Data Science in Finance and Economics

DSFE, 1(3): 253–271.
DOI: 10.3934/DSFE.2021014
Received: 20 October 2021
Accepted: 15 November 2021
Published: 19 November 2021

http://www.aimspress.com/journal/dsfe

Research article

Estimating credit default probabilities using stochastic optimisation

Dominic Joseph*

Independent researcher, Kerala, India

* **Correspondence:** Email: dom.jo@cet.ac.in.

Appendix A

Table 1. Performance test results for S&P data—Binomial Test.

			90% CI Results			95% CI Results		
Method	Rating	PD Estimate	LB	UB	Binom Test	LB	UB	Binom Test
Pluto Tasche	1	2.22%	0.00%	7.14%	TRUE	0.00%	14.29%	TRUE
Pluto Tasche	2	2.22%	0.65%	4.58%	FALSE	0.00%	4.58%	TRUE
Pluto Tasche	3	2.28%	1.50%	3.10%	FALSE	1.39%	3.32%	FALSE
Pluto Tasche	4	2.72%	2.09%	3.36%	FALSE	1.98%	3.47%	FALSE
Pluto Tasche	5	4.33%	3.47%	5.24%	TRUE	3.33%	5.37%	TRUE
Pluto Tasche	6	4.43%	3.51%	5.39%	FALSE	3.35%	5.63%	FALSE
Pluto Tasche	7	12.22%	9.42%	15.20%	TRUE	8.81%	15.81%	TRUE
Pluto Tasche	8	9.02%	0.00%	17.24%	TRUE	0.00%	20.69%	TRUE
CAP Calibration	1	0.17%	0.00%	0.00%	TRUE	0.00%	0.00%	TRUE
CAP Calibration	2	0.19%	0.00%	0.65%	TRUE	0.00%	1.31%	TRUE
CAP Calibration	3	0.32%	0.00%	0.64%	TRUE	0.00%	0.75%	TRUE
CAP Calibration	4	0.76%	0.44%	1.10%	FALSE	0.39%	1.21%	FALSE
CAP Calibration	5	2.23%	1.63%	2.86%	FALSE	1.50%	2.99%	FALSE
CAP Calibration	6	5.47%	4.41%	6.53%	FALSE	4.24%	6.78%	FALSE
CAP Calibration	7	7.79%	5.47%	10.33%	FALSE	5.17%	10.94%	FALSE
CAP Calibration	8	8.23%	0.00%	17.24%	TRUE	0.00%	20.69%	TRUE
QMM	1	0.03%	0.00%	0.00%	TRUE	0.00%	0.00%	TRUE

Continued on next page

			90% CI Results			95% CI Results		
Method	Rating	PD Estimate	LB	UB	Binom Test	LB	UB	Binom Test
QMM	2	0.09%	0.00%	0.65%	TRUE	0.00%	0.65%	TRUE
QMM	3	0.27%	0.00%	0.54%	TRUE	0.00%	0.64%	TRUE
QMM	4	0.71%	0.39%	1.05%	FALSE	0.33%	1.10%	FALSE
QMM	5	1.63%	1.09%	2.18%	FALSE	1.02%	2.31%	FALSE
QMM	6	3.68%	2.86%	4.57%	FALSE	2.69%	4.73%	FALSE
QMM	7	9.82%	7.29%	12.46%	TRUE	6.69%	13.07%	TRUE
QMM	8	24.89%	13.79%	37.93%	FALSE	10.34%	41.38%	FALSE
SA ($\alpha = 5\%$)	1	0.00%	0.00%	0.00%	TRUE	0.00%	0.00%	TRUE
SA ($\alpha = 5\%$)	2	0.00%	0.00%	0.00%	TRUE	0.00%	0.00%	TRUE
SA ($\alpha = 5\%$)	3	0.01%	0.00%	0.11%	TRUE	0.00%	0.11%	TRUE
SA ($\alpha = 5\%$)	4	0.02%	0.00%	0.11%	TRUE	0.00%	0.11%	TRUE
SA ($\alpha = 5\%$)	5	3.07%	2.38%	3.81%	FALSE	2.24%	4.01%	FALSE
SA ($\alpha = 5\%$)	6	3.07%	2.29%	3.92%	FALSE	2.12%	4.08%	FALSE
SA ($\alpha = 5\%$)	7	10.87%	8.21%	13.68%	TRUE	7.60%	14.29%	TRUE
SA ($\alpha = 5\%$)	8	11.07%	3.45%	20.69%	TRUE	0.00%	24.14%	TRUE
SA ($\alpha = 2.5\%$)	1	0.00%	0.00%	0.00%	TRUE	0.00%	0.00%	TRUE
SA ($\alpha = 2.5\%$)	2	0.00%	0.00%	0.00%	TRUE	0.00%	0.00%	TRUE
SA ($\alpha = 2.5\%$)	3	0.00%	0.00%	0.00%	TRUE	0.00%	0.00%	TRUE
SA ($\alpha = 2.5\%$)	4	0.01%	0.00%	0.06%	TRUE	0.00%	0.06%	TRUE
SA ($\alpha = 2.5\%$)	5	3.06%	2.31%	3.81%	FALSE	2.18%	3.95%	FALSE
SA ($\alpha = 2.5\%$)	6	3.07%	2.29%	3.92%	FALSE	2.12%	4.08%	FALSE
SA ($\alpha = 2.5\%$)	7	11.00%	8.21%	13.98%	TRUE	7.60%	14.59%	TRUE
SA ($\alpha = 2.5\%$)	8	11.04%	3.45%	20.69%	TRUE	0.00%	24.14%	TRUE

Table 2. Performance test results for S&P data—Likelihood Ratio Test.

Method	Rating	PD Estimate	Lik Ratio	Lik Ratio Test (5% significance)	Lik Ratio Test (2.5% significance)
Pluto Tasche	1	2.22%	0.73045	TRUE	TRUE
Pluto Tasche	2	2.22%	0.03205	FALSE	FALSE
Pluto Tasche	3	2.28%	0.00000	FALSE	FALSE
Pluto Tasche	4	2.72%	0.00000	FALSE	FALSE
Pluto Tasche	5	4.33%	0.89532	TRUE	TRUE
Pluto Tasche	6	4.43%	0.00004	FALSE	FALSE
Pluto Tasche	7	12.22%	0.93296	TRUE	TRUE
Pluto Tasche	8	9.02%	0.49545	TRUE	TRUE
CAP Calibration	1	0.17%	0.97643	TRUE	TRUE
CAP Calibration	2	0.19%	0.74605	TRUE	TRUE
CAP Calibration	3	0.32%	0.05129	FALSE	FALSE
CAP Calibration	4	0.76%	0.00000	FALSE	FALSE
CAP Calibration	5	2.23%	0.00009	FALSE	FALSE
CAP Calibration	6	5.47%	0.00000	FALSE	FALSE
CAP Calibration	7	7.79%	0.05781	FALSE	FALSE
CAP Calibration	8	8.23%	0.57600	TRUE	TRUE
QMM	1	0.03%	0.99564	TRUE	TRUE
QMM	2	0.09%	0.87311	TRUE	TRUE
QMM	3	0.27%	0.08366	FALSE	TRUE
QMM	4	0.71%	0.00000	FALSE	FALSE
QMM	5	1.63%	0.00000	FALSE	FALSE
QMM	6	3.68%	0.00398	FALSE	FALSE
QMM	7	9.82%	0.58859	TRUE	TRUE
QMM	8	24.89%	0.00639	FALSE	FALSE
SA ($\alpha = 5\%$)	1	0.00%	1.00000	TRUE	TRUE
SA ($\alpha = 5\%$)	2	0.00%	0.99780	TRUE	TRUE
SA ($\alpha = 5\%$)	3	0.01%	0.92689	TRUE	TRUE
SA ($\alpha = 5\%$)	4	0.02%	0.68103	TRUE	TRUE
SA ($\alpha = 5\%$)	5	3.07%	0.10244	FALSE	TRUE
SA ($\alpha = 5\%$)	6	3.07%	0.08341	FALSE	TRUE
SA ($\alpha = 5\%$)	7	10.87%	0.92648	TRUE	TRUE
SA ($\alpha = 5\%$)	8	11.07%	0.32083	TRUE	TRUE
SA (2.5%)	1	0.00%	1.00000	TRUE	TRUE
SA (2.5%)	2	0.00%	0.99944	TRUE	TRUE
SA (2.5%)	3	0.00%	0.98261	TRUE	TRUE
SA (2.5%)	4	0.01%	0.86061	TRUE	TRUE
SA (2.5%)	5	3.06%	0.09497	FALSE	TRUE
SA (2.5%)	6	3.07%	0.08481	FALSE	TRUE
SA (2.5%)	7	11.00%	0.95112	TRUE	TRUE
SA (2.5%)	8	11.04%	0.32343	TRUE	TRUE

Table 3. Performance test results for simulated portfolio data—Binomial Test.

			90% CI Results			95% CI Results		
Method	Rating	PD Estimate	LB	UB	Binom Test	LB	UB	Binom Test
Pluto Tasche	1	3.17%	0.00%	0.00%	TRUE	0.00%	100.00%	TRUE
Pluto Tasche	2	3.20%	0.00%	16.67%	TRUE	0.00%	16.67%	TRUE
Pluto Tasche	3	3.41%	0.00%	10.00%	TRUE	0.00%	15.00%	TRUE
Pluto Tasche	4	3.56%	0.00%	9.52%	TRUE	0.00%	14.29%	TRUE
Pluto Tasche	5	5.06%	0.00%	12.50%	TRUE	0.00%	18.75%	TRUE
Pluto Tasche	6	3.94%	0.00%	18.18%	TRUE	0.00%	18.18%	TRUE
Pluto Tasche	7	3.03%	0.00%	10.00%	TRUE	0.00%	10.00%	TRUE
Pluto Tasche	8	24.80%	0.00%	100.00%	TRUE	0.00%	100.00%	TRUE
CAP Calibration	1	0.36%	0.00%	0.00%	TRUE	0.00%	0.00%	TRUE
CAP Calibration	2	0.40%	0.00%	0.00%	TRUE	0.00%	0.00%	TRUE
CAP Calibration	3	0.62%	0.00%	5.00%	TRUE	0.00%	5.00%	TRUE
CAP Calibration	4	1.33%	0.00%	4.76%	TRUE	0.00%	9.52%	TRUE
CAP Calibration	5	3.38%	0.00%	12.50%	TRUE	0.00%	12.50%	TRUE
CAP Calibration	6	7.37%	0.00%	18.18%	TRUE	0.00%	27.27%	TRUE
CAP Calibration	7	10.03%	0.00%	20.00%	TRUE	0.00%	25.00%	TRUE
CAP Calibration	8	10.52%	0.00%	50.00%	TRUE	0.00%	50.00%	TRUE
QMM	1	5.23%	0.00%	100.00%	TRUE	0.00%	100.00%	TRUE
QMM	2	4.84%	0.00%	16.67%	TRUE	0.00%	33.33%	TRUE
QMM	3	4.49%	0.00%	15.00%	TRUE	0.00%	15.00%	TRUE
QMM	4	4.22%	0.00%	14.29%	TRUE	0.00%	14.29%	TRUE
QMM	5	4.03%	0.00%	12.50%	TRUE	0.00%	12.50%	TRUE
QMM	6	3.89%	0.00%	18.18%	TRUE	0.00%	18.18%	TRUE
QMM	7	3.68%	0.00%	10.00%	TRUE	0.00%	15.00%	TRUE
QMM	8	3.29%	0.00%	50.00%	TRUE	0.00%	50.00%	TRUE
SA ($\alpha = 5\%$)	1	0.00%	0.00%	0.00%	TRUE	0.00%	0.00%	TRUE
SA ($\alpha = 5\%$)	2	0.10%	0.00%	0.00%	TRUE	0.00%	0.00%	TRUE
SA ($\alpha = 5\%$)	3	2.67%	0.00%	10.00%	TRUE	0.00%	10.00%	TRUE
SA ($\alpha = 5\%$)	4	2.68%	0.00%	9.52%	TRUE	0.00%	9.52%	TRUE
SA ($\alpha = 5\%$)	5	6.16%	0.00%	18.75%	TRUE	0.00%	18.75%	TRUE
SA ($\alpha = 5\%$)	6	6.25%	0.00%	18.18%	TRUE	0.00%	27.27%	TRUE
SA ($\alpha = 5\%$)	7	6.27%	0.00%	15.00%	TRUE	0.00%	20.00%	TRUE
SA ($\alpha = 5\%$)	8	6.28%	0.00%	50.00%	TRUE	0.00%	50.00%	TRUE
SA ($\alpha = 2.5\%$)	1	0.00%	0.00%	0.00%	TRUE	0.00%	0.00%	TRUE
SA ($\alpha = 2.5\%$)	2	0.74%	0.00%	0.00%	TRUE	0.00%	16.67%	TRUE
SA ($\alpha = 2.5\%$)	3	2.74%	0.00%	10.00%	TRUE	0.00%	10.00%	TRUE
SA ($\alpha = 2.5\%$)	4	2.74%	0.00%	9.52%	TRUE	0.00%	9.52%	TRUE
SA ($\alpha = 2.5\%$)	5	6.09%	0.00%	18.75%	TRUE	0.00%	18.75%	TRUE
SA ($\alpha = 2.5\%$)	6	6.13%	0.00%	18.18%	TRUE	0.00%	27.27%	TRUE
SA ($\alpha = 2.5\%$)	7	6.17%	0.00%	15.00%	TRUE	0.00%	20.00%	TRUE
SA ($\alpha = 2.5\%$)	8	6.18%	0.00%	50.00%	TRUE	0.00%	50.00%	TRUE

Table 4. Performance test results for simulated portfolio data—Likelihood Ratio Test.

Method	Rating	PD Estimate	Lik Ratio	Lik Ratio Test (5% significance)	Lik Ratio Test (2.5% significance)
Pluto Tasche	1	3.17%	0.96830	TRUE	TRUE
Pluto Tasche	2	3.20%	0.82259	TRUE	TRUE
Pluto Tasche	3	3.41%	0.93507	TRUE	TRUE
Pluto Tasche	4	3.56%	0.46664	TRUE	TRUE
Pluto Tasche	5	5.06%	0.51369	TRUE	TRUE
Pluto Tasche	6	3.94%	0.75254	TRUE	TRUE
Pluto Tasche	7	3.03%	0.54065	TRUE	TRUE
Pluto Tasche	8	24.80%	0.56553	TRUE	TRUE
CAP Calibration	1	0.36%	0.99640	TRUE	TRUE
CAP Calibration	2	0.40%	0.97630	TRUE	TRUE
CAP Calibration	3	0.62%	0.29195	TRUE	TRUE
CAP Calibration	4	1.33%	0.75464	TRUE	TRUE
CAP Calibration	5	3.38%	0.29304	TRUE	TRUE
CAP Calibration	6	7.37%	0.97795	TRUE	TRUE
CAP Calibration	7	10.03%	0.12067	FALSE	TRUE
CAP Calibration	8	10.52%	0.80066	TRUE	TRUE
QMM	1	5.23%	0.94767	TRUE	TRUE
QMM	2	4.84%	0.74256	TRUE	TRUE
QMM	3	4.49%	0.99406	TRUE	TRUE
QMM	4	4.22%	0.40440	TRUE	TRUE
QMM	5	4.03%	0.37899	TRUE	TRUE
QMM	6	3.89%	0.74617	TRUE	TRUE
QMM	7	3.68%	0.47263	TRUE	TRUE
QMM	8	3.29%	0.93522	TRUE	TRUE
SA ($\alpha = 5\%$)	1	0.00%	1.00000	TRUE	TRUE
SA ($\alpha = 5\%$)	2	0.10%	0.99387	TRUE	TRUE
SA ($\alpha = 5\%$)	3	2.67%	0.84570	TRUE	TRUE
SA ($\alpha = 5\%$)	4	2.68%	0.56492	TRUE	TRUE
SA ($\alpha = 5\%$)	5	6.16%	0.64632	TRUE	TRUE
SA ($\alpha = 5\%$)	6	6.25%	0.93507	TRUE	TRUE
SA ($\alpha = 5\%$)	7	6.27%	0.27370	TRUE	TRUE
SA ($\alpha = 5\%$)	8	6.28%	0.87833	TRUE	TRUE
SA (2.5%)	1	0.00%	1.00000	TRUE	TRUE
SA (2.5%)	2	0.74%	0.95641	TRUE	TRUE
SA (2.5%)	3	2.74%	0.85661	TRUE	TRUE
SA (2.5%)	4	2.74%	0.55798	TRUE	TRUE
SA (2.5%)	5	6.09%	0.63866	TRUE	TRUE
SA (2.5%)	6	6.13%	0.92907	TRUE	TRUE
SA (2.5%)	7	6.17%	0.27979	TRUE	TRUE
SA (2.5%)	8	6.18%	0.88022	TRUE	TRUE

Appendix B

The following is the code snippet used as part of the research.

```
library(dplyr)
library(tidyr)

##### objective function
ls_fun = function(x,n,p,d = 0, e = 0){

  thetas       = x/n
  teststat     = 2*abs(log(dbinom(x,size = n,prob = thetas)/dbinom(x,size = n,prob
= p)))
  infer        = 1-pchisq(teststat,df = 1)<=0.025
  sum(log(dbinom(x,size = n,prob = p)))+sum(diff(p)<=0)*-d + sum(infer)*-e
}

##### simulated annealing
sim_anneal = function(fun, N, x, n, theta0, thetas, T0, Tf , var, d, e, f, alpha){

  llb    = c()
  thetab = c()
  k   = 0
  i   = 1
  dstart = d
  estart = e

  while(i<N){

    # calculate the new candidate
    theta = pmax(pmin(rnorm(length(x),mean = theta0, sd = sqrt(var)), 0.9999), 0)
    #decrease temperature
    Ti    = T0*(Tf/T0)^((i-1)/N)

    # if new point is far off, select previous point
    P1           = dbinom(x,size = n,prob = thetas)
    P2           = dbinom(x,size = n,prob = theta)
    teststat     = 2*abs(log(P1/P2))
    infer         = 1-pchisq(teststat,df = 1)<=alpha
    theta[infer] = theta0[infer]

    #calculate the negative objective function value for new and old points
    ll0 = -fun(x = x,n = n,p = theta0, d = d, e = e, f = f)
    lla = -fun(x = x,n = n,p = theta, d = d, e = e,  f = f)
    r   = exp(-(lla-ll0)/Ti)
```

```
    # accept with a uniform probability
    if(r >= runif(1,0,1) | lla<=ll0){
      theta0 = theta
      if(all(diff(theta)>=0)){
        thetab = rbind(thetab,theta)
        llb    = c(llb, lla)
        k      = ifelse( lla<=min(llb), k+1, k)
      }
    }
    i = i+1
  }

  if(!is.null(thetab)){
    TT           = data.frame(thetab)
    names(TT)    = 1:ncol(TT)
    rownames(TT) = NULL
    TT$Neg.Likelihood = llb
    TT           = dplyr::arrange(TT, Neg.Likelihood)
  }else{
    TT = NA
  }
  lls  = -fun(x = x,n = n,p = thetas, d = d, e= e)
  pars = list(lls = lls, TT = TT , k=k, dstart = dstart, estart = estart,
              dend = d, eend = e)
  return(pars)
}

##### perform 250 iterations
out = lapply(1:250, function(y){  z   = sim_anneal(fun = ls_fun,
                  N=5*10^4,
                  x = df$D,
                  n = df$Total,
                  theta0 = df$DR, #inital guess
                  thetas = df$DR, #MLE of theta
                  T0=100, #Starting temp
                  Tf = 10^-7, #Final temp
                  var = 0.00005,
                  d = 1000,
                  e = 900,
                  f = 0,
                  alpha = 0.025)
  outdat = z$TT
  if(class(outdat) == 'data.frame'){
```

```
    outdat$j = y
    outdat$iters = z$k}
  else{
    outdat = NULL
  }
  return(outdat)
})

#extract the output results
out.df  = do.call(rbind, out)
out.opt = dplyr::arrange(out.df, j, Neg.Likelihood) %>%
            dplyr::group_by(j) %>%slice(1)
out.opt = dplyr::arrange(out.opt, Neg.Likelihood)
```

Sample data:

Rating Grade	***Defaults***	***Total Customers***	***Non Defaults***	***Observed Default Rate***
Rating	D	Total	ND	DR
1	0	14	14	0.00%
2	0	153	153	0.00%
3	0	934	934	0.00%
4	0	1814	1814	0.00%
5	60	1470	1410	4.08%
6	25	1225	1200	2.04%
7	38	329	291	11.55%
8	1	29	28	3.45%

AIMS Press