Paper 163-2008

Loss Distribution Approach for the Operational Risk Economic Capital

Sabri Guray Uner, PNC Financial Services & University of Pittsburgh, Pittsburgh, PA

ABSTRACT

Following the Basel II Capital Accord, with the increased focus on operational risk as a distinct theme from credit and market risk, quantification of the operational risk has been a major challenge for the financial institutions. In this study, we propose a loss distribution based approach consistent with Basel II guidelines, to estimate the Economic Capital for the Operational Risk at the firm level. This approach accounts for both firm specific and industry related ('Financial Services') components of the operational risk. To serve this purpose, inline with the guidelines, the method utilizes internal loss data for the institution itself and the external loss data for the 'Financial Services' industry. The results suggest that, after addressing the selection bias in external data as suggested by the previous studies; proposed LDA based approach provides Economic Capital estimate within the expected range. Model's EC estimation over the past nine quarters suggests that, two components of the proposed model are consistent and comparable overtime.

INTRODUCTION

International Basel II Accord defines operational risk as the risk of loss resulting from inadequate or failed internal processes, people and systems or from external events. Despite defining qualitative guideline and quantitative data standards for the Advanced Measurement Approach (AMA), Basel II Accord does not restrict the quantification method. Under the AMA methodology, financial institutions are required to develop their own internal measurement methods that estimate the expected and unexpected operational losses based on the combined use of internal and relevant external data. Moreover, the approach should be comparable to internal rating approaches used for credit risk at one year holding period and 99.9th percentile confidence interval. And the bank should demonstrate that the approach is robust to capture potentially severe 'tail' loss events.

Since it is the most recent risk component brought to regulators' attention, there is no general consensus approach that is clearly defined and adopted by the banks. In this study, we introduce our model as a first step to quantify the operational risk at the enterprise level consistent with the guidelines. Proposed operational risk Economic Capital model (OpRisk EC Model[1]) is inline with the Basel II Advanced Measurement Approach's (AMA) quantitative criteria, by providing a risk measure generated by the bank's own operational risk measurement. The Economic Capital estimate provided by the model captures the one year total operational risk measure (EL+UL) at 99.9th percentile confidence interval, and accounts for the severe 'tail' loss events both experienced internally by the bank and by the industry peers.

Reported results[2] utilize PNC Financial Services' Internal Loss Data (ILD hereafter) and Fitch's Algo OpData®™ Industry Loss Data (ELD hereafter). The paper is organized as follows. First section provides an overview of the operational loss databases used for modeling. The next section introduces the LDA methodology along with the model's two components, 'Firm Specific' and the 'Industry Component' and discusses the results for the Economic Capital (EC hereafter). Later we study the robustness of the model components to represent the 'tail' loss events followed by the discussion to combine the two components. And the last section concludes the study.

OPERATIONAL LOSS DATA

Consistent with the Basel II Standards on Operational Risk AMA, this methodology provides the EC estimates based on the combined use of internal and the relevant external loss data. Internal loss data is essential to represent the bank's own business activities, technological processes and risk management procedures. Hence, it is utilized both by the 'Firm Specific Component' and the 'Industry Component'. However, due to limited historical internal loss data, the methodology also incorporates relevant external loss data to estimate tail events from a more robust dataset and to capture industry wide operational risk trend.

---

1 This approach is still a working model and developed by the PNC Financial Services' Inc's Corporate Performance Measurement group. Sabri Guray Uner is employed by the PNC Financial Services Corporate Performance Measurement as an Asst. V.P.

2 Some confidential results are censored or scaled without loss of generality and the quality.

INTERNAL LOSS DATA (ILD):

Internal loss data is provided by the PNC Financial Services Group, Inc. With total assets of $138.92 billion, The PNC Financial Services Group Inc. is a diversified financial services company, with businesses engaged in retail banking, corporate and institutional banking, asset management and global fund processing services. This database covers PNC's historical operational loss events with the loss amount and the event date[3]. ILD covers operational loss events dated back to December 1993. However, loss event coverage is not complete until 2002. Only operational loss events since January 2002 are used for the modeling purpose.

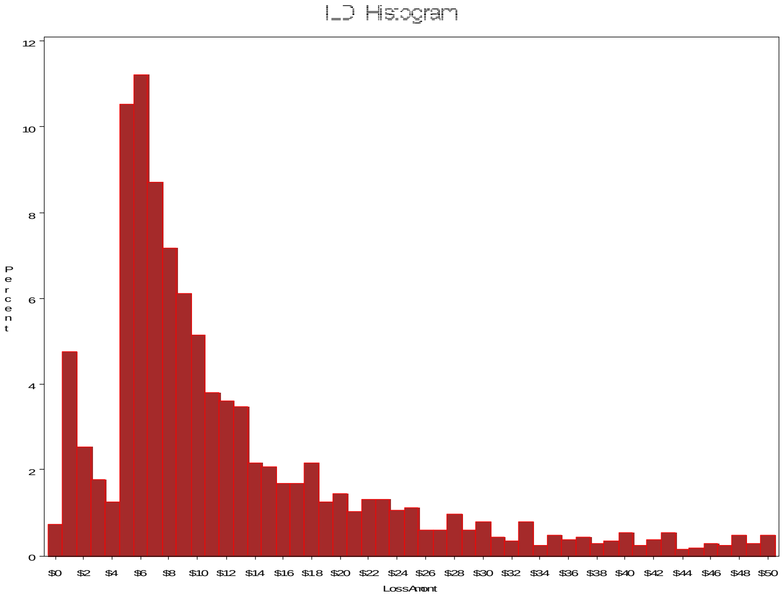


Figure 1. *ILD Loss Histogram* **(ILLUSTRATIVE)**

As evident from Figure 1, event coverage is not complete for loss amounts less than $5[4]. Operational loss events with loss amount greater or equal to $5 are used for modeling. As of December 2007, ILD sample has 1963 loss events. Annual and monthly event frequencies are provided by Figure 2-A & 2-B.

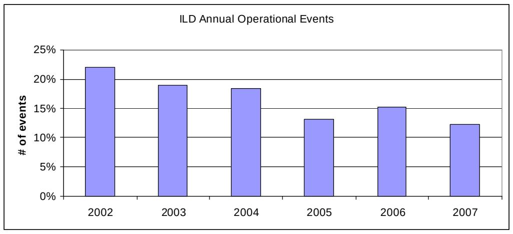


Figure 2-A. *ILD Annual Loss Events*

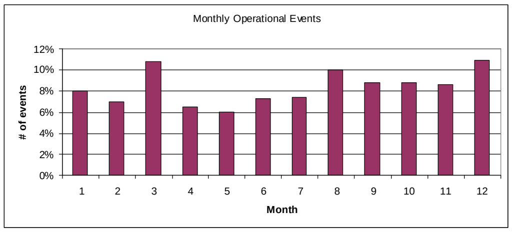


Figure 2-B. *ILD Monthly Loss Events*

Over the analysis period 2002-2007, annual number of events has an apparent declining trend. Within a given year, operational events are dispersed and not clustered in a given month or a quarter, but there is slightly more events towards the end of the year. Regarding the loss severity statistics reported in Table 1, the internal loss data has a

3 The database includes additional date attributes such as 'Discovery Date', 'Recovery Amount' and corresponding Business Line, Event Type descriptions.

4 This value is scaled for confidentiality.

heavy-tail with 99% percentile loss amount varying from $99 to as high as $63,750[5] over different periods.

			Loss Amount $									
Year	**Quarter**	**N**	**Mean**	**Median**	**Skewness**	**Minimum**	**Maximum**	**1st Pctl**	**10th Pctl**	**50th Pctl**	**90th Pctl**	**99th Pctl**
2002	1	92	$22.6	$11.1	2.46	$5.0	$137.3	$5.0	$5.6	$11.1	$75.4	$137.3
	2	99	$103.0	$12.5	9.92	$5.0	$7,859.2	$5.0	$5.2	$12.5	$50.0	$7,859.2
	3	105	$34.3	$9.9	6.49	$5.0	$817.1	$5.0	$5.4	$9.9	$48.1	$392.4
	4	136	$35.5	$11.6	9.67	$5.0	$1,336.3	$5.0	$5.5	$11.6	$57.3	$288.4
2003	1	89	$861.1	$10.0	9.12	$5.0	$63,750.0	$5.0	$5.5	$10.0	$65.6	$63,750.0
	2	86	$26.0	$11.7	7.04	$5.1	$508.5	$5.1	$5.6	$11.7	$49.3	$508.5
	3	95	$27.2	$11.5	3.68	$5.0	$247.8	$5.0	$5.7	$11.5	$50.0	$247.8
	4	100	$26.1	$11.9	6.98	$5.0	$487.5	$5.0	$5.5	$11.9	$46.2	$306.3
2004	1	100	$28.6	$10.2	5.68	$5.0	$531.6	$5.1	$5.5	$10.2	$52.2	$428.3
	2	61	$21.0	$11.2	5.43	$5.0	$249.8	$5.0	$5.4	$11.2	$42.9	$249.8
	3	88	$20.6	$10.3	8.22	$5.0	$450.0	$5.0	$5.8	$10.3	$35.9	$450.0
	4	112	$15.1	$10.0	4.68	$5.1	$131.2	$5.1	$5.8	$10.0	$24.2	$122.8
2005	1	66	$62.6	$9.9	5.85	$5.1	$1,824.8	$5.1	$5.6	$9.9	$38.8	$1,824.8
	2	46	$31.1	$12.3	6.12	$5.0	$550.0	$5.0	$6.1	$12.3	$51.4	$550.0
	3	78	$42.6	$9.3	6.52	$5.0	$1,133.9	$5.0	$5.4	$9.3	$72.4	$1,133.9
	4	69	$16.3	$10.0	2.60	$5.0	$98.6	$5.0	$5.2	$10.0	$46.9	$98.6
2006	1	75	$38.5	$10.0	4.92	$5.0	$649.7	$5.0	$5.5	$10.0	$55.8	$649.7
	2	79	$160.4	$13.8	8.86	$5.0	$9,997.1	$5.0	$6.0	$13.8	$96.1	$9,997.1
	3	67	$26.4	$11.9	7.43	$5.0	$555.0	$5.0	$5.6	$11.9	$48.8	$555.0
	4	79	$74.3	$11.4	7.05	$5.0	$2,287.6	$5.0	$5.5	$11.4	$120.9	$2,287.6
2007	1	64	$25.7	$9.6	4.02	$5.0	$275.0	$5.0	$5.2	$9.6	$50.0	$275.0
	2	61	$18.8	$9.5	3.50	$5.0	$129.8	$5.0	$5.5	$9.5	$39.8	$129.8
	3	56	$23.4	$7.8	3.84	$5.0	$255.7	$5.0	$5.2	$7.8	$65.7	$255.7
	4	60	$37.9	$8.9	4.52	$5.1	$575.0	$5.1	$5.3	$8.9	$66.1	$575.0

Table 1. *ILD Loss Severity Descriptive Statistics* **(ILLUSTRATIVE)**

FITCH'S ALGO OPDATA® EXTERNAL LOSS DATABASE (ELD):

The Algo OpData® database covers over 12,000 publicly reported operational losses from various geographical locations and industries, making it the most complete operational loss database available. Algo OpData® is a key component of the modeling process, supplementing ILD in order to populate 'tail' (high severity, low frequency) events. For each operational loss event, this database provides Event ID, Organization, Firm name, Location, Event Description, 3 Business Levels (Industry Sector and 2 Business Unit Levels), 3 Event Type Level along with the Settlement Date, Loss Amount in Local currency, Loss Amount in USD.

We apply data filtering to ensure that historical loss observations included in the modeling are relevant, i.e. observed in the 'Financial Services' industry. We also exclude events classified under "Insurance" within the 'Financial Services' industry. This procedure is similar to one used by previous studies such as Fontnouevelle et al. (2003) and Brandts (2004).

ELD covers operational loss events as early as December 1972. However, the database's coverage does not seem to be complete for operational events prior to year 1990 as evident from Table 2-A. We consider only the events since 1990 for modeling purpose. Furthermore, since the analysis period is 18 years, all the loss amounts used for modeling are adjusted by Consumer Price Index[6] and expressed in end of 2007 dollar amount.

Year	N	Year	N
1972	4	**1990**	118
1973	3	**1991**	127
1974	5	**1992**	139
1975	1	**1993**	122
1976	3	**1994**	167
1977	4	**1995**	147
1978	6	**1996**	162
1979	15	**1997**	209
1980	16	**1998**	247
1981	15	**1999**	234
1982	18	**2000**	207
1983	15	**2001**	209
1984	36	**2002**	290
1985	44	**2003**	237
1986	55	**2004**	186
1987	65	**2005**	235
1988	51	**2006**	162
1989	67	**2007**	51

Table 2-A. *ELD Annual Loss Events*

End of 2007 Q1[7] , ELD sample has 3248 relevant loss events. Descriptive statistics for the event frequency are provided on Table 2-B. Annual number of events is relatively higher over the 1997-2006 periods. There is an apparent clustering in December for the early 1990-2000 period. Median monthly number of loss events, ranges from 11 to 25.

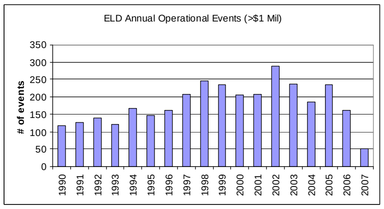
ELD covers only the publicly reported operational losses that are at least $1 Million in total amount. So the observations are left truncated. Only the operational events from the related industries with reported amounts greater or equal to $1 Million are used for modeling purpose. Figure 2 presents the histogram for the ELD loss distribution for observations less than or equal to $1 Billion.

5 Values are scaled for confidentiality.

6 U.S. Department of Labor: ftp://ftp.bls.gov/pub/special.requests/cpi/cpiai.txt

7 Latest update available for the ELD was at the end of 2007-Q1.

	Month												
Year	1	2	3	4	5	6	7	8	9	10	11	12	Annual Total
1990	5			2	3	3				2		103	**118**
1991	5	2	2		2		4	4	1	1	5	101	**127**
1992	8	6	3	2	1	4	1	1	5	3	1	104	**139**
1993	8	5	2	1	2	1	6	5	2	3	3	84	**122**
1994	23	5	2	3	2	4	5	4	4	2	6	107	**167**
1995	10	7	4	1	3	3	1	4	3	4	2	105	**147**
1996	13	1	2	4	6	6	1	3	1	3	8	114	**162**
1997	20	1	7	6	2	4	2	9	1	5	2	150	**209**
1998	16	5	9	5	4	3	6	6	3	7	3	180	**247**
1999	22	4	6	3		4	9	6	5	5	7	163	**234**
2000	21	3	4	4	5	8	2	4	4	18	3	131	**207**
2001	21	7	9	11	12	56	11	15	8	9	16	34	**209**
2002	32	33	22	19	32	22	9	14	23	22	27	35	**290**
2003	35	11	15	17	14	36	11	9	16	13	19	41	**237**
2004	29	25	24	9	10	11	15	15	10	6	5	27	**186**
2005	13	12	46	17	13	36	11	12	6	24	18	27	**235**
2006	12	12	10	6	16	14	20	11	15	18	11	16	**161**
2007	20	13	18										**51**
Monthly Median	**25**	**13**	**20**	**17**	**14**	**22**	**11**	**12**	**15**	**18**	**18**	**27**	**3248**



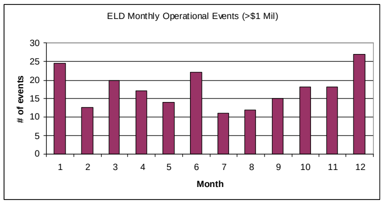


Table 2-B. *ELD Event Frequency Descriptive Statistics*

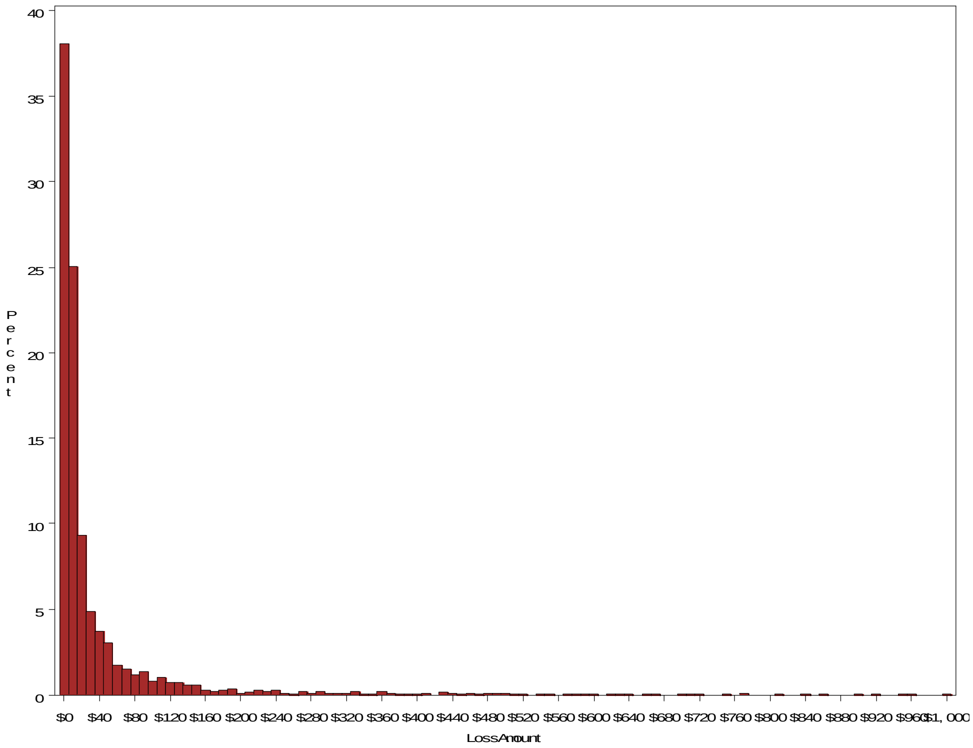


Figure 3. *ELD Loss Histogram*

Descriptive statistics for the loss severity are provided in Table 2-C. Other than being truncated, ELD presents typical characteristic of a loss data with highly skewed loss amount. Loss amount for the 99th percentile observation ranges from $158 million to $3.6 billion.

Loss Amount $											
Year	**N**	**Mean**	**Median**	**Skewness**	**Minimum**	**Maximum**	**1st Pctl**	**10th Pctl**	**50th Pctl**	**90th Pctl**	**99th Pctl**
1990	118	$48.4	$9.6	6.31	$1.5	$1,361.9	$1.5	$2.0	$9.6	$72.4	$748.3
1991	127	$270.4	$6.3	10.91	$1.5	$25,188.0	$1.5	$1.5	$6.3	$120.0	$3,677.4
1992	139	$61.4	$7.7	5.67	$1.4	$1,411.6	$1.4	$1.4	$7.7	$104.6	$1,397.8
1993	122	$33.4	$6.7	6.34	$1.4	$918.3	$1.4	$1.7	$6.7	$48.2	$712.8
1994	167	$62.3	$12.3	9.41	$1.3	$2,466.9	$1.3	$1.6	$12.3	$128.1	$582.4
1995	147	$120.4	$6.6	10.54	$1.3	$7,625.0	$1.3	$1.6	$6.6	$160.5	$1,761.9
1996	162	$45.3	$8.0	3.62	$1.3	$636.2	$1.3	$1.8	$8.0	$101.8	$548.4
1997	209	$46.8	$6.8	9.78	$1.3	$2,502.2	$1.3	$1.7	$6.8	$93.8	$388.9
1998	247	$39.5	$7.4	5.09	$1.2	$861.9	$1.2	$1.6	$7.4	$82.9	$771.9
1999	234	$48.8	$6.8	8.12	$1.2	$2,257.6	$1.2	$1.6	$6.8	$78.0	$1,141.8
2000	207	$31.7	$7.5	4.50	$1.2	$574.6	$1.2	$1.7	$7.5	$76.1	$361.8
2001	209	$61.4	$8.0	8.10	$1.1	$2,382.7	$1.1	$1.5	$8.0	$102.3	$697.8
2002	290	$58.3	$10.4	11.93	$1.1	$3,017.6	$1.1	$1.6	$10.4	$139.3	$552.7
2003	237	$144.7	$9.4	14.53	$1.1	$17,028.0	$1.1	$1.4	$9.4	$148.1	$1,428.1
2004	186	$77.1	$14.5	7.67	$1.1	$2,725.9	$1.2	$1.9	$14.5	$122.7	$1,293.6
2005	235	$87.2	$7.6	6.19	$1.0	$2,804.0	$1.0	$1.4	$7.6	$133.6	$2,205.3
2006	161	$83.7	$9.3	11.16	$1.0	$5,881.0	$1.0	$1.5	$9.3	$99.0	$1,567.1
2007	51	$23.6	$5.8	2.58	$1.3	$158.7	$1.3	$1.6	$5.8	$55.8	$158.7

Table 2-C. *ELD Loss Severity Descriptive Statistics*

LDA BASED AMA METHODOLOGY

This Operational Risk Economic Capital Model (OpRisk EC Model) is a Loss Distribution Approach (LDA) based methodology. As mentioned before, it utilizes internal (ILD) and industry (‘Financial Services’ ELD) operational loss data. Parallel with this, the model has two components to populate industry trend based on the relevant loss data characteristics observed in the ‘Financial Services’ industry and to capture the firm specific profile based on bank’s own operational loss data characteristics.

Industry Component addresses the loss distribution in two stages (Body and Tail Distribution), by utilizing ILD and the industry data (ELD). ELD supplements internal loss data in order to populate 'tail' (high severity, low frequency) events so that the ‘Industry Component’ can provide insights into risks yet to be experienced. Hence, the industry data serve the purpose of fitting the ‘tail’ events using a more robust loss data rather than a limited historical observation from ILD. This component overcomes the data limitation of ILD used in ‘Firm Specific Component’ for modeling extreme losses. ‘Industry Component’ therefore mostly serves to capture the industry wide operational risk characteristics with the robust ELD. Due to longer observation period in ELD, we adjust dollar loss amounts for the CPI level. Moreover, for the selection bias that is implicit in ELD, we implemented an econometric truncation framework to model the operational loss data.

Firm Specific Component also implements the Extreme Value Theory (EVT hereafter) due to highly skewed loss data. This model is sensitive to firm’s operational loss events through utilizing ILD, and accounts solely for the firm specific operational profile. Despite relatively limited historical loss observation, ILD is a better representation of the bank’s own business operations and risk characteristics, and serves to capture firm specific risk component.

INDUSTRY COMPONENT:

‘Industry Component’ addresses loss distribution in two stages (Body and Tail Distribution) for loss amounts below $1 Million and for loss amounts equal to or greater than $1 Million. ‘Industry Component’ utilizes both ILD and ELD in the loss distribution estimate. ELD is used for the severity tail distribution to overcome the data limitation in ILD to get robust representation of the extreme losses and also to account for loss events experienced by the industry (‘Financial Services’). ILD is used to fit severity and frequency distribution for losses below $1 Million, and to fit the frequency distribution above $1 Million.

We use NLMIXED[8] procedure from SAS/STAT® software to estimate the Maximum Likelihood Estimates (MLE hereafter) for the underlying distributions and used Chi-square statistics to choose the best-fit. Following distributions are used to estimate the distribution parameters.

8 PROC NLP is also an alternative with the SAS/OR® Software.

Body frequency distribution (ILD):	Poisson (λ)
Body severity distribution (ILD):	Log-normal (μ, σ)
Tail frequency distribution (ILD):	Poisson (λ)
Tail severity distribution (ELD):	Exponential[9] (b)

Body Distributions and Tail Frequency:
PROC NLMIXED is a convenient way to get Maximum Likelihood Estimates for various pre-defined distributions such as Normal, Bernoulli, Binomial and Poisson. The following code is an example to get MLE for the Possion "λ" parameter.

```
PROC NLMIXED data=input;
    MODEL loss_freq ~ POISSON (lambda);
    ODS OUTPUT ParameterEstimates = output_Poiss_MLE;
RUN;
```

Moreover, PROC NLMIXED is capable of fitting the user-defined distributions through general log-likelihood option in the model statement. For the severity distributions of losses below $1 Million, we define the log-likelihood for the log-normal distribution and use this option to estimate the "μ" and "σ" for the log-normal distribution. Alternatively, one can fit the normal distribution to log (Loss Amount) and apply the well known parameter conversion from normal to log-normal.

Tail Severity Distribution[10]:

Quantifying operational risk from public data (ELD) poses several challenges, the most important being that not all operational losses are publicly reported and therefore ELD has the inherited selection bias. Baud et al. (2002) propose using a random truncation framework, which is implemented by Fountnouvelle et al. (2003).

For Tail Severity distribution we implement the method applied by Fountnouvelle et al. (2003). Methodology suggested by Fountnouvelle et al. (2003) addresses the problem of sample selection bias in ELD. In this approach, truncation level for the publicly reported losses is modeled as an unobserved random variable with an assumed distribution. To model the underlying loss distribution, this model relies on EVT, which suggests that the logarithm of losses above a threshold would have an exponential distribution. We assume that the distribution of operational losses belongs to the heavy-tailed class of distributions, and that the distribution of log losses belongs to the light-tailed class. For the 'truncation point' distribution, we assume logistic distribution following the analysis by Fontnouvelle et al. (2003).

The parameters for unconditional loss distribution are then obtained by estimating the conditional likelihood function for the log losses adjusted for the reporting (selection) probability. Therefore, the likelihood function for the observed loss data can be expressed by the below equation. Consistent with the EVT, log losses are used to estimate MLE for the parameters, and hence we denote the likelihood function;

$$L(b, b, t) = \prod_{i=1}^{N} \left(\frac{\exp(-x_i / b)/b}{1+\exp(-b\,(x_i - t\,))} \Bigg/ \int_{u}^{\infty} \frac{\exp(-x/b)/b}{1+\exp(-b\,(x-t\,))}\,dx \right)$$

We estimate MLE for the underlying loss distribution parameters from the above likelihood function. To implement this method, we demand a numerical integration function within the likelihood optimization. Despite being very flexible to define user-defined distributions, PROC NLMIXED is not an alternative for this method. For this purpose, we utilize NLPxyz function within SAS/IML® software and use call QUAD function for the numerical integration. However, instead of this pre-defined 'QUAD', one can define a custom function to implement numerical integration methods such as Gaussian quadratures.

The following code estimates the "b", "β" and "τ" from the above likelihood function. Fitting this function is computationally demanding because each iteration for the optimization problem requires the numerical integration

9 We fit exponential distribution to log (Loss Amount) which is suggested by EVT and we use Logistic distribution for truncation level distribution.

10 This section details the econometric model for truncated ELD and borrows from the study by Fountnouvell et al (2003).

to be computed for the conditional probability distribution represented in the denominator.

```
PROC IML;
    Use input;
    Read all var{lossvar} into ext_data;
    ext_data = log(ext_data);

    **************************;
    *Integrand Function;
    START LOGIC(x) global(b,beta,thao);
          tail_cond = (exp(-x/ b) / b) / ( 1 + exp(- beta *(x-thao)) );
          return (tail_cond);
    FINISH LOGIC;
    **************************;

    **************************;
    *LogLikelihood Function;
    START FUNC(parm)    global(ext_data,b,beta,thao);
    b = parm[1]; beta = parm[2];     thao = parm[3];
    limits = {0 100};
          sum1 = 0;
                do i =1 to nrow(ext_data);
                      g1 = exp(-ext_data[i]/parm[1]) / parm[1];
    if g1 = 0 then g1 = 1e-6;
                      log1 = log(g1);
                      log2 = log(1 + exp(-parm[2]*(ext_data[i]-parm[3])));
                      sum1 = sum1 + log1 - log2;
                end;
    CALL QUAD (integral,"LOGIC",limits);
    LL = sum1 - nrow(ext_data)*log(integral);
    return(-LL);        *Negative LogLikelihood;
    FINISH FUNC;
    **************************;

    parm={0.1    0.1    5};
    optn = {0 2};             *MINIMIZE -LL;
    con =        {1e-6  1e-6   1e-6,  .P     .P     .P};

    * MLE Optimization for parameters b, Beta and Thao;
    CALL NLPNRR (rc, xres, "FUNC", parm ,optn , con);
    ODS OUTPUT ParameterEstimates = output_truncExpo_MLE;
QUIT;
```

CPI adjustment affect on the threshold:

When loss amounts are adjusted for the CPI level, level of truncation that is implicit in the data will be higher than $1 Million. This is a side effect of the CPI adjustment, because a $1 Million loss in 90's will be deeper in the tail than $1 Million loss from 2007. When we adjust the historical loss amounts to 2007 dollar value, observation around the truncation level will be understated and represented mostly by the events from the most recent years. Failure to account for this issue will not address the selection bias perfectly and will impose an artificial skewness in the observed sample.

To address this issue we use the following approach. We adjust loss amounts to 2007 dollar amount, but use higher truncation point, which is $1 Million adjusted for the CPI level in the first observation year. If the first year is 1990 with 160% CPI adjustment, the new implied truncation point will be $1.6 Million. This cut-off eliminates loss observations between $1 Million and $1.6 Million which are dominated by the recent years.

Estimation and Analysis:

To assess the overtime consistency of our modeling assumptions, we implement the 'Industry Component' over twelve fiscal periods (2005-Q1 to 2007-Q4). Table 3 provides parameter estimates for the underlying distributions. Parameter estimates are consistent overtime. Due to sound historical observation (1990-Q1 to 2007-Q1 from ELD) on tail severity, parameter estimates for the extreme loss values are consistent overtime.

Parameters for Industry Component													
	Period	Q1 2005	Q2 2005	Q3 2005	Q4 2005	Q1 2006	Q2 2006	Q3 2006	Q4 2006	Q1 2007	Q2 2007	Q3 2007	Q4 2007
ILD	Body Frequency												
ILD	Poisson λ	47.00	45.29	44.83	44.19	43.79	43.53	43.00	42.80	42.29	41.75	41.15	40.69
ILD	Body Severity ($ mil)												
ILD	LogNorm μ	0.040	0.040	0.040	0.040	0.040	0.041	0.041	0.042	0.042	0.042	0.042	0.042
ILD	LogNorm σ	0.044	0.044	0.044	0.043	0.045	0.047	0.046	0.048	0.049	0.048	0.048	0.049
ILD	Tail Frequency												
ILD	Poisson λ	0.27	0.25	0.27	0.25	0.24	0.25	0.24	0.25	0.24	0.23	0.22	0.21
ELD	Tail Severity ($ mil)												
ELD	Expon. "b"	0.864	0.888	0.890	0.877	0.861	0.864	0.859	0.871	0.863	0.863	0.863	0.863

Table 3. *Industry Component Parameter Estimation* **(ILLUSTRATIVE)**

Simulation Results:

Monte Carlo simulation is used to generate annual loss distribution, by combining fitted distributions and the estimated parameters. We then use 99.9% percentile level to determine the Economic Capital for the bank's operational risk. This EC captures the sum of the expected and unexpected losses at enterprise level.

We use SAS/IML software for Monte Carlo simulation, and define the modules for each distribution used for parameter estimation. We tested the convergence of the simulation result and it is attained with minimum 500,000 losses. In table 4, we report simulation results for 30 runs of 1,000,000 annual losses.

Monte Carlo simulation algorithm for the 'Industry Component' is as follows:

1) Generate Annual Loss from the Body distribution:
   i. For each month, generate number of loss events from the Poisson distribution
   ii. Calculate the annual number of loss events ($Nloss_i$) as the summation of 12 months.
   iii. For each year, generate $Nloss_i$ body losses from Log-Normal distribution
   iv. Calculate body loss amount $Bloss_i$ for that year as the summation of loss amounts for the $Nloss_i$ events.
2) Generate Annual Loss from the Tail Distribution:
   i. For each month, generate number of loss events from the Poisson distribution
   ii. Calculate number of annual loss events ($Nloss_i$)as the summation of 12 months.
   iii. For each year, generate $Nloss_i$ tail losses from exponential distribution (we actually generate log (losses) and take the exponential).
   iv. Calculate tail loss amount $Tloss_i$ for that year as the summation of loss amounts for $Nloss_i$ events.
3) Calculate annual loss as $Loss_i = Bloss_i + Tloss_i$
4) Repeat steps 1 to 3 for N years (we use 1,000,000)
5) Operational Value at Risk (OpVaR) is the annual loss, which, with a probability of 99.9%, will not be exceeded. We choose the OpVaR from the loss distribution as the $N*(1-99.9\%)$ most extreme loss.
6) Repeat steps 1 to 5 for 30 times
7) Economic Capital is reported as the mean of these 30 runs.

FIRM SPECIFIC COMPONENT:

'Firm Specific Component' employs the internal loss data (ILD) only, and sensitive to firm's own operational loss history. Therefore, this component represents bank's own operational profile and captures the risk component driven by firm specific factors.

Parallel to 'Industry Component', we use NLP and NLMIXED procedures from SAS/OR® software and SAS/STAT software to estimate MLE for the distributions. Following distributions are used to estimate distribution parameters.

Frequency distribution (ILD):	Poisson (λ)
Severity distribution (ILD):	Generalized Pareto (ξ, β)

We fit internal loss severity data to Generalized Pareto Distribution (GPD hereafter). Mean excess plot, Hill plot and Goodness of fit tests are utilized to choose the appropriate threshold level for the GPD distribution. Following section provides detailed description of the GPD fitting and the threshold choice. We compare the parameter estimates using PROC NLP, PROC NLMIXED and NLPxyz/IML procedures. The estimates are approximately same from different procedures. The following code is used to estimate shape (ξ) and scale (β) parameters for the GPD distribution using PROC NLMIXED.

```
PROC NLMIXED data=input (where = (lossAmount GT &threshold));
    parms beta=&beta xi=&epsilon;
    bounds beta >= 0 ;
    if (xi = 0) then    prob = (1/beta)*exp (-(lossAmount -&thres)/beta);
    else   prob = (1/beta)*(1+xi*( lossAmount -&thres)/beta)**((-1/xi)-1);
    LL = log(prob);
    log_loss=log(lossAmount - &thres );
    MODEL log_loss ~ general (LL);
    ODS OUTPUT ParameterEstimates = output_GPD_MLE;
RUN;
```

GPD Distribution and Threshold Analysis:
Following the "Pickands-Balkema-de Haan" theorem; for large threshold "τ", the excess distribution function is approximately distributed as[11];

$$G_{x,b}(x) = \begin{cases} 1-\left(1+\dfrac{x\,x}{b}\right)^{-1/x}, & x \neq 0 \\ 1-e^{-x/b}, & x = 0 \end{cases}$$

with β>0 scale parameter and ξ shape parameter (ξ>0 for heavy tailed).

We implement MLE method to estimate the parameters for the GPD distribution. For this purpose we again use SAS® nonlinear optimization methods such as PROC NLP and PROC NLMIXED. We also validate SAS optimization algorithm with independent codes on other platforms. SAS program utilizes nonlinear optimization method to estimate MLE for the parameters[12]. For validation purpose we use an S-Plus program from QRMLib library that is provided by Alexander J. McNeil, and Matlab program from EVIM package provided by Selcuk et al.[13]

Mean Excess Graph and Hill Plot:
We use graphical approaches (Mean Excess plot and Hill plot) to define the possible range for the threshold choice and to assess the soundness of our threshold decision.
Excess distribution over higher thresholds remains a GPD distribution with the same shape parameter but with a

11 Quantitative Risk Management by McNeil, Alexander J., Frey, Frey , Rudiger and Embrechts, Paul 275-278

12 PROC NLMIXED for Nonlinear Mixed Effect Procedure for models with fixed and random effects. Please refer to Chapter 51 of SAS User Guide for details in optimization algorithms

13 "EVIM: A Software package for extreme value analysis in MATLAB" (Faruk Selcuk, Ramazan Gencay and Abdurrahman Ulugulyagci), Studies in Nonlinear Dynamics and Econometrics, V5(3), 213-239.

scale parameter that is linearly increasing with the threshold level[14]. Thus, the linearity of the mean excess function is used as a diagnostic tool for data admitting a GPD for the excess distribution.

For loss data $X_1,\ldots,X_n$, sample mean excess function is defined as;

$$e_n^{MeanExcess}(x_k)=\frac{\sum_{i=1}^{n}(X_i-X_k)I_{\{X_i>X_k\}}}{\sum_{i=1}^{n}I_{\{X_i>X_k\}}},\qquad 2\le k\le n$$

Figure 4 provides the Mean excess plot i.e. $\{(X_{k,n},e_n(X_{k,n})):2\le k\le n\}$ for all loss observations (small plot), and for the threshold levels ranging from \$5 to \$50[15]. Overall mean excess graph is positively sloped which is consistent with our expectation and suggests a heavy tailed loss distribution. In addition, the relationship between the mean excess value and the threshold level is strongly "linear" for the threshold levels between \$5 and \$50, which is consistent with the GPD distribution assumption.

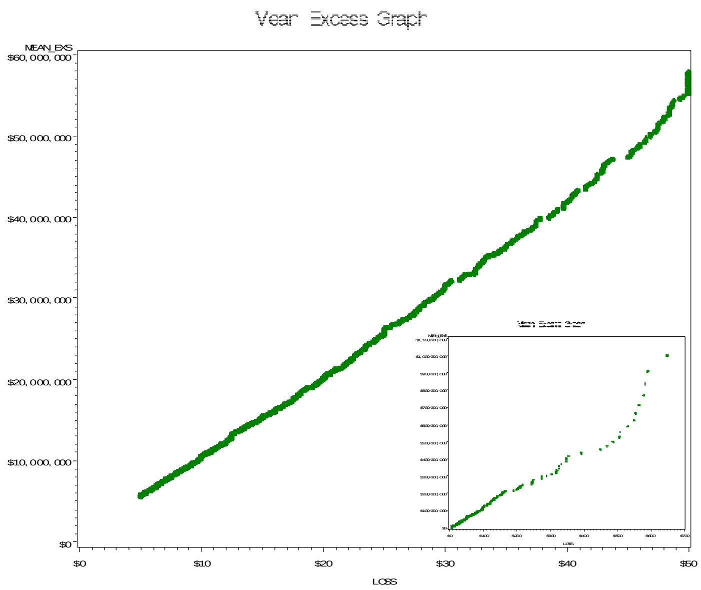


Figure 4. *Mean Excess Graph for Threshold Analysis*

Another tool to assess the threshold choice is the Hill Method. In this method, the tail index is estimated from the inverse of Mean Excess function of log (Loss Amounts), and Hill plot is the inverse Mean Excess plot for the log (Loss Amounts). For n=number of exceedances, the Tail index is estimated as the Hill estimator from the following equation;

$$\hat{\alpha}^{H}_{k,n}=\left(\frac{1}{k}\sum_{j=1}^{k}\left[\ln\left(X_{j,n}\right)\right]-\ln\left(X_{k,n}\right)\right)^{-1},\qquad 2\le k\le n$$

Figure 5 provides the Hill plot for $2\le n\le N$, where N is the sample size along with the upper and lower bounds. Shape estimator Epsilon (Inverse of the Tail index (alpha)) is very unstable up to ~200 highest losses. Inverse of Hill estimator provides the point estimate of the shape parameter for GPD, and x is relatively stable for exceedances ranging from 400 to 600 (small plot).

14 Quantitative Risk Management by Mcneil, Alexdander et al. Lemma 7.22, page 279

15 There values are rescaled for confidentiality.

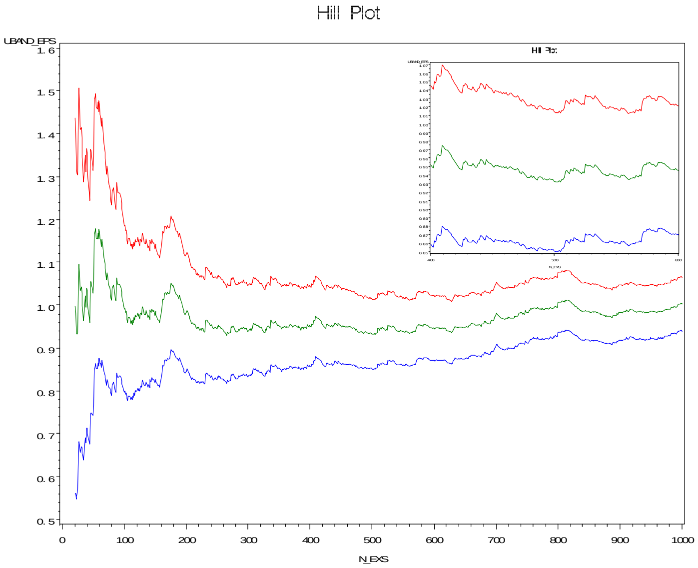


Figure 5. *Hill Plot for Threshold Analysis*

Threshold Choice:

Following the implications of the Mean Excess plot and Hill Method, we focus on $400 \le n \le 600$ range, and use \$5 - \$50 range as possible threshold level. For these threshold levels, we utilize Anderson Darling (AD) and Cramer-von-Mises (CvM) Goodness-of-Fit statistics Goodness-of-Fit statistics to choose the threshold loss level that provide the best fit. For each threshold level with fitted parameters, we follow the study by V. Choulakian and M.A. Stephens for the fit tests[16] and calculate the AD and CvM statistics as follows;

1) For loss order statistics $x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)}$, we make the transformation $z_{(i)} = F\left(x_{(i)}\right)$ using the parameter estimates, to populate the 'z' sample that will be uniformly distributed.
2) Calculate Anderson Darling (AD) and Cramer-von-Mises (CvM) statistics as follows:

$$AD = -n - \left(\frac{1}{n}\right)\sum_{i=1}^{n}(2i-1)\left[\log\left(z_{(i)}\right) + \log\left(1 - z_{(n+1-i)}\right)\right]$$

$$CvM = \sum_{i=1}^{n}\left[\frac{z_{(i)} - (2i-1)}{2n}\right]^2 + \frac{1}{12n}$$

3) Calculate the p values from the table provided by the study[17].

Both AD and CvM Goodness-of-Fit statistics are strongest for \$16 level. This threshold choice is supported by the Mean excess plot providing linear relation; by the Hill plot providing a stable region. Moreover \$16 level provides the best fit-statistics within the range suggested by Mean excess and Hill plot. For levels above ~\$40, Hill plot is unstable.

Estimation and Analysis:

To analyze overtime consistency of the assumptions, we implement 'Firm Specific Component' over different fiscal periods (2005-Q1 to 2007-Q4). Table 4 provides the estimated parameters for the underlying distributions. Parameter estimates are relatively consistent over different periods and inline with the observed loss characteristics.

16 V Choulakian, MA Stephens - Technometrics, 2001 ' Goodness-of-Fit Tests for the Generalized Pareto Distribution'

17 Refer to Table 2 from the study in footnote 12

	Parameters for Firm Specific Component												
ILD	Period	Q1 2005	Q2 2005	Q3 2005	Q4 2005	Q1 2006	Q2 2006	Q3 2006	Q4 2006	Q1 2007	Q2 2007	Q3 2007	Q4 2007
	N	1229	1275	1353	1422	1497	1576	1643	1722	1786	1847	1903	1963
	N excess	383	400	423	438	464	498	520	545	570	590	604	619
	Tail %	31.16%	31.37%	31.26%	30.80%	31.00%	31.60%	31.65%	31.65%	31.91%	31.94%	31.74%	31.53%
	Tail Frequency												
	Poisson λ	9.82	9.52	9.40	9.13	9.10	9.22	9.12	9.08	9.05	8.94	8.75	8.60
	Tail Severity ($)												
	GPD ξ	0.9432	0.9524	0.9766	0.9508	0.9542	0.9506	0.9379	0.9858	0.9775	0.9725	0.9657	0.9809
	GPD β	30858	30168	30183	30666	31031	32462	32462	32305	32082	31510	31919	31857

Table 4. *Firm Specific Component Parameter Estimation* **(ILLUSTRATIVE)**

Monte Carlo Simulation:
Monte Carlo simulation is used to generate process for the annual loss distribution by combining fitted distributions and parameters. We then use 99.9% percentile level to determine the Economic Capital for operational risk at enterprise level. Following code is an example module for the GPD generator in SAS/IML software.

```
START fRand_GPD(beta, epsilon , u);  *Generates N=ncols(u) GPD variables;
                 if -epsilon = 0 then y = log(u);
                 else Ann_Loss_Seq = (beta/epsilon)#(u##(-epsilon) - 1);
                 return(Ann_Loss_Seq);
FINISH fRand_GPD;
```

We test the convergence of the simulation and the convergence is attained above 500,000 annual losses. In Table 7, we report simulation results for 30 runs of 1,000,000 annual losses.
Monte Carlo Simulation is used to aggregate the Frequency and Severity distributions. Using the parameter estimates for Possion and GPD distributions we generate tail loss values for N simulated years. Then we choose the 99.9% VaR from the loss distribution.

Monte Carlo simulation algorithm for Internal Data Component is as follows:

1) For each month, generate loss events above the threshold from Poisson distribution, and calculate annual number of loss events ($Nloss_i$) as the summation of 12 months.

2) For each year, generate $Nloss_i$ losses above the threshold from GPD distribution. We add threshold amount to these losses because GPD generated values are excess values.

3) Calculate loss amount $loss_i$ for that year as the summation of $Nloss_i$ events.

4) Repeat steps 1 to 3 for N years (we use 1,000,000).

5) Operational Value at Risk (OpVaR) is the total annual loss which, with a probability of 99.9%, will not be exceeded. We choose the OpVaR from the aggregate loss distribution as the Nth most extreme loss. We use the proportion of the distribution above the threshold as the non-parametric estimate for the probability of loss being greater than the threshold ($\overline{F}_X(threshold)$)[18,19].

$$N^{th} = 0.001\% * Nloss \Big/ \overline{F}_X(threshold)$$

6) Repeat steps 1 to 6 for 30 times.

7) Economic Capital is reported as the mean of 30 runs.

COMBINING INDUSTRY AND FIRM SPECIFIC COMPONENTS

Both Industry and firm specific components have their own advantages and disadvantages. ‘Industry Component’ utilizes ELD data for the tail distribution which is more stable with longer historical data coverage and provides more robust estimation of ‘tail’ losses. However, ‘Industry Component’ accounts mostly for the industry risk profile, and its ability to capture structural shifts in firm specific risk characteristic is limited. On the other hand, ‘Firm Specific Component’ (with internal data ILD only) addresses this issue by utilizing the firm’s internal loss only which is a better representation of the bank’s own business

18 Operational Risk Modeling, 3.2.4 Estimating Extreme Probabilities by Harry H. Panjer p391.

19 Quantitative Risk Management, Modeling Tails and Measures of Tail Risk by Mcneil, Alexdander et al. equation 7.21, page 283

environment and operational risk profile. But this has a downside as well, which is the limited historical loss data through ILD and highly sensitive tail distribution estimate.

Fiscal Period	Q1 2005	Q2 2005	Q3 2005	Q4 2005	Q1 2006	Q2 2006	Q3 2006	Q4 2006	Q1 2007	Q2 2007	Q3 2007	Q4 2007
Industry Comp.	$150.6	$168.2	$184.1	$156.5	$135.4	$130.7	$134.0	$153.6	$146.6	$130.4	$128.8	$122.8
Firm Spec. Comp	$140.1	$146.4	$175.5	$134.2	$142.8	$151.7	$129.3	$203.9	$186.0	$172.0	$159.1	$179.1

Table 5. *Industry and Firm Specific Components' Economic Capital Estimates* **(ILLUSTRATIVE)**

Table 5 provides EC values estimated from both components for the recent twelve quarters (2005-Q1 to 2007-Q4). Even though the two components represent different risk profiles due to the dataset used for the 'tail' losses, EC estimates provided by both components are comparable up to 2006 – Q4. This observation suggests that the ILD is also a good representation of the 'tail' losses and the bank's risk profile is similar to the industry profile over this period. However, starting 2006 – Q4 we observe difference in ECs which suggests a divergence in risk profile representation by the ILD and the ELD. This deviation is due to an increase in 'Firm Specific Component' as a result of new 'tail' losses observed in ILD, which possibly overstates the tail losses due to limited historical observation.

ΔEC for 2007 - Q4		
Loss Amount	**Firm Spec. Comp.**	**Industry Comp.**
$100 million	35.0%	2.0%
$50 million	31.9%	1.1%
$20 million	25.6%	0.6%
$10 million	20.6%	-
$5 million	15.1%	-
$1 million	6.2%	-
$20 million (x5)	190.6%	-
$1 million (x5)	36.9%	-

Table 6. *Sensitivity to 'Tail' Events*

We did an exercise to assess the sensitivity of the two components due to the robustness of the tail estimates from different datasets. We introduced new loss events ranging from $1million to $100 million to ELD and ILD and computed the change in EC under the new scenarios. Table 9 reports the % change in estimated EC values for the two components. Due to more robust tail estimation from the ELD, 'Industry Component' is less sensitive to new observations in the tail. However, due to limited tail observation for the ILD, 'Firm Specific Component' is highly sensitive to additional tail losses.

Furthermore, as evident from the results when we introduce 5 events with $20 million loss instead of 1 event with $100 loss (or 5 events of $1 million instead of a $5 million), 'Firm Specific Component' is especially sensitive to clustered tail observations.

To merge the two components representing the industry and the bank's risk profiles, we use a weighting scheme to estimate the weighted Economic Capital. To address the sensitivity of the 'Firm Specific Component', we use Mean Excess values for the ILD and ELD data to weight the EC values provided by the two components according to the thickness of the ILD and ELD. Mean excess is similar to expected shortfall or conditional VaR measure. As documented by previous studies, this measure addresses the possible clustering in large losses and it is an useful tool to compare the thickness of tails of distributions on $[0,\infty]$ (e.g. see, Rockafellar and Uryasev 2002, Mikosch 2006). We calculate Mean Excess at CPI adjusted $Tail threshold level for the ELD and ILD and to estimate Weighted Economic Capital (WEC) from Industry Data Component and Internal Data Component. Figure 4 provides the WEC values along with the EC values for the industry and firm components over different periods.

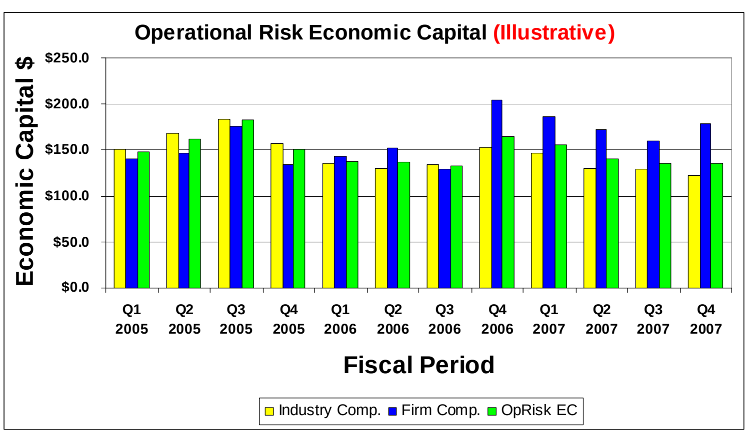


Figure 6. *Operational Risk Economic Capital*

CONCLUSION

In this study, we introduced our LDA methodology to quantify operational risk capital at firm level. Consistent with the Basel II guidelines, the methodology utilizes both internal and relevant external data. The two components capture industry (‘Financial Services’) and firm specific risk profiles. Moreover, in addition to representing the bank’s own business profile through internal data, the methodology estimates the ‘tail’ events from a more robust data by addressing the selection bias in external data.

Model’s EC estimation over the past twelve quarters (2005-Q1 to 2007-Q4) suggests that, two components of the proposed model are consistent and comparable overtime. Consistent with our expectation, ‘Firm Specific Component’ is very sensitive to additional ‘tail’ events due to limited historical observation. Furthermore, our analysis on the additional ‘tail’ losses suggest more robust and stable capital estimate for the ‘Industry Component’ through external data. Since the methodology is implemented at firm level, the model is a first step to address the data issues to quantify the operational risk and to represent risk components. A promising next step is to address the data challenges at business level and to quantify capital allocation across line of businesses.

REFERENCES

Baud, Nicolas, Antoine Frachot and Thierry Roncalli 2002 “Internal Data, External Data and Consortium Data for Operational Risk Measurement: How to Pool Data Properly?”

Brandts, Silke N., "Operational Risk and Insurance: Quantitative and Qualitative Aspects" (April 30, 2004). EFMA 2004 Basel Meetings Paper. Available at SSRN: http://ssrn.com/abstract=493082

De Fontnouvelle, Patrick, De Jesus-Rueff, Virginia, Jordan, John S. and Rosengren, Eric S., "Capital and Risk: New Evidence on Implications of Large Operational Losses" (September 2003). FRB of Boston Working Paper No. 03-5. Available at SSRN: http://ssrn.com/abstract=648164

McNeil, A., Frey, R. and Embrechts, P. (2005) Quantitative Risk Management – Concepts, Techniques and Tools, p 275-278, Princeton: Princeton University Press.

Mikosch, Thomas Non-Life Insurance Mathematics: An Introduction with Stochastic Processes, Chapter 3.2.3 p94, 2nd printing, 2006

Panjer, Harry H. Operational Risk Modeling Analytics, Chapter 3.2.4 p391, Estimating Extreme Probabilities

Ramazan Gençay & Faruk Selçuk & Abdurrahman Ulugülyagci, 2001."EVIM: A Software Package for Extreme Value Analysis in MATLAB," Studies in Nonlinear Dynamics & Econometrics, Berkeley Electronic Press, vol. 5(3), pages 213-239.

Rockafellar, Tyrrell R. and Uryasev, Stanislav P., "Conditional Value-at-Risk for General Loss Distributions" *Journal of Banking and Finance*, Volume 26, Number 7, July 2002 , pp. 1443-1471(29)

Choulakian, V. and Stephens, M. A. Goodness-of-fit tests for the generalized Pareto distribution, *Technometrics* 43(3) (2001) 478-484.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Sabri Guray Uner
PNC Financial Services
249 Fifth Ave. Pittsburgh, PA 15217
412-762-4567
412-768-3251 (fax)
sabri.uner@pnc.com