and horizons. The trend variables reflect momentum. The momentum effect is a short-term effect, and evidence of this is seen in the lower magnitude of the default parameters at longer horizons than at shorter horizons. The remaining trend variable is the NI/TA Trend. The current implementation of the CRI system retrieves net income only from annual financial statements. The default parameters for NI/TA Trend are constrained to be negative, but for most economies there is no clear relationship between the NI/TA Trend and the horizon. Once NI/TA from quarterly statements can be used, this will likely be more informative.

The ratio of the sum of market capitalization and total liabilities to total assets (M/B) can either indicate the market mis-valuation effect or the future growth effect. This default parameter is positive in most economies, indicating that higher M/B implies higher PD, and the market mis-valuation effect dominates.

Shumway (2001) argued that a high level of the idiosyncratic volatility (SIGMA) indicates highly variable stock returns relative to the market index, indicating highly variable cash flows. Volatile cash flows suggest a heightened PD, and this finding is consistent across all economies and most horizons, with the exception of India.

### 4.2. Prediction Accuracy

**In-sample and out-of-sample testing:** Various tests are carried out to test the prediction accuracy of the CRI PD forecasts. These tests are conducted either in-sample or out-of-sample.

A single calibration is conducted for the in-sample tests, using data to the end of the data sample. As an example, one-year PD forecasts are made for December 31, 2000 by using the data at or before December 31, 2000 and the parameters from the calibration. These PD forecasts can be compared to actual defaults that occurred at any time in 2001.

The out-of-sample analysis is done over time. The first calibration is conducted using only data up to the end of December 2000. For example, one-year PD forecasts can be made for December 31, 2000 using the data at or before December 31, 2000 with the parameters from this first calibration. These are PD forecasts that could have been made at the time, since the parameters are not based on data available after that date. This process is repeated every month. That is, the second calibration is conducted using only data up to the end of January 2001, and so on.

It should be noted that for these repeated calibrations based on an expanding window of data, nothing else is changed besides the dataset. In other words, the same choice of input variables and the same choice of economy dummies within the groups are used throughout all of the calibrations.

Some of the calibration groups have too few defaults in the period before December 2000 to be able to produce stable calibration results. If this is the case, the start date is advanced. Subsequently, if there are too few defaults after the start date to perform meaningful tests, only in-sample tests are performed for that calibration group. Out-of-sample tests are performed for (starting month of calibration in parentheses): China (12/2000), Japan (12/2003), India (12/2001), South Korea (12/2000), Developed Asia (12/2000), Emerging Markets (12/2000), North America group (12/2000), Western Europe 2 group (12/2002).

**Accuracy Ratio:** The accuracy ratio (AR) is one of the most popular and meaningful tests of the discriminatory power of a rating system (BCBS, 2005). The AR and the equivalent Area Under the Receiver Operating Characteristic (AUROC) are described in Duan and Shrestha (2011). In short, if defaulting firms had been assigned among the highest PD of all firms before they defaulted, then the model has discriminated well between "safe" and "distressed" firms. This leads to higher values of AR and AUROC. The range of possible AR values is in [0,1], where 0 is a completely random rating system and 1 is a perfect rating system. The range of possible AUROC values is in [0.5, 1]. AUROC and AR values are related by: $AR = 2 \times AUROC - 1$.

The AR and AUROC values for different horizons are available in the section on the technical report at the CRI web portal. In the Annex, both in-sample and out-of-sample results are available for calibration groups where out-of-sample testing could be performed. Other calibration groups include only in-sample results. The in-sample AR and AUROC are computed only from the starting date of the corresponding out-of-sample tests, so that the results between in-sample and out-of-sample are comparable. Only economies with more than 20 defaults entering into the AR and AUROC computation are listed. The PD are taken to be non-overlapping. For example, the one-year AR is based on PDs computed on 31/12/2000, 31/12/2001, … , 31/12/2009 and firms defaulting within one year of those dates, while the two-year AR is based on PDs computed on 31/12/2000, 31/12/2002, … , 31/12/2008 and firms defaulting within two years of those dates.

The AUROC values have been provided only for the purpose of comparison, if other rating systems report their results in terms of AUROC. The discussion will focus only on AR. The model is able to achieve strong AR results mostly greater than 0.80 at the one and six-month horizons for developed economies. There is a drop in AR at one and two-year horizons, but the AR are still mostly acceptable. Australia, the UK and Singapore have sharp drops in AR at the two-year horizon. Hong Kong has comparatively worse AR over all horizons as compared to other developed economies.

The AR in emerging market economies such as China, India, Indonesia, Malaysia, Philippines and Thailand 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 a developed economy, 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 less reluctant to default. The AR for the Latin American economies inside the emerging economies group are generally greater than 0.80 at horizons shorter than one year. However, these AR are for a small number of defaults.

At horizons of one and six-months, out-of-sample AR are comparable to their in-sample counterparts. At horizons of one and two-years, out-of-sample AR can be substantially lower than the in-sample AR.

Finally, the US has a sufficient number of financial firms and financial defaults to produce separate AR and AUROC. These are also listed in the Annex as out-of-sample results. The financial sector ARs are actually stronger than the non-financial sector AR. This is achieved by having only minimal differences between how financial and non-financial firms are treated.

The AR is a test of discriminatory power, or how well the rating system ranks firms in terms of credit worthiness. In a separate article included in the GCR Volume 2, we provide a more qualitative check on the CRI PD in which we compare the behaviour of CRI PD to the rating actions of external credit rating agencies such as Moody's and S&P for some well known default cases.

**Aggregate defaults:** The time series of aggregate predicted number of defaults and actual number of defaults in each calibration group are also available in the Annex.

## V. ONGOING DEVELOPMENTS

The CRI can be developed along a number of directions. We now comment on obvious ones that in our view are likely to bring meaningful and measurable benefits. Besides modifications to the current modeling framework of the forward intensity, a change in modeling platform will be undertaken if another model proves more promising in terms of accuracy and robustness of results. For this type of development we also rely on the collective efforts by the worldwide credit research community to challenge and improve the existing modeling platform.

The current CRI default prediction model is based on the econometric platform of modeling forward intensities developed by Duan *et al*. (2012). As noted by them, the forward-intensity model exhibits systematic bias in predicting longer-term defaults of the US corporate sector. In general, it overestimates (underestimates) defaults when default rates were low (high). Introducing a frailty variable to the model to capture default contagion appears to be one possibility to further improve the model.

In addition, some of the likely future developments of CRI fall in the domain of further infrastructure developments at RMI. For example, by end 2012, all exchange-listed firms in all economies around the globe should be covered. Furthermore, in terms of variable selection, more experiments are needed to identify common risk factors and RMI specific attributes that are more indicative of defaults in different economies. Also in terms of grouping, further tests should be conducted, especially as new economies will be covered. It is also worth noting that all variables used thus far in the CRI implementation are the quantitative type. Soft credit information as reflected in qualitative opinions of credit analysts may add an important dimension to future improvement. To this end, the CRI has been conducting a continuous credit analysts survey, and at this point of writing there are about 100 analysts participating in the survey. It is quite obvious that we have to expand this base of this survey in order to allow meaningful incorporation into the default prediction.

*******************************************

The RMI Credit Research Initiative is premised on the concept of credit ratings as a "public good". Being a non-profit undertaking allows a high level of transparency and collaboration that other commercial credit rating systems can not replicate. The research and support infrastructure is in place and researchers from around the world are invited to contribute to this initiative. Any methodological improvements that researchers develop will be incorporated into the CRI system. In essence, the initiative operates as a "selective wikipedia" where many can contribute but implementation control is retained.

If you have feedback on this technical report or wish to work with us in this endeavor, please contact us at rmicri@globalcreditreview.com

## REFERENCES


Anderson, L., J. Sidenius and S. Basu (2003), All your Hedges in one Basket. *Risk*, pp. 67–72.

Crosbie, P. and J. Bohn (2003), Modeling Default Risk. Moody's KMV technical document.

Duan, J.-C. (1994), Maximum Likelihood Estimation Using Price Data of the Derivative Contract. *Mathematical Finance*, 4, pp. 155–167.

Duan, J.-C. (2000), Correction: "Maximum Likelihood Estimation Using Price Data of the Derivative Contract". *Mathematical Finance*, 10, pp. 461–462.

Duan, J.-C. (2010), Clustered Defaults. National University of Singapore Working Paper.

Duan, J.-C. and K. Shrestha (2011), Statistical Credit Rating Methods, *Global Credit Review*, pp. 43–64.

Duan, J.-C., J. Sun and T. Wang (2012), Multiperiod Corporate Default Prediction — A Forward Intensity Approach. *Journal of Econometrics*, forthcoming (DOI: 10.1016/j.jeconom.2012.05.002).

Duan, J.-C. and E. Van Laere (2012), A Public Good Approach to Credit Ratings: From Concept to Reality. *Journal of Banking & Finance*, forthcoming (DOI: 10.1016/j.jbankfin.2012.03.012).

Duan, J.-C. and T. Wang (2012), Measuring Distance-to-Default for Financial and Non-financial Firms. *Global Credit Review*, pp. 95–108.

Duffie, D., A. Eckner, G. Horel and L. Saita (2009), Frailty Correlated Default. *Journal of Finance*, 64, pp. 2089–2123.

Duffie, D., L. Saita, and K. Wang (2007), Multi-Period Corporate Default Prediction with Stochastic Covariates. *Journal of Financial Economics*, 83, pp. 635–665.

Merton, R. C. (1974), On the Pricing of Corporate Debt: The Risk Structure of Interest Rates. *Journal of Finance*, 29, pp. 449–470.

Shumway, T. (2001), Forecasting Bankruptcy More Accurately: A Simple Hazard Model. *Journal of Business*, 74, pp. 101–124.

# APPENDIX A

**Table A.1** ISO codes for economies currently covered by the CRI and the group that each economy is calibrated in.

| ISO Code | Economy | Calibration Group |
|---|---|---|
| ARG | Argentina | Emerging |
| AUS | Australia | Developed Asia-Pacific |
| AUT | Austria | Europe |
| BEL | Belgium | Europe |
| BRA | Brazil | Emerging |
| CAN | Canada | North America |
| CHE | Switzerland | Europe |
| CHL | Chile | Emerging |
| CHN | China | China |
| COL | Colombia | Emerging |
| DEU | Germany | Europe |
| DNK | Denmark | Europe |
| ESP | Spain | Europe |
| FIN | Finland | Europe |
| FRA | France | Europe |
| GBR | United Kingdom | Europe |
| GRC | Greece | Europe |
| HKG | Hong Kong | Developed Asia-Pacific |
| IDN | Indonesia | Emerging |
| IND | India | India |
| ISL | Iceland | Europe |
| ITA | Italy | Europe |
| JPN | Japan | Developed Asia-Pacific |
| KOR | South Korea | Developed Asia-Pacific |
| MEX | Mexico | Emerging |
| MYS | Malaysia | Emerging |
| NLD | Netherlands | Europe |
| NOR | Norway | Europe |
| PER | Peru | Emerging |
| PHL | Philippines | Emerging |
| PRT | Portugal | Europe |
| SGP | Singapore | Developed Asia-Pacific |
| SWE | Sweden | Europe |
| THA | Thailand | Emerging |
| TWN | Taiwan | Developed Asia-Pacific |
| USA | United States | North America |
| VEN | Venezuela | Emerging |

**Table A.2** The stock indices used for each economy in computing the first common variable.

| Country | Stock Exchange | Period Used |
|---|---|---|
| ARG | Buenos Aires Stock Exchange Merval Index | |
| AUS | All Ordinaries Index | |
| AUT | Austrian Traded ATX Index | |
| BEL | Belgian All Shares Return Index | |
| BRA | Brazil Bovespa Stock Index | |
| CAN | S&P/TSX Composite Index | |
| CHE | SPI Swiss Performance Index | |
| CHL | Santiago Stock Exchange IPSA Index | |
| CHN | Shanghai Stock Exchange Composite Index | |
| COL | FTSE All World Series Colombia Local | |
| DEU | CDAX Performance Index | |
| DNK | OMX Copenhagen 20 Index | |
| ESP | IBEX 35 Index | |
| FIN | OMX Helsinki Index | |
| FRA | CAC 40 Index | |
| GBR | FTSE 100 Index | |
| GRC | Athex Composite Share Price Index | |
| HKG | Hang Seng Index | |
| IDN | Jakarta Composite Index | |
| IND | BSE Sensex 30 Index | |
| ISL | OMX Iceland All-Share Price Index | |
| ITA | Italy Stock Market BCI Comit Global | |
| JPN | Nikkei 500 | |
| KOR | KOSPI Index | |
| MEX | Mexico Bolsa Index | |
| MYS | FTSE Bursa Malaysia KLCI | |
| NLD | AEX Index | |
| NOR | OBX Price Index | |
| PER | Bolsa de Valores de Lima General Sector Index | |
| PHL | PSEI-Philippine Stock Exchange Index | |
| PRT | PSI General Index | |
| SGP | Straits Times Index | 1/10/2008–Present |
| | Straits Times Old Index | 8/31/1999–1/9/2008 |
| SWE | OMX Stockholm All-Share Index | |
| THA | Stock Exchange of Thailand Index | |
| TWN | Taiwan Taiex Index | |
| USA | S&P 500 Index | |
| VEN | Caracas Stock Exchange Stock Market Index | |

*A blank Period Used column indicates that there is only a single index that is used throughout the whole period.

**Table A.3** The interest rates used for each economy as the second common variable.

| Country | Short Term Interest Rate | Period Used |
|---|---|---|
| ARG | Argentina Deposit 90 Day | |
| AUS | Australia Dealer Bill 90 Day | |
| AUT | Germany 3 Month Bubill | 1/1/1999–Present |
| | — | –12/31/1998 |
| BEL | Germany 3 Month Bubill | 1/1/1999–Present |
| | — | –12/31/1998 |
| BRA | Andima Brazil Govt Bond Fixed Rate 3 Months | 4/3/2000–Present |
| | Brazil CDB (up to 30 Days) | 10/10/1994–3/31/2000 |
| CAN | Canada Treasury Bill 3 Month | |
| CHE | — | |
| CHL | Chile TAB UF Interbank Rate 90 Days | |
| CHN | China Time Deposit Rate 3 Month | |
| COL | Colombia CD Rate 90-Day | |
| DEU | Germany 3 Month Bubill | 5/25/1993–Present |
| | Germany Interbank 3 Month | 1/2/1986–5/24/1993 |
| DNK | Denmark Interbank 3 Month | |
| ESP | Germany 3 Month Bubill | 1/1/1999–Present |
| | — | –12/31/1998 |
| FIN | Germany 3 Month Bubill | 1/1/1999–Present |
| | — | –12/31/1998 |
| FRA | Germany 3 Month Bubill | 1/1/1999–Present |
| | — | –12/31/1998 |
| GBR | UK Treasury Bill Tender 3 Month | |
| GRC | Germany 3 Month Bubill | 1/1/2001–Present |
| | — | –12/31/2000 |
| HKG | Hong Kong Exchange Fund Bill 3 Month | |
| IDN | Indonesia SBI 90 Day | 7/10/2003–Present |
| | Indonesia SBI/DISC 90 Day | 1/1/1985–7/9/2003 |
| IND | India T-Bill Secondary 91 Day | |
| ISL | — | |
| ITA | Germany 3 Month Bubill | 1/1/1999–Present |
| | — | –12/31/1998 |
| JPN | Japan Treasury Discount Bills 3 Month | 7/10/1992–Present |
| | Japanese Government Bond Interest Rate-1 Year Maturity | 9/24/1974–7/9/1992 |
| KOR | Korea Commercial Paper 91 Day | |
| MEX | Mexico Cetes 2ND MKT. 90 Day | 6/26/1996 – Present |
| | Mexico Cetes 91 Dat AVG.RET.AT AUC. | 3/9/1989–6/25/1996 |
| MYS | Malaysia Deposit 3 Month | |
| NLD | Germany 3 Month Bubill | 1/1/1999–Present |
| | — | –12/31/1998 |
| NOR | Norway Govt Treasury Bills 3 Month | 6/27/1995–Present |
| | Norway Interbank 3 Month(effective) | 1/2/1986–6/26/1995 |
| PER | Peru Savings Rate | |
| PHL | Philippine Treasury Bill 91 Day | |
| PRT | Germany 3 Month Bubill | 1/1/1999–Present |
| | — | –12/31/1998 |
| SGP | Singapore T-Bill 3 Month | |
| SWE | Sweden T-Bill 3 Month | 5/25/1993–Present |
| | Sweden Treasury Bill 90 Day | 4/25/1989–5/24/1993 |
| THA | Thailand Repo 3 Month(BOT) | |
| TWN | Taiwan Money Market 90 Day | |
| USA | US Generic Govt 3-Month Yield | |
| VEN | Venezuela Overnight | |

*A blank Period Used column indicates that there is only a single interest rate that is used throughout the whole period.

**Table A.4** The interest rates used for each economy in the DTD calculation.

| Country | Interest Rate Name | Period Used |
|---|---|---|
| ARG | Argentina Deposit 90 Day (PA.) | |
| AUS | Australia Govt. Bonds Generic Mid Yield 1 Year | |
| AUT | German Government Bonds 1 Year BKO | 1/1/1999–Present |
| | Austria VIBOR 12 Month | 6/10/1991–12/31/1998 |
| BEL | German Government Bonds 1 Year BKO | 1/1/1999–Present |
| | Belgium Treasury Bill 1 Year | 4/2/1991–12/31/1998 |
| BRA | Andima Brazil Govt Bond Fixed Rate 1 Year | 4/3/2000–Present |
| | BRAZIL CDB (UP TO 30 DAYS) | 10/10/1994–3/31/2000 |
| CAN | Canada Treasury Bill 1 Year | |
| CHE | Swiss Interbank 1 Year (ZRC:SNB) | |
| CHL | Chile TAB UF Interbank Rates 360 Days | 8/1/1996–Present |
| | Chile TAB UF Interbank Rate 90 Days | 11/2/1992–7/30/1996 |
| CHN | China Household Savings Deposits 1-Year Rate | |
| COL | Colombia Government Generic Bond 1 Year Yield | 3/1/2001–Present |
| | Colombia CD Rate 360-Dat | 7/12/1993–2/8/2001 |
| DEU | German Government Bonds 1 Year BKO | 1/10/1995–Present |
| | Germany Interbank 12 Month | 11/2/1990–1/9/1995 |
| DNK | Denmark Government Bonds 1 Year Note Generic Bid Yield | 6/1/2008–Present |
| | Denmark Euro-Krone 1 Year(FT/ICAP/TR) | 6/14/1985–5/31/2008 |
| ESP | German Government Bonds 1 Year BKO | 1/1/1999–Present |
| | Spain 12 Month Treasury Bill Yield | 11/30/1992–12/31/1998 |
| | Spain Interbank 12 Month | 12/19/1991–11/29/1992 |
| FIN | German Government Bonds 1 Year BKO | 1/1/1999–Present |
| | Finland Interbank Close 12 Month | 4/2/1992–12/31/1998 |
| FRA | German Government Bonds 1 Year BKO | 1/1/1999–Present |
| | France Treasury Bill 12 Months | 1/3/1989–12/31/1998 |
| GBR | UK Govt. Bonds 1 Year Note Generic | 9/12/2001–Present |
| | UK Govt. Liability Nominal Spot Curve 12 Month | 12/13/1985–9/11/2001 |
| GRC | German Government Bonds 1 Year BKO | 1/1/2001–Present |
| | Greece Treasury Bill 1 Year | 1/2/1990-12/31/2000 |
| HKG | HKMA Hong Kong Exchange Fund Bill 12 Month | |
| IDN | Indonesia SBI 90 Day | 7/10/2003–Present |
| | Indonesia SBI/DISC 90 Day | 1/1/1985–7/9/2003 |
| IND | India T-Bill Secondary 1 Year | |
| ISL | Iceland Interbank 12 Month | 2/1/2000–Present |
| | Iceland Interbank 3 Month | 8/4/1998–1/31/2000 |
| | Iceland 90-day CB Notes | 5/12/1987–8/3/1998 |
| ITA | German Government Bonds 1 Year BKO | 1/1/1999–Present |
| | Italy Bots Treasury Bill 12 Month Gross Yields | 9/5/1994–12/31/1998 |
| | Italy T-Bill Auction Gross 12 Month | 3/31/1987–9/4/1994 |
| JPN | Japan Treasury Bills 12 Month | 12/14/1999–Present |
| | Japanese Government Bond Interest Rate-1 Year Maturity | 9/24/1979–12/13/1999 |
| KOR | Korea Monetary Stabilization Bonds 1 Year | |
| MEX | Mexico Cetes 2ND MKT. 360 Day | 6/26/1996 –Present |
| | Mexico Cete 91 DAY AVG.RET.AT AUC. | 3/9/1989– 6/25/1996 |
| MYS | Bank Negara Malaysia 1 Year Govt. Securities Indicative YTM | 6/21/2005–Present |
| | Malaysia Deposit 1 Year | 1/1/1985–6/20/2005 |
| NLD | German Government Bonds 1 Year BKO | 1/1/1999–Present |
| | Netherland Interbank 1 Year | 1/2/1987–12/31/1998 |
| NOR | Norway Govt Treasury Bills 12 Month | 7/1/1997–Present |
| | Norway Interbank 1 Year | 1/2/1986–6/30/1997 |
| PER | Peru Savings Rate | |
| PHL | Philippine Treasury Bill 364 Day | |

(*Continued*)

**Table A.4** (*Continued*)

| Country | Interest Rate Name | Period Used |
|---|---|---|
| PRT | German Government Bonds 1 Year BKO | 1/1/1999–Present |
| | Portugal 1-Year-LISBOR-Act/365 Day convention | 8/16/1993–12/31/1998 |
| SGP | Singapore T-Bill 3 Month | |
| SWE | Sweden Interbank 1 Year | 5/25/1993–Present |
| | Sweden Treasury Bill 1 Year Note | 4/25/1989–5/24/1993 |
| THA | Thailand Govt. Bond 1 Year Note | 8/7/2000–Present |
| | Thailand Deposit 12 Month(KT) | 1/2/1991–8/6/2000 |
| TWN | Taiwan Deposit 12 Month | |
| USA | US Treasury Constant Maturities 1 Year | |
| VEN | Venezuela Overnight | |

*A blank Period Used column indicates that there is only a single interest rate that is used throughout the whole period.

**Table A.5** Exits classified as "Defaults".

| Default | |
|---|---|
| **Action Type** | **Subcategory** |
| Bankruptcy filing | Administration, Arrangement, Canadian CCAA, Chapter 7, Chapter 11, Chapter 15, Conservatorship, Insolvency, Japanese 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, Other, Unknown |
| Delisting | Bankruptcy |
| Default Corporate Action | Bankruptcy, Coupon & Principal Payment, Coupon Payment Only, Debt Restructuring, Interest Payment, Loan Payment, Principal Payment, ADR (Japan only), Declared Sick (India Only), Unknown |

**Table A.6** Exits classified as "Other Exits".

| Other Exits | |
|---|---|
| **Action Type** | **Subcategory** |
| Delisting | Unknown, Acquired/Merged, Assimilated with underlying shares, Bid price below minimum, Cancellation of listing, End of When-issued trading, Expired, 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, Matured, Not available, Not current in required filings, NP/FP finished, Privatized, Reorganization security called for redemptions, the company's request, Scheme of arrangement, Insufficient spread of holders, Selective capital reduction of the company |