Likelihood Approaches to Low Default Portfolios

Adjustment of Alan Forrest's Method to the Multi-Year Period Design

Andrija Djurovic

www.linkedin.com/in/andrija-djurovic

Low Default Portfolio

- Qualitative descriptions of an Low Default Portfolio (LDP) leave ample room for interpretation, as different individuals may have varying opinions on whether a given portfolio qualifies as an LDP.
- The low number of defaults within a LDP undermines estimates' reliability and statistical validity for quantitative risk parameters based on historical default experience.
- LDPs are not necessarily low-data portfolios. The scarcity of defaults needs to be considered in relation to the size of the portfolio producing them.
- Regulators may be concerned that Probability of Default (PD) estimates based solely on simple historical averages or judgmental considerations may underestimate the bank's capital requirements due to default scarcity.
- Alan Forrest (AF) proposed Likelihood approaches for estimating the PD for the LDP. Compared to other methods, these approaches introduce an additional level of flexibility and utilize the data more efficiently.
- Despite the additional flexibility, comparing them with other methods is recommended.

Alan Forrest's Approach

- Compared to other approaches, AF proposed using the Likelihood and Likelihood Ratio instead of the probability of a class of data outcome. This change is crucial as it establishes a direct connection with the classical theory of statistical inference and its well-known approximations, which are valid for high default cases.
- The proposed method introduces additional flexibility and provides a framework for tailoring the approach to specific needs.
- In his paper, AF demonstrated the use of the likelihood approach for various cases, including single- and multiple-grade-level estimation with or without consideration of asset correlation and cases involving no or some defaults.
- Andrija Djurovic extended the proposed approach to one of the most prevalent designs in practice - portfolio-level multi-year period estimation, which considers asset and year-to-year correlation.
- The following slides describe the proposed adjustment to the multi-year design.

Adjustment of AF's Method to the Multi-Year Period Design

1. The change in the value of obligor i's assets over a year t is given by:

$$y_{i,t} = \sqrt{\rho} * z_t + \sqrt{1-\rho} * \epsilon_{i,t}$$

where ρ is asset correlation, z and ϵ_i systemic and idiosyncratic factor, respectively.

2. Given the number of years (T) and year-to-year correlation (θ) we simulate the systemic factor for each year as follows:

$$z_{t,t\leq T} = \theta * z_{t-1} + \sqrt{(1-\theta^2)} * \epsilon_t$$

where $z_{t,t=1}$ and ϵ_t are drawn from the standard normal distribution (`N[0, 1]`).

3. Using the $z_{t,t\leq T}$ we define the conditional PD_{c_t} for each year t, $t \leq T$ as follows:

$$PD_{c_t} = N\left[\frac{N^{-1}(PD) + z_t * \sqrt{\rho}}{\sqrt{1-\rho}}\right]$$

where PD comes from the range 0 and 1, and N and N^{-1} represent the distribution and quantile function of the normal standard distribution, respectively.

Adjustment of AF's Method to the Multi-Year Period Design cont.

4. For the multi-year period T, simulated $z_{t,t\leq T}$, and the PD that ranges from 0 to 1 we calculate the Likelihood as follows:

$$LL = \prod_{t}^{T} (PD_{c_t})^{d_t} * (1 - PD_{c_t})^{n_t - d_t}$$

   where d_t and n_t represent the number of defaults and obligors at year t, respectively.
5. To obtain the Expected Likelihood for the selected PD, we repeat steps from 2 to 4 N times and calculate the average value of the N simulated Likelihoods.
6. Repeat step 5 for the PD in the range from 0 to 1 to obtain the values of the Likelihood.
7. Given the Likelihoods from step 6, we calculate the Log-Likelihood Ratio as:

$$-2 * log(\frac{LL}{max(LL)})$$

   where LL is the Expected Likelihood for different PDs.
8. Finally, calculate the upper bound of the interval for which Log-Likelihood Ratio is lower than a cutpoint defined as:

$$cp = \begin{cases} -2 * log(1 - cl) & : \sum_{i=t}^{T} d = 0 \\ \chi^2_{(p=cl, df=1)} & : \sum_{i=t}^{T} d > 0 \end{cases}$$

   where cl is selected confidence level and $\chi^2_{(q=cl, df=1)}$ is quantile of the χ^2 distribtuon for the probability p equal to cl and 1 degree of freedom.

Simulation Dataset

Number of obligors in each grade per year:

```
##   GRADE Y1 Y2 Y3 Y4 Y5
## 1     A  8  7  6  4  1
## 2     B 24 25 25 24 24
## 3     C 37 37 36 37 35
## 4     D 23 24 25 26 25
## 5     E  5  6  4  4  5
## 6     F  1  3  3  2  5
## 7     G  1  1  2  2  3
```

Number of defaults in each grade during the year:

```
##   GRADE Y1 Y2 Y3 Y4 Y5
## 1     A  0  0  0  0  0
## 2     B  0  0  0  0  0
## 3     C  0  0  0  0  0
## 4     D  0  0  0  0  0
## 5     E  1  0  0  0  0
## 6     F  0  1  0  0  0
## 7     G  0  0  0  1  1
```

Summary

Number of obligors per year: 99, 103, 101, 99, 98

Number of defaults per year: 1, 1, 0, 1, 1

Default rate per year: 0.0101, 0.0097, 0, 0.0101, 0.0102

Average default rate (Y1-Y5): 0.00802

R Code Extract

```
#...
#inputs
n <- c(99, 103, 101, 99, 98)    #number of obligors
d <- c(1, 1, 0, 1, 1)           #number of defaults
theta <- 0.30                   #year-to-year correlation
rho <- 0.12                     #asset correlation
N <- 1e4                        #number of simulations
cl <- 0.75                      #confidence level
pd.r <- seq(from = 0.000,
            to = 0.03,
            by = 0.0001)        #pd range
#random seed
set.seed(975)
#pd interval simulation
sim.res <- pd.interval(pd.r = pd.r,
                       n = n,
                       d = d,
                       rho = rho,
                       theta = theta,
                       cl = cl,
                       N = N)
#pd of the maximum likelihood
sim.res[["pd.mll"]]
```

```
## [1] 0.0106
```

```
#upper bound of the pd interval
max(sim.res[["pd.i"]])
```

```
## [1] 0.0212
```

Python Code Extract

```
#...
#inputs
n = np.array([99, 103, 101, 99, 98])     #number of obligors
d = np.array([1, 1, 0, 1, 1])            #number of defaults
theta = 0.30                             #year-to-year correlation
rho = 0.12                               #asset correlation
N = int(1e4)                             #number of simulations
cl = 0.75                                #confidence level
pd_r = np.arange(start = 0,
                 stop = 0.030001,
                 step = 0.0001)          #pd range
#random seed
np.random.seed(678)
#pd interval simulation
sim_res = pd_interval(pd_r = pd_r,
                      n = n,
                      d = d,
                      rho = rho,
                      theta = theta,
                      cl = cl,
                      N = N)
#pd of the maximum likelihood
sim_res["pd_mll"]
```

```
## 0.0108
```

```
#upper bound of the pd interval
np.max(sim_res["pd_i"])
```

```
## 0.021400000000000002
```

Simulation Result Visualization