MODEL VALIDATION SAMPLE DOCUMENT

Contents

Abstract

This document summarizes the methodology used in Yang and Du [2016] for CCAR stress testing and provides multiple validation tests. For completeness, some complementary material from Yang and Du [2015], Rutkowski and Tarca [2015] are incorporated herein as well.

1 Background

Under Basel II, authorized deposit taking institutions are required to hold adequate capital for their credit risks. This can be done using either the standardized approach or the internal rating based (*IRB*) approach. Despite the fact that IRB involves administrative complications, it often leads to less capital requirements. IRB implements the asymptotic single factor (*ASRF*) model under which every single credit exposure accounts for only an infinitesimal amount to the total portfolio exposure. In addition, ASRF is an asset value model which means that a firm's assets value solely determine if it is going to default. Modeling the asset value via a geometric Brownian motion, we have:

$$\log A(t) = \log A(0) + \sqrt{t} \cdot \left(\sqrt{t} \cdot \left(\mu - \frac{\sigma^2}{2} \right) + \sigma W \right) \text{ where } W \sim N(0,1).$$

Fixing a risk measurement horizon $[0, \tau]$, we are able to formally define the (unconditional) default event:

$$D := \{W < \Phi^{-1}(p_\tau)\}$$

for some $p_\tau \in [0,1]$. We either assume that p_τ can be sourced from market data or it could be estimated based on the historical default frequencies. Note that $p_\tau = \mathbb{P}(D)$.

Under IRB, p_τ depends on the current rating of the entity (obligor) under consideration. More precisely, consider the following available ratings:

$$R_1, \cdots, R_K \text{ where } R_1 \text{ is the best and } R_K \text{ the worst } i.e., \text{ default rate.} \tag{1}$$

p_τ is defined depending on the entity's rating at the beginning of the horizon $[0, \tau]$. Dropping τ for notational convenience, p_i denotes the probability of default where entity has rating R_i. We also make the following assumption:

$$p_{i,j} \text{ is a decreasing function of } |i - j|. \tag{2}$$

Namely, transition to closer rating is considered more likely. Now consider a portfolio compromising of n entities where their assets values are modeled via $W_i \sim N(0,1)$ for $i \in [1, n]$. We assume that W_i are conditionally independent. More precisely, denoting the systematic risk factor by the random variable s, we suppose that $W_i|_s$ are independent and furthermore both s and $W_i|_s$ follow a mean-zero Gaussian distribution. We simplify somewhat further and write

$$W_i = \rho_i \cdot s + \sqrt{1 - \rho_i} \cdot \epsilon_i \text{ where } \rho_i \in (0,1) \text{ and } s, \epsilon_i \sim N(0,1).$$

Parameters ρ_i are called the asset correlation and assumed to be fixed across any rating level. Notice that

$$Cov(W_i, W_i) = \sqrt{\rho_i \rho_j} \tag{3}$$

Yang and Du [2016, 2015] estimates parameters ρ_i as a key step in the modeling process.

2 Methodology

In this section, we explain the methodology used in Yang and Du [2016] (Henceforth, referred to as YD16 for brevity).

2.1 Advanced Internal Rating Based Approach (AIRB)

YD16 considers a more granularized version of IRB. Under AIRB, instead of only the default transition probabilities, transition between different ratings are also considered. In detail, YD16 denotes by p_{ij} the probability transition from rating i, at the start of a horizon, to rating j at the end of the horizon. Notice that $p_{iK} = p_i$ by definition where p_i is defined in the paragraph preceding Equation (2). Moreover, similarly as above, there exists constant values b_{ij} for each fixed horizon so that for entity A_t with rating R_i at the beginning of the horizon

$$W_t < b_{i,K-j+1} \text{ implies transition from } R_i \text{ to } R_j \text{ at the end of the horizon.}$$

It may be defined that $b_{i,K} := +\infty$.

2.2 Macroeconomic Variables

YD16 decomposes the systematic risk factor s into two components: the first component is expressed in terms of macroeconomic factors and the second component is simply composed of a Gaussian noise. In detail, considering $(x_1, \cdots, x_m)$ to be a list of macroeconomic factors, YD16 writes

$$s = -\lambda \cdot ci(\boldsymbol{x}) - \epsilon \cdot \sqrt{1-\lambda^2}. \tag{4}$$

where $ci(\boldsymbol{x})$[1], called the credit index is the normalized version of $c_{a_1,\cdots,a_m}(\boldsymbol{x}) := \sum_{\ell=1}^{m} a_\ell x_\ell$ *i.e.*,

$$\mathbb{E}\left[ci(\boldsymbol{x})\right] = 0 \text{ and } \operatorname{Var}\left(ci(\boldsymbol{x})\right) = 1.$$

Notice that for

$$\tilde{a}_\ell := \frac{a_\ell}{\sqrt{\operatorname{Var}[c_{a_1,\cdots,a_m}(\boldsymbol{x})]}} \text{ for } \ell \in [1, m], \tag{5}$$

it holds that $ci(\boldsymbol{x}) = \sum_{\ell=1}^{m} \tilde{a}_\ell \tilde{x}_\ell$.

2.3 Parameter Estimation

The following equation could be easily obtained using some standard calculus of normally distributed variables (henceforth denoted by SCONDV).

$$p_{i,j}(s) = \Phi\left[\frac{b_{i,K-j+1} - s\sqrt{\rho_i}}{\sqrt{1-\rho_i}}\right] - \Phi\left[\frac{b_{i,K-j} - s\sqrt{\rho_i}}{\sqrt{1-\rho_i}}\right] \quad \text{{\scriptsize use (4) and definition of transition}} \tag{6}$$

Therefore, denoting by $d_{i,j}(s)$ the probability of transition from R_i to R_t for some $t \geq j$, given the systematic risk factor s, we have that

$$d_{i,j}(s) = \Phi\left[\frac{b_{i,K-j+1} - s\sqrt{\rho_i}}{\sqrt{1-\rho_i}}\right].$$

Step 1: Estimation of $b_{i,j}$. Using SCONDV, we obtain that, for $j > 1$,

$$\Phi^{-1}\left(b_{i,K-j+1}\right) = \mathbb{E}_s[d_{i,j}(s)] \approx \frac{n_{i,j}}{n_i} \tag{7}$$

[1] bold letters indicate multi-dimensional vectors

Here $n_{i,t}$ denotes the proportion of transitions from R_i to R_j at the beginning of the horizon. Also, $n_i := n_{i,1} + \cdots + n_{i,K}$. Approximation in (7) is obtained via a simple MLE argument. Notice that if $j = 1$, then $d_{i,j}(s) = 1$ by definition. Finally, $n_{i,j}$ are provided by market data.

Step 2: Estimation of ρ_i. In Yang and Du [2015], Section 4.1, through maximizing the probability of exactly c downgrades, they obtain an estimation for the parameter

$$r_i := \sqrt{\frac{\rho_i}{1-\rho_i}} \tag{8}$$

So far, estimations of values $b_{i,j}$ and ρ_i (via (8)) are obtained.

Step 3: Estimation of $\tilde{a}_\ell$ (5) using MLE. Appealing to SCONV, the following is true:

$$d_{i,K}(\boldsymbol{x}) := \mathbb{E}_s\left[d_{i,K}(s)|\boldsymbol{x}\right] = \Phi\left(b_{i,1}\sqrt{1+\tilde{r}_i^2} + \tilde{r}_i \cdot \sum_{\ell=1}^{m} \tilde{a}_\ell \tilde{x}_\ell\right) \tag{9}$$

where for λ from (4), it is defined

$$\tilde{r}_i(\lambda) := \frac{r_i \cdot \lambda}{\sqrt{1 + r_i \cdot (1-\lambda^2)}}. \tag{10}$$

YD16 takes 2nd derivatives to show that (9) is concave in $\tilde{a}_\ell$ for $\ell \in [1, m]$. Next, it considers $\lambda_0 < \lambda_1 < \cdots < \lambda_M$ with $\lambda_{u+1} - \lambda_u$ small enough. Via MLE based on three different likelihood functions Eq (3.1, 3.2, 3.3)[2], YD16 obtains global maximum likelihood estimations for $\tilde{a}_\ell$ for $\ell \in [1, m]$ for each λ_u for $u \in [1, M]$. Comparing M different values, YD16 finds the global maximum likelihood estimation for λ and $\tilde{a}_\ell$ for $\ell \in [1, m]$. For future reference, assume that max occurs at

$$\lambda_{u^*}, \tilde{a}_1, \cdots, \tilde{a}_m. \tag{11}$$

Step 4: Estimation of $\tilde{r}_i$: After Step 3 above, YD16 calculates $ci(\boldsymbol{x})$. Then using (9) and the same likelihood functions, namely Eqs. (3.1, 3.2, 3.3), it finds approximation of $\tilde{r}_i$.

Before, we finish this section, we emphasize that YD16 utilizes non-linear methods from SAS library to solve some of these MLE optimization problems.

2.4 Experiments

2.4.1 Data and Model Pre-Setup

The data for experiments comes from the historical quarterly rating transition frequency for a US commercial portfolio. There are seven distinct ratings $R_1, \cdots, R_7$. YD16 considers a list of nine different macro economic factors[3]. For each economic variable and for any given time, YD16 produces four *horizon-specific* macroeconomic factors. Namely, if $x_i(\tau_0)$ denotes the value of x_i at the beginning of horizon τ_0 and consequent horizons are separated by $\tau > 0$, then for τ_0 horizon, it considers $x_i(\tau_0), x_i(\tau_0 - \tau), x_i(\tau_0 - 2\tau)$ and $x_i(\tau_0 - 3\tau)$. Hence, in total, there are $36 = 9 \times 4$, horizon-specific macroeconomic variables at any specific horizon. Finally, due to computational complexity, YD16 only consider three[4] macroeconomic variables, at a time, for each model. This way, $\binom{9\times 4}{3} \approx 7000$ different models are tried out!

[2] See Section 3 for a discussion on these likelihood functions and their usages

[3] Each variable should pass the unit test to ensure non-stationarity

[4] In the paper, they write four, but that seems to be a typo!

2.4.2 Variable Selection

Putting together all the estimated parameters from the previous section, YD16 will obtain estimations for

$$p_{i,j}(\boldsymbol{x}) := \mathbb{E}_s\left[p_{i,j}(s)|\boldsymbol{x}\right] \text{ where } p_{i,j}(s) \text{ is calculated in (6)} \tag{12}$$

for any given $\boldsymbol{x} \in \mathbb{R}^3$ representing the macroeconomic variables. All the 7000 different models are trained using logistic regression.

2.4.3 Best Model Criterion

The best model is picked based on its performance during the 2008 financial crisis; namely, [2008Q1 - 2009Q4]. It is emphasized that for multi-horizons data samples, the likelihood functions for each period is calculated separately. The final likelihood function is the sum of all these functions.

2.4.4 Model Performance and Scenario Loss Projection

Based on the model they picked in Subsection 2.4.3, they compare the model's predicted default rates for each rating, $R_1, \cdots, R_7$, separately. They measure the goodness of their method using Mean Absolution Deviation from the historical default rates. Next, they compute their model's loss generation for the given scenarios provided by Fed. Scenarios are expressed in terms of the macroeconomic variables and are partitioned into three different categories: based, adverse and servery adverse. Notice that the loss function is defined as below:

$$L_n := \sum_{q=1}^{n} w_q \eta_q \mathbf{1}_{D_q}$$

Here w_q is the exposure weight of asset q. Note that $\sum_{q=1}^{n} w_q = 1$. Moreover, η_q denotes the loss given default at the horizon. Notice that

$$\mathbb{E}[L_n] = \sum_{q=1}^{n} w_q \eta_q p_{i_q,K} \text{ where asset } q \text{ belongs to } R_{i_q}.$$

Rutkowski and Tarca [2015] explains the credit risk modeling mathematical foundations.

3 Model Validation

In this section, we raise some questions regarding the methodology introduced in Yang and Du [2016]. Furthermore, we design some validation steps to test the correctness of their method.

3.1 Weighted Loss Function

YD16 picks their best model based on the performance over a specific horizon, namely the period between 2018Q1 and 2009Q4. See Subsection 2.4.3. This raises the following question:

Why not using a weighted loss function for training to assign higher weights to the stress periods?

A model that performs well on [2018Q1 - 2009Q4] does not yield similar performance for other periods. In view of this, we propose the following validation criterion:

Validation Test: Replace the logistic loss function used for training in Section 4 Yang and Du [2016] with a weighted version where higher weights are assigned to the horizon [2018Q1 - 2009Q4] while keeping every other model's parameter the same. We then compare the resulting top 10 models with Page 10, Table 2. We repeat the same experiments with varying weights distribution. Will we get similar top 10 models? How do their perform in terms of their MAD and RSQ values?

3.2 Asset Correlation

Even though, conditional independence between asset values may seem to be a common assumption, we would like to put its soundness to test. Notice that a fixed parameter ρ_t is considered for the entire set of obligors with rating R_t. We propose the following validation criterion.

Validation Test: Suppose that for each $t \in [1,7]$, there exist L_t asset of rating R_t. We select all pairs of available assets from rating R_t for $t = 1, \cdots, 7$. Hence, there will be $\left(\sum_{t=1}^{7} L_t\right)^2$ of such pairs. For a pair q_1 and q_2, we note that

$$\operatorname{Cov}\left(\log A_{q_1}(t) \cdot \log A_{q_2}(t)\right) = \sigma_{q_1}\sigma_{q_2} \cdot t \cdot \operatorname{Cov}(W_{q_1} W_{q_2}) \underset{\text{use } (3)}{=} \sigma_{q_1}\sigma_{q_2} \cdot t \cdot \sqrt{\rho_{q_1}\rho_{q_2}} \tag{13}$$

We now need to confirm that (13) (or some approximation thereof) holds. To this end, we regress the left-hand-side of $\log A_{q_1}(t) \cdot \log A_{q_2}(t)$ to obtain the best linear fit. Then using historical volatilities for assets q_1 and q_2, we obtain that

$$\operatorname{Cov}\left(\log A_{q_1}(t) \cdot \log A_{q_2}(t)\right) \approx \sigma_{q_1}\sigma_{q_2} \cdot t \cdot \sqrt{\rho(q_1, q_2)}.$$

Next, for each pair $i, j \in [1,7]$, we consider the following parameters:

$$\rho(q_1, q_2) \text{ where } q_1, q_2 \text{ have rating } R_i \text{ and } R_j \text{ respectively.} \tag{14}$$

We have gathered enough information to provide the validation criterion's statement concretely:

Will $\rho(q_1, q_2)$ as described in (14) exhibits low standard deviation for each pair of $i, j \in [1,7]$.

3.3 Merged Rating Levels

Why authors of Yang and Du [2016] decided to consider the 7 rating levels? For example, due to lack of default rate among the higher ratings (*e.g.,* R_1 and R_2), it may make sense to merge some of these ratings together and perform a similar experiments. On Page 8, Sec 3.1, they mention low default rates enforce a different choice for the likelihood function. However, merging ratings together could have been another solution as well. We propose the following validation criterion:

Validation Test: Keeping everything else the same, will merging the rating levels to reduce the considered ratings in the modeling prove useful in terms of performance? Merging could be done in various forms such as combining the top 2 ratings, last 2,3 ratings, combining some ratings in the middle, etc.

3.4 Risk Sensitivity

The last 6 columns of Table 2 describes non-consistent risk sensitivity w.r.t the model performance. For example, R_1 jumps for a very low risk of 0.02 to 0.23 to the next best model and the MAD and RSQ remains the same. We propose the following validation criterion.

Validation Test: Impose additional constraints to keep the risk sensitivies lower for better ratings. For example, in best model of Table 2, r_3 is twice r_4. Constraints such as the following should be added to the optimization problems during model fitting:

$$r_1 < (1-\kappa) \cdot r_2 < \cdots < (1-\kappa)^6 \cdot r_7$$

for $\kappa > 0$. Some few values of κ is recommended. We suggest $\kappa = 0, 0.001, 0.01, 0.1$. The resulting models' performances should be compared with Table 2.

3.5 Monotonicity of Transition Probabilities

Authors Yang and Du [2016] suggest that they impose the monotonicity of p_i values (2) in their optimization solving. They also miss to include $p_{i,i}$ in the chain of inequalities at the bottom of Page 7. We would like to see more clearly how this condition is imposed and whether the final transition probabilities satisfy these constraints.

3.6 Sensitivity to Market Data

Yang and Du [2016] deploys (7) to obtain estimation of $b_{i,j}$. We know that Φ decays extremely fast. In detail, we have that

$$\Phi'(x) = \frac{1}{\sqrt{2\pi}} \cdot \exp(\frac{-x^2}{2}).$$

Hence, even a substantial deviation in the estimation of $\frac{n_{i,j}}{n_i}$ leads to an infinitesimal deviation in the approximation of $b_{i,j}$. In view of this observation and to test the robustness of their technique, we propose the following validation criterion.

Validation Test: Replace each $n_{i,j}$ by $n_{i,j} \cdot (1 \pm \alpha_{i,j})$ where $\alpha_{i,j}$ is chosen uniformly random from $[0, \epsilon]$ for some $\epsilon > 0$. Will the resulting models perform similarly for various choices of ϵ? Notice that $n_{i,j}$ are also involved in the estimation of risk sensitivity parameters ρ_i.

3.7 λ Confusion

Authors Yang and Du [2016] appeals to the fact that since the likelihood function is concave in $\tilde{a}_1, \cdots, \tilde{a}_m$, global maximum could be found after trying enough many different values of λ.

After finding $\tilde{a}_1, \cdots, \tilde{a}_m$, they run another optimization problem to find $\tilde{r}_1, \cdots, \tilde{r}_m$. This is done despite the fact that $\tilde{r}_i$ is determined based on ρ_i and λ. See Eq. (10). We propose the following validation criterion.

Validation Test: In Section 2.3 (Parameter Estimation), drop Step 4 and use λ_u^* in (11) to estimate $\tilde{r}_i$. Simply plug λ_u^* into (10). Compare the resulting models with Table 2.

3.8 Lack of Baseline Methods

Authors Yang and Du [2016] did not compare their model's performance with some other obvious baseline methods. For example, on Page 12, Figuer 1, YD16 illustrates the performance of their model for transition probabilities $p_{i,7}$ for $i = 1, \cdots, 7$. On the other hand, an IRB also would provide estimations for the same default rates probabilities.

Validation Test: Does the AIRB approach of Yang and Du [2016] outperform a similar IRB approach? Furthermore, Monte Carlo simulation could be used for estimation of transition probabilities $p_{i,j}(\boldsymbol{x})$ (12). The Gaussian parameter ϵ in (4) can be generated synthetically to produce a MC simulation. Does incorporation of MC's techniques help with better performance?

3.9 Predictive Models

Authors Yang and Du [2016] have only considered logistic regressions for their model fitting. Other algorithms such as random forests proved to be quite successful in practice. We thus propose the following validation criterion.

Validation Test: Different standard methods such as random forests should be tried for model fitting and the resulting models need to be compared with Table 2. More advanced model such as LSTM are also recommended.

3.10 Computational Complexity

Authors Yang and Du [2016] limits their model to only three macroeconomic factors at a time due to excessive computational cost. We thus provide the following validation criterion.

Validation Test: Use some GPU programming, if possible, to expand the numerical results to more complex models! Limitation on only a few number of macroeconomic parameters does not seem ideal and counter intuitive. The main novelty of the method in Yang and Du [2016] is to use macroeconomic factors in modeling credit risk factors ...

References

Marek Rutkowski and Silvio Tarca. Regulatory capital modeling for credit risk. *International Journal of Theoretical and Applied Finance*, 18(05):1550034, 2015.

Bill Huajian Yang and Zunwei Du. Stress testing and modeling of rating migration under the vasicek model framework-empirical approaches and technical implementation. 2015.

Bill Huajian Yang and Zunwei Du. Rating transition probability models and ccar stress testing: Methodologies and implementations. 2016.