Publication Date: Apr 26, 2022
Effective Date: Apr 26, 2022
Addendum 3 to the CRI Technical Report (Version: 2021, Update 1)

This document updates the Technical Report (Version: 2021, Update 1) and details the implementation of default correlations (DCs) to produce default-rate (or default-number) distributions for different credit portfolios and horizons. Default-rate and default-number distributions are interchangeable given the total number of companies at the assessment time, and the NUS-CRI website only releases default-number distributions. The changes have been implemented for the release starting 26 Apr 2022.

The DCs between entities in a credit portfolio of interest are essential to producing default-rate distributions that go beyond marginal PDs and POEs. As can be expected, inclusion of DCs among corporate entities makes default-rate distributions far more right-skewed, reflecting a much higher chance that obligors in a typical credit portfolio may default together due to their reactions to a common shock such as commodity price, interest rate, exchange rate, etc.

Calculation of DCs by NUS-CRI is largely based on the methodology of Duan and Miao (2016)[1] except for factor model selection and part of re-calibration. Beyond research, computing for so many NUS-CRI portfolios (regions, countries/economies, sectors) on a regular basis requires of algorithmic improvements and additional computing resources. Prior to this release, the NUS-CRI default-rate distributions were computed under the assumption of zero default correlations.

Computing default-rate distributions comprises four steps: (1) identify a set of predetermined credit risk factors, estimate the factor model, and produce the factor model residuals; (2) estimate the time series dynamics of the predetermined credit risk factors and individual factor model residuals; (3) construct a sparse correlation matrix for the factor model residuals after taking out their individual time series effect; (4) further calibrate the model to the term structure of PDs at the time of application to take advantage of the information embedded in longer-term PDs. The details of the four steps are as follows.

I. Generating the common factor model

The first step is to estimate the factor model by using the transformed one-month PDs and POEs. Let n_t be the total number of firms with PDs and POEs at time t, and T be the time point for credit analysis. Let $p_{i,t}(l)$ and $q_{i,t}(l)$ denote the l-month PD and POE of firm i at month t for $i = 1,\cdots,n_t$ and $t = 1,\cdots,T$. Because PDs and POEs are naturally bounded between 0 and 1, the following logit transformation is first applied to the PDs and POEs to take them back to the real line:

$$P_{i,t} = \ln\{p_{i,t}(1)/[1 - p_{i,t}(1)]\} \text{ and } Q_{i,t} = \ln\{q_{i,t}(1)/[1 - q_{i,t}(1)]\}$$

The firm's likelihood of default is influenced by the global, industry-specific and econ-specific credit factors. Thus, credit cycle indices (CCIs) are created by taking the logit-transformed global, industry and economy median PDs and POEs as credit factors. The way of constructing CCIs is akin to creating stock market indices except that medians are used instead of means. To be more relevant to the economic environment, other factors, including FX rate and interest rate, will also be considered. HP filter, a method to decompose time series, is applied to FX rate and interest rate respectively to

[1] Duan, J.C., W. Miao, 2016, Default Correlations and Large-Portfolio Credit Analysis. Journal of Business and Economic Statistics 34, 51-65.

generate cycle and trend. Only rate cycles are included in the factor model. In total, there are around 420 potential common factors[2] to select from every month. The PD- and POE-CCIs as constructed above are certainly correlated. The industry-specific and economy-specific CCIs are thus orthogonalized using the global PD-CCI and POE-CCI pair. The FX rate and interest rate factors are also orthogonalized using the global CCI pair as well as the corresponding economy-specific CCI pair.

The firms are divided into 19 groups based on their industry sectors according to NUS-CRI Industry Classification Standard 2020. Then variable selection is performed separately for the PD and POE series to obtain at most 15 common factors for each group using the zero-norm variable selection technique of Duan (2019) where the selection target is to maximize the average R^2 of 1000 randomly selected members of the group.[3] For a PD factor model, the global PD-CCI and the industry-specific PD-CCI (the industry to which the majority of the companies belong), are always in the factor model as the 'must-include' variables in the variable selection algorithm. A POE factor model is likewise constructed with global POE-CCI and the same industry-specific POE-CCI as the 'must-include' variables. The selected factor model is applicable to all members of a group.

For individual firms, each one is expected to respond to credit risk factors in a different manner, and the channels of influence can be identified by regressing each firm's transformed one-month PDs and POEs on the common factors. A factor model is constructed as below, where $\boldsymbol{F}_t{}^{(P)}$ and $\boldsymbol{F}_t{}^{(Q)}$ refer to the selected factors in PD and POE factor models respectively.

$$P_{i,t} = \beta_{0,i}{}^{(P)} + \boldsymbol{\beta}_i{}^{(P)}\boldsymbol{F}_t{}^{(P)} + \varepsilon_{i,t}{}^{(P)}, \qquad t = 1, \dots, T \quad (1)$$

$$Q_{i,t} = \beta_{0,i}{}^{(Q)} + \boldsymbol{\beta}_i{}^{(Q)}\boldsymbol{F}_t{}^{(Q)} + \varepsilon_{i,t}{}^{(Q)}, \qquad t = 1, \dots, T \quad (2)$$

II. Considering time dynamics

The next step is to estimate time dynamics for both factors and individual factor-model residuals by using VAR (Vector Auto-Regression) and AR (Auto-Regression) model. The first-order AR model is used to account for the autocorrelation in the residual dynamics:

$$\varepsilon_{i,t}{}^{(P)} = \mu_i{}^{(P)} + \rho_i{}^{(P)}\varepsilon_{i,t-1}{}^{(P)} + e_{i,t}{}^{(P)}, \qquad t = 1, \dots, T \quad (3)$$

$$\varepsilon_{i,t}{}^{(Q)} = \mu_i{}^{(Q)} + \rho_i{}^{(Q)}\varepsilon_{i,t-1}{}^{(Q)} + e_{i,t}{}^{(Q)}, \qquad t = 1, \dots, T \quad (4)$$

For global PD-CCI and POE-CCI pair, a first-order two-dimensional vector autoregression (VAR) is used to capture the time dynamics. In this factor dynamic model, $\boldsymbol{F}_t$ represents the global CCI vector, $\boldsymbol{A}$ is a time-invariant 2-by-2 square matrix and $\boldsymbol{E}_t$ is a vector of two error terms.

$$\boldsymbol{F}_t = \boldsymbol{A}\boldsymbol{F}_{t-1} + \boldsymbol{E}_t, \qquad t = 1, \dots, T \quad (5)$$

The dynamic models for other common factors including industry/economy CCIs, FX rates and interest rates, a first-order AR model is built for each of them. Here Y_t denotes the corresponding factor, b_0 is the intercept, b is the coefficient of the lag-1 term, and e_t is the error term.

---

[2] The 420 potential common factors include 282 PD- and POE-CCIs (141 for PD- and POE-CCIs respectively, which include 1 global factor, 11 industry factors and 129 economy factors) and 138 other factors, which include 65 FX cycles and 73 interest rate cycles. The total number of potential factors may change every month due to data availability.

[3] Duan, J.C., 2019, Variable Selection with Big Data based on Zero Norm and via Sequential Monte Carlo, National University of Singapore working paper. Duan's paper deals with variable selection for one regression, but we apply it to a multiple-equation setting where the selected regressors are common to all equations. The selection target is to maximize an equally-weighted average R^2 over all equations.

$$Y_t = b_0 + bY_{t-1} + e_t, \qquad t = 1, \ldots, T \quad (6)$$

III. Constructing the sparse residual correlation matrix

To better capture co-movements of credit risks across firms, supplementary information on co-movements from the PD and POE residuals, ${\varepsilon_{i,t}}^{(P)}$ and ${\varepsilon_{i,t}}^{(Q)}$ in equations (1) and (2), are also extracted. We stack together all AR model residual terms ${e_{i,t}}^{(P)}$ and ${e_{i,t}}^{(Q)}$ in equations (3) and (4) to form a matrix of size $T \times 2n_T$ at time T. Having employed the common risk factors, we have reasons to assume that most pairs of AR residual terms are uncorrelated, resulting in a sparse correlation matrix. Due to missing data, only pairwise correlations are available, and statistical significance can be used as a simple way to filter out insignificant pairwise correlations. The SCAD-thresholding element-wise method is then applied to achieve both sparsity and positive semi-definiteness.

IV. Calibrating to PD term structures

Up until this point, only historical time series of one-month PDs and POEs have been utilized in the estimation of the factor model with sparsely correlated residuals. Since term structures of PDs are available in the NUS-CRI database, further calibration of the factor model to term structures of PDs can take advantage of the additional information.

The factor model in equations (1)-(2), the factor dynamics in equations (5)-(6), the residuals dynamics model in equations (3)-(4), and the sparse residual correlation matrix can be combined to generate future paths of the one-month PDs and POEs for any group of obligors over any horizon of interest. With one set of simulated paths in place, one can then compute by the standard survival/default formula the randomized individual default probabilities conditional on the paths. These conditional individual default probabilities can then be averaged over the simulated paths to arrive at PDs for different horizons and obligors, which should in principle match up with their observed term structure of PDs. In reality, model misspecification and estimation errors will prevent two sets to match exactly. We thus conduct a re-calibration of the factor model's individual residuals dynamics, i.e., ${\varepsilon_{i,t}}^{(P)}$ and ${\varepsilon_{i,t}}^{(Q)}$, to minimize the mismatch. For firm i, re-calibration is to search for the six unknown parameters in equations (3)-(4), i.e., $\theta_i \coloneqq ({\mu_i}^{(P)}, {\mu_i}^{(Q)}, {\rho_i}^{(P)}, {\rho_i}^{(Q)}, {\sigma_i}^{(P)}, {\sigma_i}^{(Q)})$ with σ_i being the standard deviation of e_i, to minimize the gap between the NUS-CRI PD term structure and the one deduced from the model.

Impact on default-rate distributions

Figures 1 and 2 show the impact of incorporating DCs into default-rate distributions. Each of these graphs show default-rate distributions with and without DCs for one of three regions under one of two future horizons, namely global, United States, and China over 3 and 12 months in Mar-2022. Evident through these graphs, incorporation of DCs tends to make the default-rate distribution much more right-skewed, reflecting a general tendency for firms in the same economy/sector to co-move as far as credit risk is concerned. The long right tail in these graphs suggests that ignoring DCs can cause a severe under-assessment of the chance for joint defaults to occur, a real concern for any credit portfolio. The conclusion applies to either short or longer horizon.

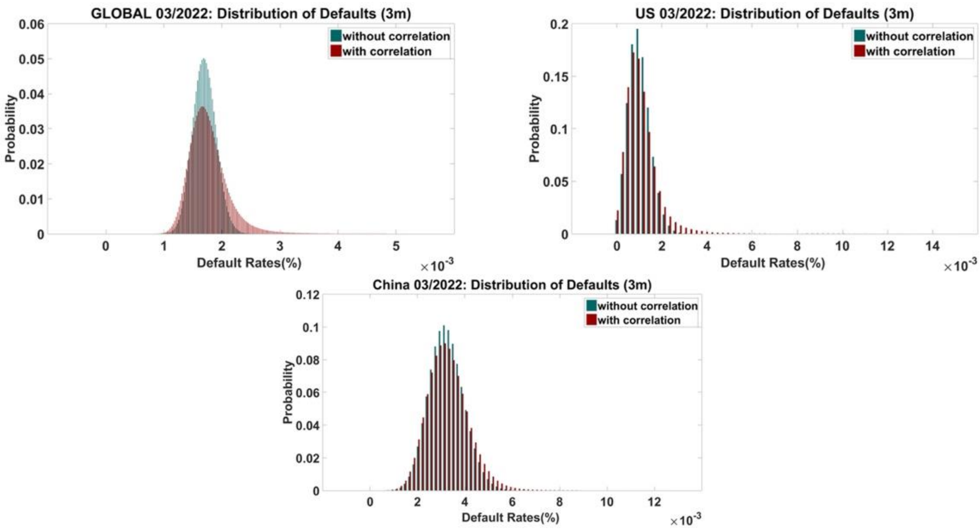

Figure 1: Correlated and non-correlated default-Rate distribution for the global economy, US, and China (3-month horizon)

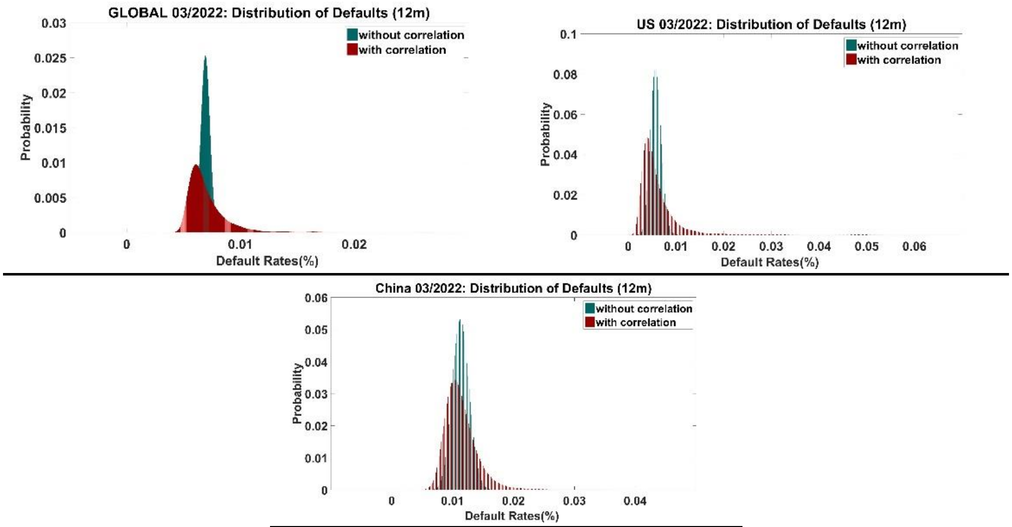

Figure 2: Correlated and non-correlated default-rate distribution for the global economy, US, and China (12-month horizon)