Finance and Economics Discussion Series
Divisions of Research & Statistics and Monetary Affairs
Federal Reserve Board, Washington, D.C.

Predicting Operational Loss Exposure Using Past Losses

Filippo Curti and Marco Migueis

2016-002

Please cite this paper as:
Curti, Filippo, and Marco Migueis (2016). "Predicting Operational Loss Exposure Using Past Losses," Finance and Economics Discussion Series 2016-002. Washington: Board of Governors of the Federal Reserve System, http://dx.doi.org/10.17016/FEDS.2016.002.

Predicting Operational Loss Exposure Using Past Losses[1]

Filippo Curti[2]

Federal Reserve Bank of Richmond

Marco Migueis[3]

Federal Reserve Board

August 28, 2015

Abstract

Operational risk models, such as the loss distribution approach, frequently use past internal losses to forecast operational loss exposure. However, the ability of past losses to predict exposure, particularly tail exposure, has not been thoroughly examined in the literature. In this paper, we test whether simple metrics derived from past loss experience are predictive of future tail operational loss exposure using quantile regression. We find evidence that past losses are predictive of future exposure, particularly metrics related to loss frequency.

Keywords: operational risk, tail risk, quantile regression
JEL: G21, G28, G32

[1] The opinions expressed in this manuscript belong to the authors and do not represent official positions of the Federal Reserve Bank of Richmond, the Federal Reserve Board, or the Federal Reserve System. The authors thank Robert Stewart for helpful suggestions.

[2] Filippo.curti@rich.frb.org.
[3] Marco.a.migueis@frb.gov.

1 – Introduction

Large, internationally active US bank holding companies (BHCs) are required to calculate their operational risk capital according to the Advanced Measurement Approach (AMA), which relies on banks' internal models to estimate exposure at the 99.9 confidence level.[4] Similarly, large US BHCs are required to estimate operational losses under stressed conditions for the annual Comprehensive Capital Analysis and Review (CCAR) required by the Federal Reserve.[5] To estimate exposure for both exercises, US BHCs frequently rely on the Loss Distribution Approach (LDA), an actuarial modeling framework within which past loss frequencies and severities are used to forecast exposure. A critical assumption of such a framework is that exposure in the past is the same as exposure in the future. In this paper, we test whether the use of past losses really improves the forecasting of exposure. In particular, we use quantile regression to assess whether past losses add value in predicting future tail losses.

The literature includes various papers discussing the factors that predict operational loss exposure (e.g., Chernobai et al. 2011, Cope et al. 2012, Wang and Hsu 2013). However, not much attention has been devoted to the fundamental assumption underlying LDA models: do past losses help predict future exposure, particularly tail exposure?

To assess whether past operational losses add value in predicting operational risk exposure, we performed quantile regressions where the high quantiles of the industry distribution of annual operational losses are forecasted using metrics calculated from past losses and other financial variables. We used a variety of alternative explanatory variables and specifications, including regressions with firm and time fixed effects, and found that loss metrics help forecast tail operational loss exposure. In particular, we found that average loss frequency above $100k is a statistically significant predictor of future operational losses all the way to the 99th quantile of the operational loss distribution. This relation is robust to a variety of specifications. Other loss metrics, such as average total losses, are also predictive of future tail exposure but stop being statistically significant at lower quantiles. Therefore, our results show that past losses are useful in predicting operational loss tail exposure.

Our results also show that firm size, measured through gross income or total assets, is predictive of tail loss exposure. As banks grow, their tail operational losses also grow.

The remainder of this paper is organized as follows: Section 2 discusses the model used for the conditional quantiles of the operational loss distribution; Section 3 describes the data used;

[4] Code of Federal Regulations, Title 12, Federal Reserve System, Part 217, Subpart E.

[5] Board of Governors of the Federal Reserve System, 2014, "Comprehensive Capital Analysis and Review 2015 Summary Instructions and Guidance."

Section 4 describes the quantile regression methodology, presents the main empirical results, and provides multiple robustness checks; finally, Section 5 concludes.

2 – Model for Conditional Quantiles of the Operational Loss Distribution

Historical operational losses are frequently used to estimate operational loss exposure. The logic of such an approach is that past operational losses proxy for the risk profile and risk management of firms and, thus, that if such risk profile and risk management remain stable, the loss profile in the future should be similar to the loss profile in the past. The goal of this study is to assess whether past operational losses do help predict operational loss exposure, particularly tail exposure. To study how the tail of the operational loss distribution behaves, we assume the following specification for the conditional quantiles of the operational loss distribution:

$$OL_{i,t}^{q} = \alpha^{q} + \beta^{q} LossMetrics_{i,t-1} + \gamma^{q} GrossIncome_{i,t-1}$$

where $OL_{i,t}$ are the annual operational losses of bank i in year t, $LossMetrics_{i,t-1}$ are loss metrics from bank i at year t-1, and $GrossIncome_{i,t-1}$ is the three year rolling average of gross income for bank i ending at year t-1.

A variety of loss metrics are considered in this paper, including average total annual losses, the standard deviation of total quarterly losses, and the average annual loss frequency above certain thresholds (such as $100k and $1MM). Average total losses are an obvious loss metric to consider because they are the natural estimate of the first moment of the annual operational loss distribution. Similarly, the standard deviation of total losses is a natural metric to consider, as it reflects the variation of total operational losses; we have chosen to use the standard deviation of quarterly total losses instead of the standard deviation of annual total losses because the quarterly statistic provides a more granular measurement of variation. Finally, we also explore whether the average frequency of losses above certain thresholds is predictive of the tail of the operational loss distribution. While average frequency metrics ignore most of the information concerning loss severity, such metrics may still be more robust predictors of tail operational loss exposure than average losses or the standard deviation of losses because average frequency is more stable than the average or the standard deviation of total losses, as average frequency does not fluctuate significantly when a few tail losses enter a bank's loss data.

The operational risk literature has shown that financial statement variables help predict operational losses. Chernobai et al. (2011) showed that the market value of equity, the return on equity, the tier 1 capital ratio, and other financial measures are predictive of the frequency of operational losses. Similarly, Abdymomunov (2014) showed that operational losses are positively related to total assets for some event types and negatively related for other event types. In this paper, we primarily rely on gross income to proxy for the impact of firm size on operational loss exposure. Gross income is the proxy indicator used in the Basel II standardized approaches for operational risk capital (Basel Committee on Banking Supervision 2006). Perhaps surprisingly, the relationship between gross income and operational losses has not been thoroughly examined in the literature. The regressions of Cope et al. (2011) show that gross income appears to be negatively related to operational loss severity; however, the authors do not emphasize this result in their conclusions. Also, when the Basel II Accord was published, the Basel Committee in Banking Supervision did not provide analysis supporting the use of gross income a proxy for operational risk. For these reasons and given the use of gross income in the Basel framework, we are interested in assessing whether gross income predicts operational loss exposure. As a robustness check, we have also performed the regressions using total assets instead of gross income, and results are qualitatively similar.

3 – Data

The analysis in this paper includes 31 bank holding companies that participated in CCAR 2015 (list provided in Annex 1). Two types of data are used: operational loss event data, obtained from Federal Reserve's Y-14Q regulatory report; and financial statement data, including gross income and total assets, obtained from Federal Reserve's Y-9C regulatory report. Loss information is used from 2000 or as far back as available in the Y-14Q reports.[6] Matching up all available loss data with financial statement data is possible for all institutions in our sample; for most institutions, financial statement data is available for a much longer period than loss data.

Calculating gross income involves two steps: first, for each year we sum item 3 ("Net interest income") and item 5.m ("Total noninterest income") from the schedule HI of the Y-9C report; second, we average over a rolling window of three years, excluding any negative values, as prescribed by the Basel II framework. For example, to calculate the average gross income for 2014, we average the gross income of 2012, 2013, and 2014. According to the Basel II Accord, the goal of averaging gross income over a three year window is to stabilize the resulting capital estimates; we use the three year average gross income because we agree that increasing the stability of this metric is sensible for our purposes and because we want to be consistent with the measure used in the Basel framework. Total assets for a given year are simply the figure reported in item 12 ("Total assets") of the schedule HC of the Y-9C report at the end of the fourth quarter of the year of interest.

Table 1 presents the descriptive statistics of the loss data, gross income and total assets used in the regressions of this paper. Descriptive statistics are presented for ratio of loss metrics to total

[6] Some banks have reported losses before 2000. However, such data is likely incomplete, and so we have opted to initiate our sample at 2000.

assets, instead of simply for the loss metrics, to preserve the confidentiality of banks loss information. Each bank year combination is an independent observation.

Table 1
Descriptive Statistics

Variable	N	Mean	St Dev	10th Prct	Median	90th Prct	Coeff of variation
Annual Losses/Total Assets	211	0.0011	0.0015	0.0002	0.0005	0.0025	1.4286
Avg Annual Losses/Total Assets	211	0.0007	0.0007	0.0001	0.0006	0.0016	0.9128
Std Dev Quarterly Losses/Total Assets	204	0.0004	0.0005	0.0000	0.0002	0.0012	1.2342
10^9*Avg Frequency Above $20k/Total Assets	211	1.6023	1.0385	0.4572	1.4503	2.8291	0.6481
10^9*Avg Frequency Above $100k/Total Assets	211	0.3455	0.2079	0.1283	0.2982	0.6206	0.6017
10^9*Avg Frequency Above $1Mln/Total Assets	211	0.0405	0.0241	0.0158	0.0375	0.0712	0.5951
10^9*Avg Frequency Above $10Mln/Total Assets	211	0.0058	0.0041	0.0012	0.0051	0.0113	0.7025
Gross Income (Billion $)	211	28.58	38.01	3.44	10.38	103.12	1.3300
Total Assets (Billion $)	211	526.65	697.27	63.25	174.17	1,868.35	1.3240

Operational risk is significant for large US BHCs. On an average year, a large US BHC faces losses close to 0.11% of its total assets; while once in ten years BHCs in our sample lose at least 0.25% of their total assets. Given that large US BHCs are highly leveraged, this can represent a significant hit to their capital base. On average, large US BHCs funded 8.8% of their total assets with tier 1 capital in 2014Q3.[7] Thus, once every ten years, a BHC in our sample suffered operational losses that would erode more than 2.8% of their capital base if their tier 1 to assets ratio stood at the industry average.

[7] Board of Governors of the Federal Reserve System, 2015, "Comprehensive Capital Analysis and Review 2015: Assessment Framework and Results."

4 – Regression Methodology and Results

a) Main Regression Results

The conditional quantiles of the industry operational loss distribution are estimated through quantile regression.[8] Losses from all CCAR bank holding companies are pooled in the regressions. Loss metrics at year t-1 are calculated using the loss data reported by banks until year t-1.

Due to large size differences of banks in our sample, the conditional distribution of annual operational losses is likely heteroscedastic. To assess this issue, first, we performed an ordinary least squares (OLS) regression of operational losses on year t on average loss frequency above $100k up to year t-1 and average gross income for years t-3 to t-1; then, we calculated the correlation between absolute values of residuals of this regression and banks total assets – this correlation is 55.8%. Therefore, regression residuals increase significantly in absolute value with banks total assets, which implies that residuals are not identically distributed and, thus, that un-weighted regressions are inefficient.

To increase the efficiency of the estimation procedure, we divide all observations by the bank's total assets at the end of year t-1. Repeating the procedure of calculating an OLS regression on the transformed data, and then calculating the correlation between regression residuals and total assets, we find a correlation of 22.3%. So, for the OLS regression, the normalization we employed significantly diminished heteroscedasticity and, thus, increased regression efficiency. This increase in efficiency is likely to also apply to the quantile regressions we are interested in. So, we divide all observations by banks total assets at the end of year t-1 in the quantile regressions calculated in this paper; the quantile regressions we estimate produce the following conditional quantile fitted values:

$$\left(\frac{\widehat{OL_{\iota,t}}}{Assets_{\iota,t-1}}\right)^{q} = \frac{\widehat{\alpha^{q}}}{Assets_{i,t-1}} + \widehat{\beta^{q}}\,\frac{LossMetrics_{i,t-1}}{Assets_{i,t-1}} + \widehat{\gamma^{q}}\,\frac{GrossIncome_{i,t-1}}{Assets_{i,t-1}}$$

Empirical bootstrapping is used to estimate the confidence intervals of model parameters. We employ the (x,y)-pair bootstrap technique, whereby pairs of observations from the original sample are sampled with replacement and model parameters are re-estimated multiple times. Parameter confidence intervals are calculated according to the percentile method, whereby confidence boundaries correspond to the appropriate percentiles of the distribution of the bootstrapped parameter estimates.[9] Note that when this technique is used, confidence intervals

[8] See Koenker and Basset (1978), Koenker and Hallock (2001), and Koenker (2005) for descriptions of the theory and use of quantile regression.

[9] For further description of how to employ bootstrapping techniques to estimate parameter uncertainty in quantile regressions see Koenker (2005).

generally will not be symmetrical around the coefficient estimate because estimators are not normally distributed in small samples. Our uncertainty estimates are based on 1000 re-samples.

Tables 2 presents the results of regressions where the 90th quantile of operational losses are forecasted using the different loss metrics described in Section 2 jointly with gross income. Tables 4 and 5 provide regression results for the 95th and 99th quantiles, respectively.

Table 2

	Quantile Regressions – 90th Quantile				
Explanatory Variables	(1B)	(2B)	(3B)	(4B)	(5B)
10^9×Avg Frequency Above 100k	0.004** [0.001,0.008]	--	--	--	--
10^9×Avg Frequency Above 1MM	--	0.032*** [0.012,0.047]	--	--	--
10^9×Avg Frequency Above 10MM	--	--	0.112*** [0.059,0.204]	--	--
Avg Annual Losses	--	--	--	1.461** [0.378,1.976]	--
Std Dev Quarterly Losses	--	--	--	--	1.108* [-0.239,2.529]
Gross Income	0.029*** [0.007,0.047]	0.033*** [0.017,0.048]	0.044*** [0.023,0.054}	0.026*** [0.011,0.046]	0.037*** [0.020,0.65]
Q. Reg. Objective Function	0.0673	0.0675	0.0684	0.0680	0.0687
Note: N = 211 for all regressions except (5B). N = 204 for regression (5B). Coefficient 95% confidence intervals in brackets. *** = significant at 1%; ** = significant at 5%; * = significant at 10%.					

Table 3

	Quantile Regressions – 95th Quantile				
Explanatory Variables	(1C)	(2C)	(3C)	(4C)	(5C)
10^9×Avg Frequency Above 100k	0.007** [0.001,0.010]	--	--	--	--
10^9×Avg Frequency Above 1MM	--	0.041*** [0.010,0.096]	--	--	--
10^9×Avg Frequency Above 10MM	--	--	0.120*** [0.016,0.385]	--	--
Avg Annual Losses	--	--	--	1.252 [-0.551,3.474]	--
Std Dev Quarterly Losses	--	--	--	--	-0.337 [-2.038,4.935]
Gross Income	0.029*** [0.009,0.089]	0.038*** [0.016,0.088]	0.070*** [0.029,0.095]	0.039*** [0.011,0.097]	0.095*** [0.021,0.144]
Q. Reg. Objective Function	0.0490	0.0499	0.0504	0.0526	0.0528
Note: N = 211 for all regressions except (5C). N = 204 for regression (5C). Coefficient 95% confidence intervals in brackets. *** = significant at 1%; ** = significant at 5%.					

Table 4

	Quantile Regressions – 99th Quantile				
Explanatory Variables	(1D)	(2D)	(3D)	(4D)	(5D)
10^9×Avg Frequency Above 100k	0.010** [0.000,0.027]	--	--	--	--
10^9×Avg Frequency Above 1MM	--	0.057* [-0.005,0.189]	--	--	--
10^9×Avg Frequency Above 10MM	--	--	0.121 [-0.080,0.564]	--	--
Avg Annual Losses	--	--	--	1.636 [-1.385,11.536]	--
Std Dev Quarterly Losses	--	--	--	--	1.044 [-5.066,4.303]
Gross Income	0.083* [-0.008,0.133]	0.108** [0.015,0.207]	0.156*** [0.034,0.257]	0.108*** [0.034,0.270]	0.151*** [0.079,0.282]
Q. Reg. Objective Function	0.0162	0.0170	0.0177	0.0182	0.0175
Note: N = 211 for all regressions except (5D). N = 204 for regression (5D). Coefficient 95% confidence intervals in brackets. *** = significant at 1%; ** = significant at 5%; * = significant at 10%.					

Simple summary metrics of past losses are predictive of future operational losses, even in the tail of the distribution.[10] Average annual total losses are predictive of future losses until somewhere between the 90th and the 95th quantile. Similarly, the standard deviation of quarterly losses is weakly predictive of future 90th quantile losses, as it is statistically significant at 10%, but it is not predictive of 95th quantile losses and above.

Frequency measures of past losses perform even better than the average and standard deviation of past total losses in predicting the tail of future total losses.[11] Of the different loss metrics considered, the regressions including average loss frequency above $100k result in the lower values of the quantile regression objective function for the 90th, 95th, and 99th quantiles; this implies that, when considered together with gross income, average loss frequency above $100k appears to be the best predictor of future tail losses among the metrics considered. Also, unlike other loss metrics, average frequency above $100k is statistically significant at 5% in predicting operational losses up to the 99th quantile. An additional loss event above $100k per year implies, approximately, a $3.7MM larger 90th quantile, a $7.4MM larger 95th quantile, and a $10.1MM larger 99th quantile of the distribution of total annual operational losses.

The superiority of frequency measures relative to measures based on total losses (such as the average annual total losses or the standard deviation of quarterly total losses) is likely the outcome of frequency measures being more stable proxies for risk exposure, as they do not fluctuate significantly when new tail losses are incurred. The higher stability of frequency metrics is demonstrated by their lower coefficient of variation: the coefficients of variation for average frequency above $100k and for average frequency above $1MM are approximately equal to 60%, while the coefficient of variation of average total annual losses is approximately 91% and the coefficient of variation of the standard deviation of quarterly total losses is approximately 123%. Likewise, frequency metrics based on lower thresholds likely perform better than frequency measures based on higher thresholds – average frequencies above $100k and $1MM perform better than average frequency above $10MM – because average loss frequencies at lower thresholds are more stable metrics of risk exposure. Moreover, by focusing on smaller losses, the lower threshold frequency measures provide more granular information on the exposure of banks, which appears to be connected with tail exposure.

Gross income, in combination with any of the loss metrics considered, is a statistically significant predictor of tail operational losses for all quantiles considered. When used in combination with average loss frequency above $100k, a $100 increase in gross income implies a $2.9 increase of

[10] See Annex 2 for results of the median regression.

[11] Average frequency above $20k was also considered as an explanatory variable. However, it led to inconsistent results, likely due to the lower quality of loss data between $20k and $100k. Banks' collection practices of smaller dollar losses are often inconsistent across the sample period and generally less reliable.

the 90th and 95th quantiles and an $8.3 increase of the 99th quantile of the distribution of total annual operational losses.

In summary, both metrics of past losses and gross income are predictive of future operational tail risk, and the best forecast is obtained when they are combined.

b) Robustness checks

i) Firm and time fixed effects

To assess the robustness of our results, we explore whether the coefficients associated with loss metrics and gross income remain significant when firm heterogeneity and time-specific systematic differences are controlled for. For this purpose, we run quantile regressions with firm and time fixed effects. Similar to the main regressions, we divided both the dependent and the explanatory variables by total assets to decrease the heteroscedasticity of data, and thus the fitted conditional quantiles follow the expression below:

$$\left(\frac{\widehat{OL_{\iota,t}}}{Assets_{\iota,t-1}}\right)^{q} = \frac{\widehat{\alpha_{\iota}^{q}}}{Assets_{i,t-1}} + \frac{\widehat{\tau_{t}^{q}}}{Assets_{i,t-1}} + \widehat{\beta^{q}}\,\frac{LossMetrics_{i,t-1}}{Assets_{i,t-1}} + \widehat{\gamma^{q}}\,\frac{GrossIncome_{i,t-1}}{Assets_{i,t-1}}$$

In most cases, the magnitude of coefficients does not change significantly when fixed effects are introduced. However, their statistical significance generally diminishes.[12] Table 5 presents the results of the fixed effects regressions when average frequency above $100k and gross income are used as explanatory variables for the 90th, 95th, and 99th quantiles.

Table 5

	Quantile Regressions		
Explanatory Variables	90th Qtl	95th Qtl	99th Qtl
10^9×Avg Frequency Above 100k	0.004* [-0.001,0.008]	0.007 [-0.001,0.010]	0.004 [-0.000,0.024]
Gross Income	0.015** [0.001,0.042]	0.012* [-0.000,0.055]	0.046* [-0.000,0.086]
Note: N = 211 for all regressions. Coefficient 95% confidence intervals in brackets. ** = significant at 5%; * = significant at 10%.			

The coefficients of the average frequency above $100k have a similar magnitude in the quantile regressions forecasting the 90th and 95th quantiles of the operational loss distribution when fixed effects are included and when fixed effects are not included; the statistical significance of the

[12] The same empirical bootstrapping technique is used to estimate statistical significance as in the non-fixed effects regressions.

coefficients is reduced, but the coefficient is still significant at 10% in the regression of the 90th quantile and almost statistically significant at 10% in the 95th quantile regression. The coefficient of the average frequency above $100k in the 99th quantile regression has a much smaller magnitude in the fixed effects regression than in the non-fixed effects regression. Nevertheless, and despite the large standard error of the coefficient, the coefficient of the 99th quantile fixed effects regression is still close to significant at 10%.

The reduction of the significance of past losses of the loss metrics in the fixed effects regression indicate that, once firm-specific and year-specific heterogeneity are controlled for, past losses add less value in predicting future losses. A possible explanation for the reduced significance of past losses is that the risk profile and risk management characteristics that past losses proxy do not fluctuate significantly with the fluctuation of the simple loss metrics we considered; and thus, in a fixed effects regression where the coefficient of past losses is estimated from within-firm variation, the effect of the firm risk profile and risk management characteristics, proxied by our simple loss metrics, on operational losses cannot be distinguished from firm fixed effects.

While these results question the robustness of the results of the previous regressions, they do not imply that past losses are not useful metrics to consider. What drives firm-specific effects and how they evolve through time is unclear, but past losses appear to be good proxies for them.

ii) Accounting for risk management measures

We also investigated whether the predictive ability of past losses is robust to the inclusion of risk management metrics. Chernobai et al. (2011) and Wang and Hsu (2013) have shown that governance and risk management measures are predictive of operational losses. Replicating their methodology is beyond the scope of this paper; nevertheless, we are interested in assessing whether the risk management metrics collected by the Federal Reserve add value in predicting operational losses.

We focus our analysis on the Federal Reserve's composite risk management rating for BHCs and its subcomponents: board and senior management oversight; policies, procedures and limits; risk monitoring and management information systems; and internal Controls. The ratings reflect the effectiveness of the banking organization's risk management and controls. The values of this metrics vary between one and five – one being the best rating and five the worst rating – and are based on the evaluations of Federal Reserve examiners. Table 6 presents the descriptive statistics of the risk management metrics considered in our robustness test.

Table 6
Descriptive Statistics

Rating	N	Mean	St Dev	10^{th} Prct	Median	90^{th} Prct	Coeff of variation
Composite	191	2.31	0.56	2.00	2.00	3.00	0.24
Board and Senior Management Oversight	191	2.25	0.50	2.00	2.00	3.00	0.22
Policies, Procedures and Limits	191	2.25	0.46	2.00	2.00	3.00	0.20
Risk Monitoring and Management Information Systems	191	2.47	0.56	2.00	2.00	3.00	0.23
Internal Controls	191	2.25	0.46	2.00	2.00	3.00	0.20

None of the metrics proved predictive of future losses, and their inclusion in the forecasting regressions did not change substantially the magnitude of the coefficients of past loss metrics. As an example, we present the regression results when the Federal Reserve measures of risk management are combined with average loss frequency above \$100k and gross income to predict the 95^{th} quantiles of the annual operational loss distribution. To ensure that the risk management metrics are in a similar scale to the dependent variable, the metrics are scaled by total assets, and so the estimated quantile function is

$$\left(\frac{\widehat{OL_{i,t}}}{Assets_{i,t-1}}\right)^q = \frac{\widehat{\alpha^q}}{Assets_{i,t-1}} + \widehat{\theta^q} RiskMng_{i,t-1} + \widehat{\beta^q}\frac{AvgFreqAbv100k_{i,t-1}}{Assets_{i,t-1}} + \widehat{\gamma^q}\frac{GrossIncome_{i,t-1}}{Assets_{i,t-1}}$$

Table 7 presents the regression results for the 95^{th} quantile of the annual operational loss distribution.

Table 7

	Quantile Regressions				
Explanatory Variables	95^{th} Qtl	95^{th} Qtl	95^{th} Qtl	95^{th} Qtl	95^{th} Qtl
Composite	-0.003 [-0.077,0.139]	--	--	--	--
Board and Senior Management Oversight	--	-0.055 [-0.103,0.015]	--	--	--
Policies, Procedures and Limits	--	--	-0.007 [-0.073,0.113]	--	--
Risk Monitoring and Management Information Systems	--	--	--	-0.038 [-0.104,0.029]	--
Internal Controls	--	--	--	--	-0.013 [-0.084,0.157]
10^9×Avg Frequency Above 100k	0.007** [0.001,0.010]	0.006** [0.001,0.011]	0.005** [0.001,0.010]	0.007** [0.001,0.009]	0.007** [0.000,0.011]
Gross Income	0.033** [0.007,0.090]	0.036*** [0.008,0.096]	0.050*** [0.009,0.090]	0.039*** [0.010,0.093]	0.029*** [0.008,0.090]
Note: N = 199 for all regressions. Coefficient standard errors in parenthesis. *** = significant at 1%; ** = significant at 5%; * = significant at 10%.					

The Federal Reserve risk management metrics do not have a statistically significant relation with future operational losses; at the same time, the magnitude and statistical significance of the coefficient of the frequency of losses above $100k does is not substantially change in comparison to the regressions that did not include risk management metrics. So, the predictive ability of loss frequency is robust to the inclusion of the Federal Reserve's risk management metrics.

iii) Using total assets instead of gross income

Total assets are highly correlated with gross income (ρ = 95.9%), and thus using one or the other in the quantile regressions results in qualitatively similar results. Table 8 presents the regressions results when average frequency above $100k and total assets are used to predict the 90^{th}, 95^{th}, and 99^{th} quantiles of the annual operational loss distribution.

Table 8

	Quantile Regressions		
Explanatory Variables	90^{th} Qtl	95^{th} Qtl	99^{th} Qtl
10^9×Avg Frequency Above 100k	0.004*** [0.001,0.008]	0.007** [0.000,0.013]	0.013* (-0.002,0.020]
Total Assets	0.001*** [0.000,0.003]	0.002** [0.000,0.006]	0.005** [0.000,0.011]
Note: N = 211 for all regressions. Coefficient 95% confidence intervals in brackets. *** = significant at 10%; ** = significant at 5%; * = significant at 10%.			

Total assets are statistically significant in predicting tail quantiles of the operational loss distribution. Once average loss frequency above \$100k is controlled for, a \$1000 increase in total assets results in a \$1.4 increase in the 90^{th} quantile, a \$2 increase in the 95^{th} quantile, and a \$4.9 increase in the 99^{th} quantile of the annual operational loss distribution. The magnitude and statistical significance of the coefficients associated with average frequency above \$100k are not significantly affected by using total assets instead of gross income. In what concerns tail operational risk, total assets and gross are similarly informative measures of bank size.

iv) Block bootstrapping

To assess the statistical robustness of our results we used block bootstrapping, a technique that addresses dependence within clusters of observations by bootstrapping clusters of observations as a block, rather than independently (Cameron et al. 2008). In our implementation of the block bootstrap procedure, in each replication we sampled with replacement from the 31 BHCs for which we have data, until the bootstrap sample included data for 31 firms. Given that the number of years of data available varies across firms, our final bootstrap samples have a variable number of observations but generally close to our total sample size of 211.

Coefficient standard errors and confidence intervals do not change significantly when observations from a firm are treated as a block and drawn together in the resampling procedure. Table 9 compares the (x,y)-pair bootstrapping standard errors and confidence intervals for the estimates of the coefficient of average loss frequency above \$100k in the quantile regressions with the statistics obtained using block bootstrapping.

Table 9

	Quantile Regressions		
10^9×Avg Frequency Above 100k	90^{th} Qtl	95^{th} Qtl	99^{th} Qtl
Coefficient Estimate	0.0037	0.0074	0.0101
	(X,Y)-Pair Bootstrapping		
Lower Bound of 95% Conf Interval	0.0013	0.0013	0.0003
Upper Bound of 95% Conf Interval	0.0079	0.0095	0.0272
	Block Bootstrapping		
Lower Bound of 95% Conf Interval	0.0005	0.0016	-0.0002
Upper Bound of 95% Conf Interval	0.0079	0.0087	0.0268
Note: N = 211 for all regressions.			

Confidence intervals grow a little wider for 90^{th} and the 99^{th} quantile regressions, but shrink a bit for the 95^{th} quantile regression. Overall, the change is small. Therefore, accounting dependence between observations of the same firm does not significantly affect the uncertainty of estimates in this sample.

An additional robustness analysis around the simultaneous use of multiple loss metrics in forecasting the tail of the operational loss distribution is presented in Annex 3.

5 – Conclusion

Operational risk practitioners have typically relied on past operational losses to model the distribution of future operational losses. In this paper we provide evidence that past operational losses are a useful metric to predict operational loss exposure, including tail exposure. Metrics associated with loss frequency prove the most robust in forecasting exposure, likely because they are more stable proxies for risk exposure as they do not fluctuate significantly when new tail losses are incurred. However, financial regulators should interpret these findings with caution. Using frequency metrics to measure exposure can lead to undesirable incentives, as breaking loss events into smaller loss events or aggregating them into larger loss events would have capital implications.

Unsurprisingly, the analysis shows that loss metrics are statistically stronger predictors of operational loss exposure at lower quantiles. Thus, these results suggest that setting operational risk capital requirements at a lower confidence level than 99.9% - and then scaling up capital requirements to maintain conservatism – may be a desirable change to capital framework, to increase estimation stability and accuracy.

Also, our analysis shows that once firm-specific and year-specific effects are accounted for, past loss metrics lose statistical significance. This finding raises the possibility that other firm and year specific characteristics may better predict operational loss exposure. However, this finding does not deny the usefulness of operational losses in predicting exposure, when other better measures of firm-specific exposure are not available.

Finally, our analysis shows a robust link between firm size, measured either through gross income or total assets, and operational loss exposure. This suggests that a risk sensitive capital framework for operational risk should account for firm size.

<u>**References**</u>

Abdymomunov, Azamat, 2014, "Banking Sector Operational Losses and Macroeconomic Environment," Working Paper, Federal Reserve Bank of Richmond.

Basel Committee on Banking Supervision, 2006, "International Convergence of Capital Measurement, A Revised Framework."

Cameron, Colin, Jonah Gelbach, and Douglas Miller, 2008, "Bootstrap-based Improvements for Inference with Clustered Errors," *The Review of Economic Studies*, Vol. 90, 414-427.

Chernobai, Anna, Philippe Jorion, and Fan Yu, 2011, "The Determinants of Operational Risk in U.S. Financial Institutions," *Journal of Financial and Quantitative Analysis*, Vol. 46, 1683-1725.

Cope, Eric, Mark Piche, and John Walter, 2012, "Macroenvironmental determinants of operational loss severity," *Journal of Banking & Finance*, Vol. 36, 1362-1380.

Koenker, Roger, 2005, *Quantile Regression*, Econometric Society Monographs, Cambridge University Press.

Koenker, Roger, and Gilbert Bassett, Jr., 1978, "Regression Quantiles," *Econometrica*, Vol. 46, 33-50.

Koenker, Roger, and Kevin Hallock, 2001, "Quantile Regression," *Journal of Economic Perspectives*, Vol. 15, 143-156.

Wang, Tawei, and Carol Hsu, 2013, "Board Composition and operational risk events of financial institutions," *Journal of Banking & Finance*, Vol. 37, 2042-2051.

Annex 1

List of bank holding companies included in the analysis:

Ally Financial Inc.
American Express Company
Bank of America Corporation
The Bank of New York Mellon Corporation
BB&T Corporation
BBVA Compass Bancshares, Inc.
BMO Financial Corp.
Capital One Financial Corporation
Citigroup Inc.
Citizens Financial Group, Inc.
Comerica Incorporated
Deutsche Bank Trust Corporation
Discover Financial Services
Fifth Third Bancorp.
The Goldman Sachs Group, Inc.
HSBC North America Holdings Inc.
Huntington Bancshares Incorporated
JPMorgan Chase & Co.
KeyCorp
M&T Bank Corporation
Morgan Stanley
MUFG Americas Holdings Corporation
Northern Trust Corporation
The PNC Financial Services Group, Inc.
Regions Financial Corporation
Santander Holdings USA, Inc.
State Street Corporation
SunTrust Banks, Inc.
U.S. Bancorp.
Wells Fargo & Co.
Zions Bancorporation

Annex 2

Results of median regression

Tables A1 presents the results of regressions where the median of the operational loss distribution is predicted using the different loss metrics described in Section 2 together with gross income.

Table A1

	Quantile Regressions – 50th Quantile				
Explanatory Variables	(1A)	(2A)	(3A)	(4A)	(5A)
10^9×Avg Frequency Above 100k	0.001*** [0.000,0.002]	--	--	--	--
10^9×Avg Frequency Above 1MM	--	0.012*** [0.06,0.016]	--	--	--
10^9×Avg Frequency Above 10MM	--	--	0.043*** [0.022,0.067]	--	--
Avg Annual Losses	--	--	--	0.386*** [0.232,0.613]	--
Std Dev Quarterly Losses	--	--	--	--	0.426*** [0.219,0.833]
Gross Income	0.007*** [0.004,0.012]	0.006*** [0.003,0.011]	0.009*** [0.006,0.013]	0.007*** [0.005,0.010]	0.009*** [0.006,0.013]
Q. Reg. Objective Function	0.0741	0.0729	0.0749	0.0730	0.0721
Note: N = 211 for all regressions except (5A). N = 204 for regression (5A). Coefficient 95% confidence intervals in brackets. *** = significant at 1%; ** = significant at 5%.					

The metrics of loss experience considered are statistically significant predictors of the median of the annual operational loss distribution. In the case of the median, the regression including average frequency above \$1MM appears to perform the best; nevertheless, the regression specifications considered have similar predictive power.[13]

[13] The objective function of the quantile regressions assumes the lowest value for the quantile regression including the standard deviation of quarterly operational losses. However, this regression has a smaller sample than the other regressions, and so results are not comparable. When the regressions using other explanatory variables are restricted to the sample for which the standard deviation of quarterly losses is available, the regression including the standard deviation of quarterly losses performs the worst.

Gross income is also always statistically significant in predicting median losses; its coefficient varies between 0.006 and 0.009.

Annex 3

Using multiple loss metrics

In the main regressions of this paper, only one loss metric is used at a time. We have followed this approach because the loss metrics considered are highly correlated, and so it is not possible to robustly use them in combination to forecast future losses. Table A2 presents the linear correlation between the loss metrics used in this paper.

Table A2

Correlation Matrix	Avg Frequency Above 100k	Avg Frequency Above 1MM	Avg Frequency Above 10MM	Avg Annual Losses	Std Dev Quarterly Losses
Avg Frequency Above 100k	1	--	--	--	--
Avg Frequency Above 1MM	0.9579	1	--	--	--
Avg Frequency Above 10MM	0.8993	0.9706	1	--	--
Avg Annual Losses	0.8949	0.9148	0.9075	1	--
Std Dev Quarterly Losses	0.8378	0.8289	0.8180	0.9691	1
Note: N = 211 for all correlation coefficients except those including the standard deviation of quarterly losses. N = 204 for the correlation coefficients including the standard deviation of quarterly losses.					

To demonstrate how combining multiple loss metrics leads to non-robust and inconsistent results, we present two examples in Tables A3 and A4. Table A3 presents regression results when loss frequency above $100k, average annual losses and gross income are combined to forecast the 90th, 95th, and 99th quantiles of the annual operational loss distribution; while

Table A3

	Quantile Regressions		
Explanatory Variables	90^{th} Qtl	95^{th} Qtl	99^{th} Qtl
10^9×Avg Frequency Above 100k	0.002* [-0.000,0.007]	0.007* [-0.000,0.011]	0.009 [-0.004,0.030]
Avg Annual Losses	0.629 [-0.277,1.704]	0.470 [-1.576,1.732]	-1.873 [-4.923,4.710]
Gross Income	0.023** [0.002,0.043]	0.023** [0.005,0.090]	0.071* [-0.003,0.172]
Note: N = 211 for all regressions. Coefficient 95% confidence intervals in brackets. *** = significant at 1%; ** = significant at 5%; * = significant at 10%.			

When frequency above $100k is combined with average annual losses in forecasting quantiles of the operational loss distribution, unsurprisingly, the magnitudes of the coefficients of both metrics are reduced in comparison to when the metrics are used separately. But the coefficient estimates also become more uncertain, particularly the coefficient estimates associated with average annual losses. In the 99^{th} quantile regression, the coefficient associated with average annual losses even becomes negative, although not statistically significant. Thus, combining loss frequency above $100k with average annual losses does not appear to be a promising approach to predict tail operational loss exposure.

Table A4 presents the regression results for the same quantiles when average annual losses are combined with the standard deviation of quarterly losses and gross income.

Table A4

	Quantile Regressions		
Explanatory Variables	90^{th} Qtl	95^{th} Qtl	99^{th} Qtl
Avg Annual Losses	2.502*** [0.654,5.750]	3.286** [0.317,10.426]	0.327 [-1.258,24.652]
Std Dev Quarterly Losses	-2.135 [-5.782,0.175]	-3.056 [-10.877,0.866]	0.729 [-26.276,7.346]
Gross Income	0.028** [0.003,0.044]	0.039** [0.004,0.095]	0.147* [-0.013,0.207]
Note: N = 204 for all regressions. Coefficient 95% confidence intervals in brackets. *** = significant at 1%; ** = significant at 5%; * = significant at 10%.			

Similar to when average frequency above $100k and annual average losses are combined to forecast quantiles of the operational loss distribution, when annual average losses and the standard deviation of quarterly losses are combined the uncertainty associated with coefficient estimates increases. In this case, the coefficient estimates associated with average annual losses increase substantially in the regressions of the 90th and 95th quantiles, while the coefficient estimates with the standard deviation of quarterly losses are negative (but not significant). These results are likely spurious and driven by the very high correlation between average annual losses and the standard deviation of quarterly losses (96.9%). Again, we believe these results demonstrate that combining multiple loss metrics in forecasting tail exposure produces non-robust results due to the small sample available and the high correlation between loss metrics.