Hedging the risk of productivity loss in financial operations

Ashish Pandey

Abstract

In Operations analytics, key parameter is how business operation runs in relation to the assigned performance targets. Key dimension here is performance rate which defines the throughput for a given unit of time. Some level of productivity loss is always accounted for during planning and thus risk here is negative deviance from the expected productivity loss. It's important for any organization to manage tradeoff between costs of deploying additional resources Vs having effective risk controls implemented. This becomes more of a cost benefit analysis and to strike balance between both becomes important. Paper leverages Loss distribution approach from operational risk space to model Risk of unexpected productivity loss. The approach results in a Full time equivalent contingency plan in form of human capital requirement for worst case scenario and provides an optimal resource plan to manage risk. If required resource pool is less than the optimal, it exposes business to the risk of increased Turnaround times, while if the same is above optimal level it adds to overall cost of the firm.

Modeling approach:

(Data is hypothecated for illustration purpose & tool used is R)

a) Lognormal distribution and EVT: Models magnitude of productivity loss
b) Poisson & Negative binomial distribution: Models frequency of productivity loss events
c) Convolution technique: combines frequency and loss magnitude distributions using Montecarlo simulation
d) Correlation among operational process is used to derive the business level contingency plan
e) Accounting for shrinkage in capacity utilization of human capital

Keywords: Operations, risk, productivity, loss, optimization, workforce, Convolution, Insurance

Key Messages:

1) In terms of resource planning in financial operations it is not just important to plan for the expected scenarios but also to be ready with a contingency plan for any unexpected scenario.

2) Unexpected productivity losses can be a major area of concern in financial operations like claim processing, NAV related processes like Surrenders and underwriting that may be exposed to legal and operational risk due to increased turnaround times.

3) The Paper proposes the use of Loss distribution approach to model productivity losses in financial operations. It provides management with an optimal resource plan to hedge the risk of productivity losses.

• *Ashish Pandey is a Certified FRM from GARP (Global association of Risk Professionals) with extensive experience into financial Services domain. E-mail: ashish.pandeysims@gmail.com*

1) Productivity Loss Measurement

Let $Mx1$=target process rate for an operations function x1. Process rate is output per unit of time, for example in claims processing function it will be no. of claims processed per hour. In Insurance domain Claim processing, surrenders and Underwriting are key operations functions where turnaround times are really low and any delay in processing can pose significant operational and legal risk.

$\mathcal{R}x1$=realized process rate for function x1.

$K1, K2 \ldots \ldots Kn$ Denote individual cases being processed in that function and $t1, t2 \ldots \ldots tn$ be the individual time measured in hours to process the cases.

Thus $\mathcal{R}x1 = \sum_{i=1}^{n} Ki / \sum_{i=1}^{n} ti$

$\sum_{i=1}^{n} Ki$ Is sum of cases being processed and $\sum_{i=1}^{n} ti$ is sum of time taken to process those cases.

Productivity loss measured in hours is then defined as

$$Ploss(x1) = \sum_{i=1}^{n} ti - \left(\frac{\sum_{i=1}^{n} Ki}{Mx1}\right) \ldots..for\ \mathcal{R}x1 < Mx1$$

$$Ploss(x1) = 0 \ldots for\ \mathcal{R}x1 \geq Mx1$$

Let $wt1, wt2 \ldots \ldots wtn$ be the weekly time intervals over which we measure frequency of $Ploss(x1)$ and average magnitude impact of those $Ploss(x1)$ in hours

Data table structure (rate)

Week	Frequency of $Ploss(x1)$	Average $Ploss(x1)$
1	30	7 hours
2	22	5 hours

Table: 1

The risk for the firm here is unexpected productivity losses that can happen due to high Absenteeism and Turnover, operational issues, system breakdown, sudden change in the mix of complexity of cases and many more. Since productivity loss here is defined in terms of the additional hours being used by the team when performance rate was below the guided target rate, the contingency plan for unexpected productivity losses is derived in terms of additional hours or resources that BU head should plan for such a situation if arises.

2. Modeling the average magnitude of Ploss(x1) distribution

2.1) Descriptive statistics of productivity loss distribution

```
> library(fitdistrplus)
> library(evir)
```

Name of the dataset is “rate”. Productivity loss variable is “loss”.

```
> fit_n=fitdistr(rate$loss,densfun ="normal")
```

```
> hist(rate$loss,xlim = c(5,14),breaks=20,col = 8,probability = TRUE,main="Productivity Loss Distribution")
```

```
>curve(dnorm(x,fit_n$estimate[1],fit_n$estimate[2]),col="red",lwd=2,add = T)
```

Statistical summary of Productivity loss distribution

Min	1st Qu.	Median	Mean	3rd Qu.	Max.
4.778	6.588	7.733	8.410	9.286	19.434

Table: 2

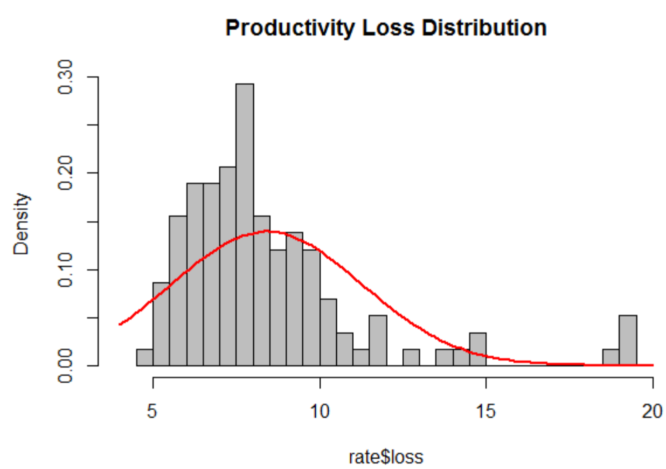


Fig: 1

We assume that there will be some level of variability in productivity and thus we will not consider initial 2% of productivity loss distribution for our analysis.

```
> Lower_Cutoff=qnorm(0.02,mean = fit_n$estimate[1],sd=fit_n$estimate[2],lower.tail=TRUE,log.p=FALSE)

> rate=subset(rate,rate$loss>Lower_Cutoff)
```

Next step in the process is to model magnitude of productivity loss distribution. We will separately model the body and tail of distribution.

Body of the distribution is modeled using Logarithmic distribution because sample data we have collected would be sufficient enough around the location of the log dist. to model the same, log distribution is bounded by "0" on lower end which makes it a good candidate for application in comparison to a hypothesized normal distribution. A normal distribution can take negative values while the same is not with the log distribution. Another valid reason for application of log distribution is that productivity loss values are skewed to the right with fat tails, which in general would be the case, be it with operational losses or financial losses.

When we are measuring risk of an event happening through probability distribution of a random variable we are majorly concerned with probability of occurrence of rare events and their magnitude impact. These events can occur, may be once in a year or a span of 5 years.

for example: sudden increase of claim levels or surrenders to the extreme end of distribution would be rare in nature which might happen due to increase in claims in an event of a natural calamity like hurricanes, surrenders can increase suddenly if the broad financial market crashes in a given day or a week.

Surrender initiation and approval process has to be completed before the market close cut-off time for all cases received before a defined interval (generally 3 PM). Organization here bears the risk of compensating customers for any loss of NAV due to delay in surrender processing. In a normal market scenario it may not be a big risk for an Insurance firm but in a bear market or a recession this can hurt the organization badly.

Underwriting is a risk management process where underwriters would study the individual cases to decide on level of sum assured or cover the life assured would be eligible for. This process is core to an insurance firm as it involves the process of strict financial underwriting which includes Anti-money laundering checks, KYC norms and financial health of the life assured. If there is sudden increase in the business of insurance firm in terms of gross premium collection it would add additional pressure on the Underwriting division (If not adequately planned for) to adhere to the regulatory timelines. Regulatory norms are defined around this process in terms of adherence to turn around times, for any case that breaches the same the firm is exposed to legal and regulatory risk. Insurance firm would be obligated to honor the claim if it arises during that period where the contract was not underwritten/issued due delay in Turnaround times.

Underwriting as a process is not mechanical but based on quantitative and qualitative judgement of experienced underwriters who manage financial risk for the organization. If the complexity of the cases increase ,for example: Volume proportion mix that used to be heavy on individual life cases for a underwriting unit suddenly became heavy on Keyman insurance where sum assured levels are high or sudden increase in cases where Anti money laundering checks would consume time. These situations would be rare in nature and the combined productivity losses in such a scenario can expose the organization to major legal, regulatory, operational and financial risk.

To model the risk of extreme events we will use Extreme value theory. EVT is a branch of statistics that deals specifically with rare events and allows making statistical inferences from cases where data points are scarce. There are two modeling techniques in the same 1) GEV (Generalized extreme value distribution) which is the block maxima approach and the other is 2) POT (Peak over threshold) where data points above a particular threshold are considered for modeling. POT converges to Generalized Pareto distribution as the threshold level increases commonly known as GPD distribution.

2.2) Extreme value theory (GPD): to model the tail of the distribution.

2.2.1) Distributions function for GPD

Z: Stochastic variable.

Function of distribution of Variable Z: $F(Z) = P(Z \leq z)$

$'H'$: Threshold level where the excess events are given by $Y = Z - H$ and have the following distribution

(1) $FH(y) = P(Z - H \leq y | Z > H) = (F(y + Z) - F(H))/(1 - F(H))$

Cumulative distribution function

(2) $G\varepsilon, \beta(y) = [1 - \left(1 + \frac{\varepsilon y}{\beta}\right)^{-\frac{1}{\varepsilon}}, \varepsilon \neq 0]$

(3) $G\varepsilon, \beta(y) = [1 - exp\left(-\frac{y}{\beta}\right), \varepsilon = 0]$

2.2.2) *Modeling the tail of the distribution:* Here we will use Extreme value distribution on the tail of $Ploss(x1)$. Gradient of the Mean excess plot gives the information about the nature of tail and possible threshold value.

> meplot(loss)

Mean excess plot gives the positive gradient indicating fat nature of the tail.
Possible threshold value =8

> meplot_cutoff=8

> gpdfit_dist=gpd(loss,threshold = meplot_cutoff)
> gpdfit_dist$par.ests
The above function provides parameter estimates for the fitted GPD distribution.
ε =0.1939016
β =2.1533231
> plot(gpdfit_dist) gives exponential quantiles chart and functional distribution of stochastic variable denoting productivity loss magnitude above the defined threshold value. Below charts (Fig: 2) show reasonable fit to data

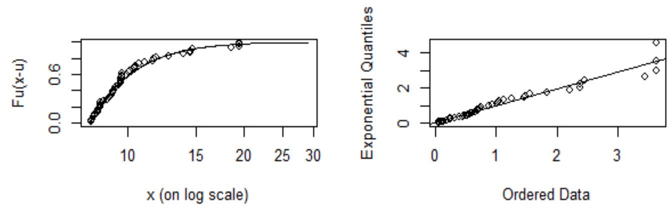


Fig: 2

2.3) Logarithmic distribution: to model the body: here we will fit logarithmic distribution to find estimates that are location and scale parameters.

$$\text{PDF}= \frac{1}{x\sigma\sqrt{2\pi}} e^{-\left(\frac{((\ln(x)-\mu)^2}{2\sigma^2}\right)}$$

First we will subset our data to select for productivity loss values below the threshold level. Fitdist allows modeling of logarithmic distribution using the maximum likelihood parameter estimation.

> loss1=subset(loss,loss<meplot_cutoff)>
loss2=data.frame(loss1)
> fit_ln_loss=fitdist(loss1,"lnorm")
> summary(fit_ln_loss)
Parameter estimates for meanlog and sdlog output is given below

Meanlog=1.8956249
Sdlog=0.1313317

Fig: 3 gives reasonable fit to the hypothesized logarithmic distribution.

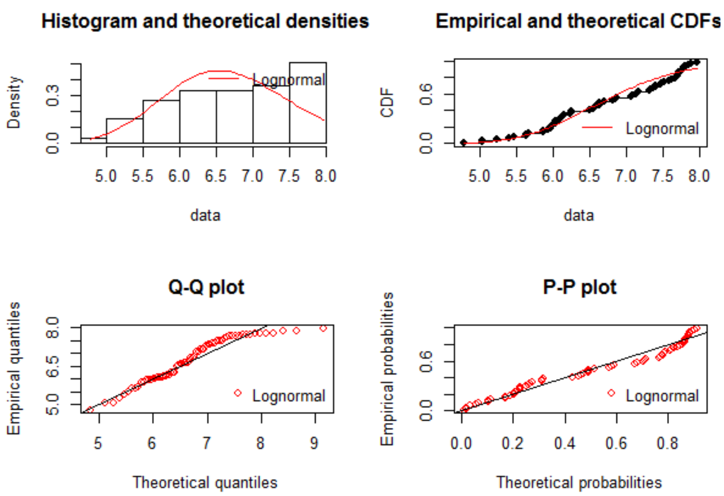


Fig: 3

>plnorm(7,meanlog=fit_ln_loss$estimate[1],sdlog=fit_ln_loss$estimate[2],lower.tail = TRUE,log.p=FALSE)

=64.9%

Above function gives cumulative probability distribution to the left of the tail for magnitude impact of productivity loss.

3) Modeling the Frequency of productivity loss events

Frequency of productivity loss events can be modeled using Poisson distribution and Negative binomial distribution. It gives probability of a given number of events occurring in a defined interval of time.

Poisson distribution is given by

$$\rho(x \text{ events in interval } t) = \lambda^x e^{-\lambda}/x!$$

Parameter estimate given by λ is the average number of events for a given time interval.

We will use the MASS, vcd and fitdistrplus libraries to analyze the Poisson and Negative-Binomial distribution of the Frequency of productivity loss distribution

```
> set.seed(1000)
> library(MASS)
> library(vcd)
> library(fitdistrplus)
```

```
> Poisson_estimate=fitdistr(freq,"poisson")
> Poisson_estimate
> lambda_pois=Poisson_estimate$estimate
[1] 36.06897

> sd_pois=Poisson_estimate$sd
```

Lambda value of 36.06897 is received basis fitdistr function with a standard deviation of 0.5576194

```
> ci95=c(Poisson_estimate$estimate+c(-1,1)*1.96*Poisson_estimate$sd)
```

ci95 gives the 95% confidence interval limits of Poisson_estimate Lambda
lower 95% confidence limit=34.97603
higher 95% confidence limit=37.16190
Negative binomial estimate can be derived using the fitdist function

```
> nbinomial_estimate=fitdist(freq, "nbinom")
```

It gives mu estimate of 36.06861 with a standard error of 0.6149304
goodfit function of vcd package can also be used to find Poisson and Negative Binomial estimates of location and standard error.

```
> gf_pois=goodfit(freq, type="poisson")
> gf_nbinomial=goodfit(freq, type = "nbinomial")
```

Frequency plot of Poisson distribution > plot(gf_pois, type = "standing", scale="raw")

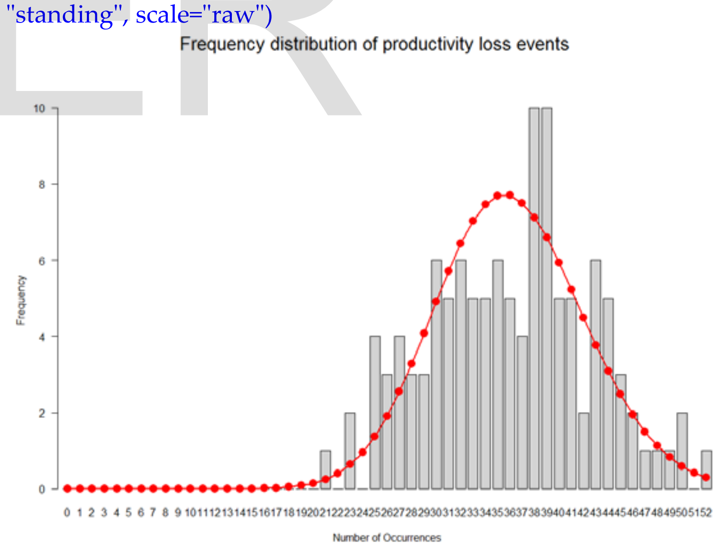


Fig: 4

> plot(gf_pois, type = "hanging", scale = "sqrt") gives the hanging bar distribution chart

ISSN 2229-5518odd page

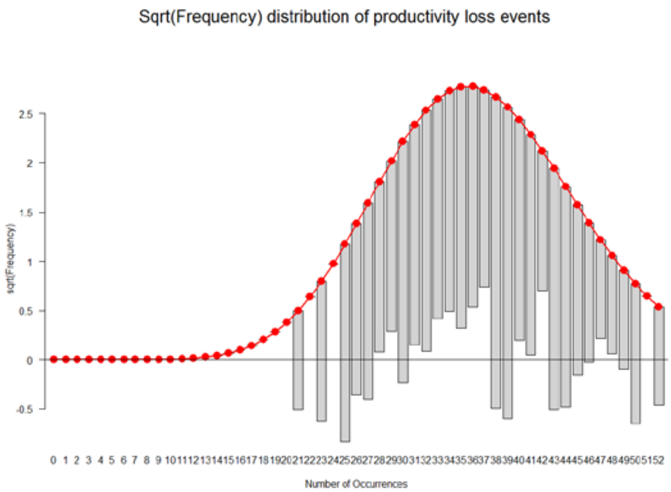


Fig: 5

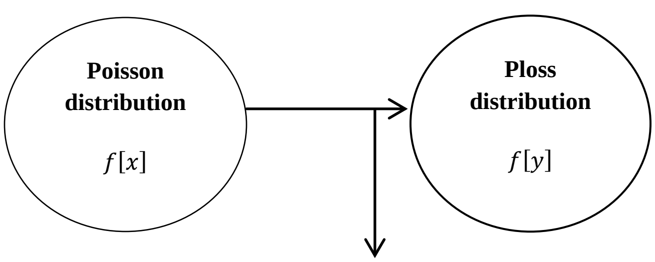
Next step we will modify the Poisson_estimate to account for the assumption of productivity loss cutoff of 2% since there will always be some level of variability in the process.

```
> Poisson_estimate= Poisson_estimate$estimate / (1 - plnorm(q=Lower_Cutoff, meanlog=fit_ln_loss$estimate[1], sdlog=fit_ln_loss$estimate[2]))
```

4) Convolution using Monte Carlo simulation method

Convolution is a statistical technique that combines two distributions to form a third distribution. If there are two functions $f[x]$ and $f[y]$ then convolution is denoted by $f[x] * f[y]$ and is defined as integral of the product of $f[x]$ and $f[y]$ after one is randomly selected or changed. Convolution definition assuming discrete time function

$$C[n] = f[x] * f[y] = \sum_{i=-\infty}^{\infty} f[xi] * f[y-i]$$



Select random value $\omega i = 2$

$$Ploss(x1i) = 4, Ploss(x2i) = 9 \ then$$

$$\sum Ploss(xij) = 13 \ hours$$

It randomly selects one random number ωi from the Poisson distribution, e.g. $\omega i = 2$. Then it extracts ωi random numbers from the distribution of magnitude of $Ploss(xij)$

$$\text{e.g. } Ploss(x1i) = 4, Ploss(x2i) = 9,$$
$$\sum Ploss(xij) = 13 \ hours$$

Integral of *Ploss(xij)* is taken to arrive at the magnitude of productivity loss for weekly time interval.
This process is then simulated million times to create the distribution of weekly magnitude of productivity loss.

```
> G_Pareto_estimate=gpdfit$par.ests
> e=fit_ln_loss$estimate
> quantile_loss_magnitude = function(p, e, G_Pareto_estimate, u){
+  Fu = plnorm(u, meanlog=e[1], sdlog=e[2])
+  x = ifelse(p<Fu,
+  qlnorm( p=p, meanlog=e[1], sdlog=e[2] ),
+  qgpd( p=(p - Fu) / (1 - Fu) , xi=G_Pareto_estimate[1], mu=meplot_cutoff, beta=G_Pareto_estimate[2]) )
+  return(x)
+ }
> # Random sampling function
> random_generator_loss_magnitude = function(n, e, G_Pareto_estimate, u){
+  r = quantile_loss_magnitude(runif(n), e, G_Pareto_estimate, u)
+ }
> set.seed(1234)
> simulation_count = 1000000
> pois_estimate = Poisson_estimate
> loc_log = fit_ln_loss$estimate[1] # Meanlog estimate [1]
> sd_log = fit_ln_loss$estimate[2] # Sdlog estimate [2]
> xiGPD= G_Pareto_estimate[1]
> betaGPD= G_Pareto_estimate[2]
> mk = rep(0,Simulation_count)
> freq = rpois(Simulation_count, pois_estimate)
> for(i in 1:Simulation_count)
+  mk[i] = sum(random_generator_loss_magnitude(n=freq[i], e=c(loc_log,sd_log), G_Pareto_estimate=c(xiGPD,threshold,betaGPD),u=meplot_cutoff))
```

Above process completes the convolution process to derive the simulated weekly productivity loss distribution

```
> summary(mk)
```

Summary of simulated weekly productivity loss distribution

Table: 3

```
> fit_mk=fitdistr(mk,densfun ="normal")
> hist(mk,xlim = c(75,500),breaks=100,col = 8,probability =
TRUE,main="Weekly Productivity Loss Distribution")
>
curve(dnorm(x,fit_mk$estimate[1],fit_mk$estimate[2]),col="
red",lwd=2,add = T)
```

Worst case scenario of Productivity loss is calculated by taking 95% quantile estimate of the below distribution (Fig: 6)

```
> WCScenario_x1=quantile(mk,0.95)
```

Expected scenario of Productivity loss is calculated by taking 50% quantile estimate of the below distribution (Fig: 6)

```
> Expected_Ploss_x1=quantile(mk,0.5)
```

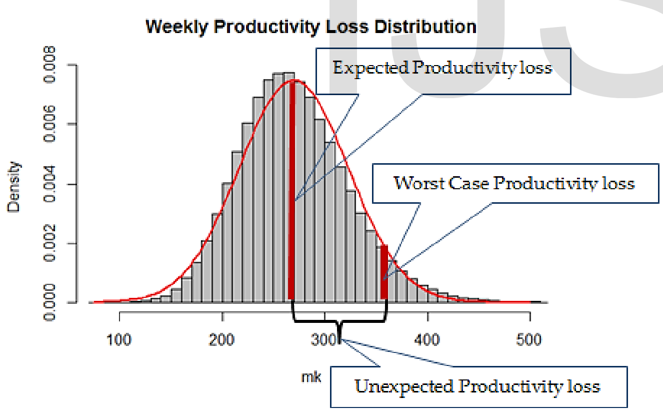


Fig: 6

After running the simulation we get the final weekly productivity loss distribution given above in Fig: 6.

5) Calculating the contingency plan for process x1

Basis the final distribution of productivity loss we will derive 3 components

5.1) Worst case Productivity loss @95% confidence level

Min	1st Qu.	Median	Mean	3rd Qu.	Max.
78.79	232.63	265.97	269.70	302.39	1047.99

for process function x1 (wcx1)

`> WCScenario_x1=quantile(mk,0.95)`= 362.4235 hours

5.2) Expected Productivity loss @50% confidence level for process function x1 (ex1)

`> Expected_Ploss_x1=quantile(mk,0.5)`=265.9857 hours

5.3) Unexpected Productivity loss is calculated as the difference between Worst case Productivity loss @95% & Expected Productivity loss @50% (ux1)

`>Unexpected_Ploss_x1=-Expected_Ploss_x1+WCScenario_x1`=96.4378 hours

Assuming 8 working hours per day equals 40 hours per week that an associate would work on claim processing cases. With an Unexpected productivity loss of 96.4378 hours it relates to 2.41 FTE (Full time equivalent~40 hours)

`> Contingency_plan_x1=Unexpected_Ploss_x1/40`=2.41 FTE.

The above process was done for one single process x1 for an operations function. Similarly we can do the same for other process x2

Let us assume that for process x2

Worst case Productivity loss @95% confidence level (wcx2) =335.9132 hours

Expected Productivity loss @50% confidence level (ex2) =257.336 hours

Unexpected Productivity loss (ux2) =78.57721 hours

`> Contingency_plan_x2=Unexpected_Ploss_x2/40`=1.96443 FTE

Assuming operations function has only two processes then the aggregate unexpected productivity loss is calculated using correlation between them. If there is negative or no correlation between processes then the final required FTE

Min	1st Qu.	Median	Mean	3rd Qu.	Max.
0.5100	0.6075	0.6700	0.6636	0.7200	0.8200

would be less than that of the sum of all the processes. If there is a perfect correlation then the final required FTE would be the sum of required FTE for all the processes. For the ease of application let's assume there is 50% correlation between processes x1 and x2. Final FTE required for overall operations level to plan for unexpected productivity loss is calculated below

- **5.4) Worst case Productivity loss for entire operations**=SQRT (wcx1^2+ wcx2^2+2* wcx1* wcx2*50%) =604.9226 hours
- **5.5) Expected Productivity loss for entire operations** =SQRT (ex1^2+ ex2^2+2* ex1* ex2*50%) =453.2305 hours
- **5.6) Unexpected Productivity loss for entire operations** (ux)=604.9226-453.2305=151.692 hours which is equal to an FTE worth of 3.79 that business unit needs to maintain to manage the unexpected productivity loss.

6) Accounting for Shrinkage in capacity utilization

In the above analysis of calculation for contingency planning we have assumed 8 working hours to arrive at the human capital requirement in form of FTE (Full time equivalent) where 1 FTE is equivalent to 40 hours' worth of effort of human capital requirement.

We can expect machines to run at 100% capacity utilization while the same will not be the case with human resource capital. Of the 8 hours available in a day we might expect the effort on core work to be around less than 8. If we assume utilization percentage of 80% it means the effort on core work out of 8 available hours comes to 6.4 hours which leads to 32 hours/week actually spent on core productive work. In order to account for shrinkage in capacity utilization we will analyze the Utilization levels for the referenced weekly time periods before. Dataset contains two variables "week" and "Utilization"

```
> summary(utilization)
```

Next step we will study the distribution of the historical Utilization levels using normal distribution assuming them to be random in nature.

```
> fit_n_Utilization=fitdistr(utilization,densfun ="normal")
```

> fit_n_Utilization This gives the location estimate of 0.663620690~66.36% and standard deviation estimate of 0.080701627~8.07%

```
> hist(utilization,xlim = c(0.51,0.84),breaks=40,col = 8,probability = TRUE,main="Capacity Utilization Distribution")
```

```
>curve(dnorm(x,fit_n_Utilization$estimate[1],fit_n_Utilization$estimate[2]),col="red",lwd=2,add = T)
```

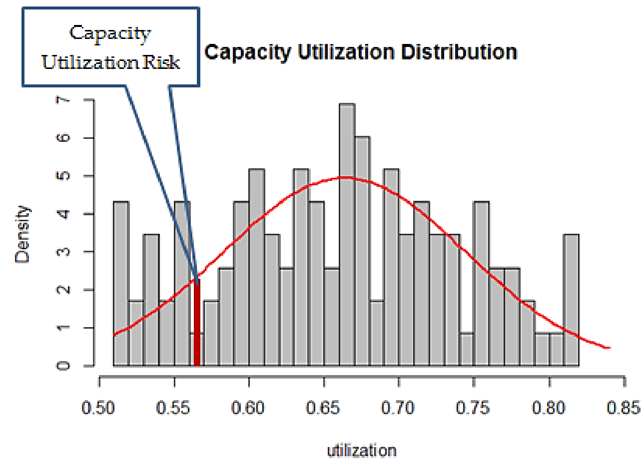


Fig: 7

Utilization level of 66.36% is the expected capacity utilization scenario but since we are dealing with risk of unexpected productivity loss we should also account for the risk of utilization falling to lower end of the distribution curve.

```
>lower_ci90_util=qnorm(0.05,mean=fit_n_Utilization$estimate[1],sd=fit_n_Utilization$estimate[2],lower.tail = TRUE,log.p=FALSE)
```

```
> lower_ci90_util
```

[1] 0.5308783

=53.08%

```
>Upper_ci90_util=qnorm(0.9,mean=fit_n_Utilization$estimate[1],sd=fit_n_Utilization$estimate[2],lower.tail = TRUE,log.p=FALSE)
```

```
> Upper_ci90_util
```

[1] 0.767044

=76.70%

So at a 90% confidence level we expect the capacity utilization to fall in the interval between 53.08% and 76.70%.

Risk here is capacity utilization falling to the level of 53.08%.

In the section 5.6 we had derived the FTE requirement to manage the risk of productivity loss to be 3.79 FTE~151.692 hours.

After accounting for Capacity utilization shrinkage the final requirement stands to be

>3.79/ lower_ci90_util

=7.1391~7 FTE

This is the optimal additional FTE plan that business operation needs to maintain to absorb the extreme negative deviation into Productivity loss.

If current plan is maintained above the level of 7 FTE to withstand unexpected scenarios then that additional delta is adding to the overall cost of the firm. If the same is maintained below 7 FTE than the business is running the risk of operations in terms of possible increase in turnaround times, unable to meet the client deliverables on time can have the impact on the Net Promoter Score (NPS) that measures the willingness of customers to recommend a company's services to others.

Conclusion

The Primary objective of the paper is to describe how using loss distribution approach a financial services firm can create a contingency plan in terms of maintaining optimal resources to manage the unexpected productivity losses that can arise due to operational inefficiencies. There is always a tradeoff between managing cost and implementing risk control initiatives. Additional FTE contingency plan proposed through this analysis helps to strike a balance between this tradeoff so that an organization can decide what could be the effective bench staff they would maintain. Loss distribution approach has its applications across financial risk domain to model the economic capital requirement for operational risk & credit risk. Productivity loss directly impacts the bottom line of the firm. Generally any business would maintain additional resources to manage the expected productivity losses. With this modeling approach expectation is to optimize additional resource requirement that the business is maintaining, and if there no contingency plan for such a scenario it highlights how effectively the risk can be hedged by maintaining the optimal resources.

References

1. Baud, N., A. Frachot, and T. Roncalli, 2003, "How to avoid over-estimating capital charge for operational risk?"
2. Introduction to R for Quantitative Finance: By: Agnes Vidovics-Dancs; Daniel Havran; Dr. Edina Berlinger; Gergely Daróczi; Márton Michaletzky; Michael Puhle; Péter Csóka; Dr. Kata Váradi; Zsolt Tulassay
3. Topics in Data Analysis Using R in Extreme Value Theory : By Helena Penalva1,Manuela Neves2 and Sandra Nunes3
4. The advanced measurement approach for banks: Jan Lubbe1
and Flippie Snyman2 (http://www.bis.org)
5. Loss Distribution Approach for operational risk∗
A. Frachot, P. Georges & T. Roncalli†
6. Stephenson, A. G. (2012): ismev: An Introduction to Statistical Modeling of Extreme Values. Original S functions written by Janet E. Heffernan with R port and R documentation. R package version 1.38, URL http://CRAN.Rproject.org/package=ismev.
7. Loss Distribution Approach in Practice∗ Antoine Frachot, Olivier Moudoulaud, Thierry Roncalli
8. Basel Committee on Banking Supervision, 2001, Operational Risk — Consultative Document, Supporting document to the New Basel Capital Accord, January.

Disclaimer: The views, opinions, findings or recommendations expressed in this paper are strictly those of the author. They do not necessarily reflect the views of any specific industry.

9. De Fontnouvelle, P., V. DeJesus-Rueff, J. Jordan, and E. Rosengren, 2003, “Using loss data to quantify operational risk”

10. Moscadelli, M. (2004) The Modelling of Operational Risk: Experience with the Analysis of the Data, Collected by the Basel Committee, Banaca d’Italia, report 517-July 2004

11. Fabio Piacenza - UniCredit Operational Risk Methodologies and Control

12. Convolution Process http://www.songho.ca/dsp/convolution/convolution.html.