CIRPÉE
Centre interuniversitaire sur le risque, les politiques économiques et l'emploi

Cahier de recherche/Working Paper **07-02**

Scaling Models for the Severity and Frequency of External Operational Loss Data

Hela Dahen
Georges Dionne

Janvier/January 2007

Dahen: Department of Finance and Canada Research Chair in Risk Management, HEC Montreal, Canada
hela.dahen@hec.ca
Dionne: Department of Finance and Canada Research Chair in Risk Management, HEC Montreal, Canada, CREF and CIRPÉE
georges.dionne@hec.ca

Please send all correspondence to:
HEC Montréal, Canada Research Chair in Risk Management, 3000, Chemin de la Côte-Sainte-Catherine, Montreal, Quebec, Canada H3T 2A7
(Phone): +514 340 6596
(Fax): +514 340 5019

We thank F. Bellavance, S. Christoffersen and B. Rémillard for their helpful comments and recommendations. We gratefully acknowledge financial support from CREF and IFM2.

Abstract:

According to Basel II criteria, the use of external data is absolutely indispensable to the implementation of an advanced method for calculating operational capital. This article investigates how the severity and frequencies of external losses are scaled for integration with internal data. We set up an initial model designed to explain the loss severity. This model takes into account firm size, location, and business lines as well as risk types. It also shows how to calculate the internal loss equivalent to an external loss, which might occur in a given bank. OLS estimation results show that the above variables have significant power in explaining the loss amount. They are used to develop a normalization formula.

A second model based on external data is developed to scale the frequency of losses over a given period. Two regression models are analyzed: the truncated Poisson model and the truncated negative binomial model. Variables estimating the size and geographical distribution of the banks' activities have been introduced as explanatory variables. The results show that the negative binomial distribution outperforms the Poisson distribution. The scaling is done by calculating the parameters of the selected distribution based on the estimated coefficients and the variables related to a given bank. Frequency of losses of more than $1 million are generated on a specific horizon.

Keywords: Operational risk in banks, scaling, severity distribution, frequency distribution, truncated count data regression models.

Résumé:

L'utilisation de données externes constitue une condition sine qua non dans l'implantation d'une méthode avancée de calcul de capital opérationnel, d'après les critères de Bâle II. Cet article vient répondre à plusieurs interrogations sur la mise à l'échelle des montants et des fréquences des pertes externes; et ce dans le but de les intégrer avec les données internes. Cette étude met en place un premier modèle explicatif des montants de pertes incluant la taille de l'entreprise, le lieu de la perte, les lignes d'affaires ainsi que les types de risque. Elle montre comment il est possible de calculer la perte interne équivalente à une perte externe, qui pourrait être subie au niveau d'une banque donnée. Les résultats de l'estimation par MCO montrent que ces variables ont un pouvoir significatif dans l'explication des montants de pertes. Ces dernières sont retenues pour le développement d'une formule de normalisation.

Un deuxième modèle est développé pour la mise à l'échelle des fréquences des pertes qui pourraient avoir lieu durant une période déterminée à partir des données externes. Nous considérons deux modèles de régression, à savoir le modèle Poisson tronqué et le modèle binomial négatif tronqué avec composante de régression. Des variables estimant la taille et la répartition géographique des activités des banques ont été introduites comme variables explicatives dans le modèle. Les résultats montrent que la distribution binomiale négative domine la distribution Poisson. Ainsi, la mise à l'échelle se fait en calculant les paramètres de la distribution retenue à partir des coefficients estimés et des variables propres à une banque donnée. Il est

donc possible de générer des fréquences de pertes de plus de 1 million de dollars sur un horizon déterminé.

Mots Clés: Risque opérationnel des banques, mise à l'échelle, distribution de sévérité, distribution des fréquences, modèles de comptage tronqué avec composante de régression.

JEL Classification: G21, G28, C30, C35.

1. Introduction

Over the recent years, there is an increasing interest from financial institutions to identify losses associated with operational risk. This is due to regulatory considerations according to Basel II accord and also due to the occurrence of huge operational losses recently. We can mention two examples of enormous operational losses sustained by the financial sector: $2.4 billion lawsuit CIBC sustained by the shareholders of Enron and a $690 million loss caused by a rogue trading activities at Allied Irish Banks. Add to these the case of Barings, the UK's oldest bank; it went bankrupt following a rogue trading activities too occasioning a loss of $1.3 billion. These examples show the scope of this risk. They also serve as an imperative warning signal to financial institutions, which must define, measure, and manage this risk. Besides the huge losses it can cause, operational risk also threatens all the activities and operations of an institution. Operational events can be linked to people, processes, systems, and external events. However, operational risk has a varying degree of impact on all units within the institution. Given its scope and complexity, the management of operational risk has become a necessity.

Aware of its disruptive potential, regulatory authorities, in June 1999, opened a debate on the development of a management framework attuned to operational risk. Such a framework would, among other things, provide for an operational loss identification and measurement of operational regulatory capital. They seek to improve on the existing rules by aligning regulatory capital requirements more closely to the underlying risks that banks face. In other words, they ensure that there is sufficient capital to cover the unexpected losses.

Research in this field is still in its embryonic stage. Hence, any and all developments will be of great use in helping financial institutions meet their short-term demands and they will also benefit other industries in their pursuit of medium-term goals. In the Basel II agreement, one of the approaches proposed for quantifying operational risk capital is the advanced method. The development of such a method requires a large database. The data can be drawn from different sources. Internal data are very useful in reflecting the real level of exposure to operational risk. Ideally, they should be the only source of statistical information. However, most bank's internal

data collection process are still in its infancy stage, and there is not enough data especially those rare ,[1] high impact losses to estimate the unexpected loss. .

Recourse to external operational loss data is therefore essential in order to supplement internal data, especially those tail events, which are generally missing from internal data. It is thus justifiable to combine these possible severe losses with the losses of the internal database of a bank, so as to reduce their "surprise" effect (the unexpected) and to calculate adequate operational risk capital. Obviously, we cannot predict the exact amount of extreme losses, which have not yet occurred. However, based on the losses recorded in the banking sector, we can make a projection for a particular bank, if our scaling takes certain factors into account.

Given this context and the fact that an external database is needed for the calculation of an operational risk capital with an advanced approach, the objective of this paper is to develop a robust method that can use external data to predict the severity as well as the frequency of losses, which a bank is exposed to. Several factors will be considered in explaining the number of losses and their severities for a given period. It is thus a matter of projecting the external losses having occurred in the industry to the level of just one bank.

The method developed in this article has been tested on the data of an external base containing operational losses of more than $1 million. However, this method is applicable to any external database. A combination of external losses scaled with the internal data of a particular bank makes it possible to measure that bank's exposure to operational risk.

This article is organized as follows. The second section describes the different approaches used to measure capital; gives the sources and characteristics of the data; and takes a brief look at the scaling methods reported in the literature. A description of the data in the external base is presented in the third section. The fourth section states the model's hypotheses. The model to be used in scaling the severity of losses is then developed in the fifth section. The next to last section develops the model for scaling frequencies. Finally, the study ends with a conclusion and a discussion of possible avenues of further research.

[1] A rare loss is defined as one which results from a highly unlikely event.

2. Context

2.1 Regulatory framework

In 2001, the Basel Committee defined operational risk as being the risk of loss resulting from inadequate or failed internal processes, people and systems or from external events. Legal risk is also included, but the definition does not take into account strategic and reputational risk. With the Basel II accord, there is now a requirement for an amount of additional regulatory capital to cover operational risk. Regulatory authorities have identified three different methods of calculating this capital. The most advanced of these three methods shows greater sensitivity in its detection of risk. In this article, we use the advanced approach in dealing with our research problem.

This more advanced and sophisticated approach relies on internal capital-calculation procedures adopted by banks. Regulatory authorities are very flexible concerning the method chosen, provided it combines adequately qualitative and quantitative criteria (internal data, relevant external data, scenario analysis, and business environment and internal control factors). The method selected must reflect the financial institution's level of exposure to operational risk and must measure the unexpected loss correctly.

Three options are proposed under the advanced approach: (1) the internal measurement approach (*IMA*); (2) the loss distribution approach (*LDA*); and (3) the Scorecard approach. In this study we focus on the loss distribution approach (*LDA*), which estimates unexpected loss or the operational value at risk by modeling the amounts and frequencies of operational losses. The correct combination of internal and external loss data is thus an important step to be considered in an advanced approach.

2.2 Sources of external data and their potential biases

Once scaled, external data can be combined with internal data to generate a database representing a particular bank's risk profile. This plays a part in implementing the advanced approach. The sources of external data are still quite limited. But we can cite at least three:

- Public data obtained from reports in the media and magazines on losses of over $1 million. There are two bases of external data on the market (such as Fitch). The problem with this type of data is that the base only contains very severe losses having occurred in large financial institutions. Recourse to this type of base will not make up for the scarcity of data on certain types of risk (business disruptions and system failures), but it does supplement the base with data on extreme losses, which occur rarely. Such losses will form distribution tails, since the internal data of most financial institutions contain no historical record of large losses, which might occur. When combining internal and external data, special treatment is needed to correct data-linked biases.

- Data provided by insurance brokers (such as Willis, Aon and Marsh) have to do with losses claimed by financial institutions. The major advantage of this source is its reliability. Since data of this sort are collected directly from financial institutions, there is minimum selection bias. However, this source has the disadvantage of containing different collection thresholds, sometimes unobservable, which hinge on variations in insurance policy deductibles. The second limitation of this source resides in the specific nature of the types of risk collected. In fact, only insurable losses will be included in this base.

- Non-public data obtained by compiling internal data from banks, which have agreed to share their information, thus constituting a consortium, like ORX (Operational Riskdata Exchange Association). However, given the confidentiality of the information shared, only the statistics and analyses pertaining to the losses are available to participants. The advantage of this source of data is its reliability. The collection threshold is much lower than that of the preceding sources. This makes the loss amounts more comparable, especially when the member banks are almost of the same size. However, the major disadvantage of this source of data is that it does not allow event-by-event access to the losses. Therefore, these data can not be used to construct a base combining internal and external data.

External data contain many biases, such as:

- Selection bias: Only very large losses are published. This bias is linked to the nature of the databases available and is thus difficult to correct.

- Control bias: Losses come from banks with different control environments. There are unfortunately no variables capable of estimating quality control for the banks found in external bases. So, it is not possible to correct this bias with the information available.

- Collection bias: When data are drawn from different sources, variations in thresholds may cause biases. Frachot and Roncalli (2002) and Baud, Frachot, and Roncalli (2002) describe how internal data can be compared with external data having different collection thresholds.

- Scale bias: Losses come from banks of different sizes (assets, revenues, number of employees…) located in different countries. Our research is concerned with correcting this bias.

2.3 Literature review

Little research has been done to find a solution to the scale problem. Shih, Samad-Khan, and Medapa (2000) have introduced the institution's size as the main scaling factor. They have shown that the relation between operational losses and firm size is non-linear. In effect, the relation between the logarithm of the scale factor and the loss amounts is stronger than the one between losses and the gross scale variable. Besides, a bank twice as large as another will not, on average, suffer losses two times higher than those of the smaller bank. Shih, Samad-Kahn, and Medapa (2000) effectively suppose the relation to be as follows:

$$L = R^{\alpha} \times F(\theta)$$

where:

L: loss amounts;

R: total revenue of the firm where the loss occurred;

α : a scaling factor;

θ: a vector representing all the risk factors not explained by R. $F(\theta)$ is thus a multiplying residual term which is not explained by fluctuations in revenue.

Taking the logarithm of this equation, we obtain a linear relation. It is thus possible to estimate the scaling factor α and the logarithm of the function of the other factors $F(\theta)$ which constitute the regression's constant term.

Total revenue is the only risk factor included in the model, which estimates firm size. Most of the variability in losses is thus probably caused by other factors such as type of business line, quality of management, and effectiveness of the control environment. In this same study, it has been shown that size explains only a small portion (about 5%) of the loss amounts.

Along the same lines, Hartung (2004) has developed a normalization formula, which makes it possible to calculate the equivalent of an external loss for a given bank. The formula used is

$$Loss_{adj} = loss_{org}\left\{1 + a\left[\left(\frac{Scal.Param\left(Loss_{adj}\right)}{Scal.Param\left(Loss_{org}\right)}\right)^{b} - 1\right]\right\}$$

where:

$Loss_{adj}$: the loss amount adjusted for a given bank;

$Loss_{org}$: the original loss amount for a reference bank;

$Scal.Param\left(Loss_{adj}\right)$: a scaling parameter for a given bank;

$Scal.Param\left(Loss_{org}\right)$: a scaling parameter for a reference bank;

a, b: adjustment factors such that $a \in [-1;1], b \in [0;1]$.

The scaling parameter was assigned based on the cause of the event. Examples of this parameter are revenues, number of employees or quality of risk management. Hypotheses have been formulated concerning the value of the adjustment factors in relation to the scaling parameter. This scaling model's limitations consist, on the one hand, in the lack of any theoretical justification of the formula used and, on the other hand, in the absence of a suitable method for estimating the adjustment factors.

According to the study by Na (2004), the loss amount can be broken down into a common component and an idiosyncratic component. The component common to all the banks or business lines captures all the changes in the macroeconomic, geopolitical, and cultural environment, whereas the idiosyncratic component covers all the factors specifically linked to the line of business or the loss event. A power relationship has been inserted between this last component and a size estimator. A normalization formula has been developed to find, for lines of

business B1 and B2 of a given bank, the equivalent loss amount of a loss taken as a reference point. The formula is as follows:

$$\frac{L_{T,B1}}{\left(R_{idiosyncratic}\right)_{T,B1}^{\lambda}} = \frac{L_{T,B2}}{\left(R_{idiosyncratic}\right)_{T,B2}^{\lambda}}$$

$L_{T,Bi}$: a loss amount having occurred at date *T* at the bank or in the business line *Bi*;

$\left(R_{idiosyncratic}\right)_{T,Bi}$: the revenue of the bank or the business line at date *T*, constituting the only estimator of the idiosyncratic component;

λ : a scaling factor.

This model can be improved by introducing scaling factors other than firm size. Our model takes into consideration size, location, business line, and risk type.

Once severity has been scaled, it is also important to determine the frequency with which normalized losses occur on a particular time horizon. Very few studies have attempted this type of scaling. Some studies have indeed developed normalization models for severity but without considering any scaling for frequencies (Shih, Samad-Khan, and Medapa, 2001; Hartung, 2004). Hartung (2004) groups the frequency of losses in four banks along a nine-year horizon. The bank used as a reference will have a distribution identical to that of the four banks grouped together. These banks are not necessarily comparable. Several factors enter into play when determining the frequency of losses.

Na (2004) has developed a model for scaling frequencies which is equivalent to the one used to scale severity. This model stipulates that the frequency of losses can be broken down into a common and idiosyncratic component estimated by size. He concludes that size is a significant factor in explaining the variability of the number of losses. However, the model's main limitation is that it does not account for the discrete character of the frequency data.

In this study, we develop a count data regression model. A model of this type can take the discrete and non-negative character of the data into account. Two models will thus be tested: the Poisson and negative binomial (Klugman, Panjer, and Willmot, 1998; Cruz, 2001). The regression component contained in the model allows us to take into account certain factors related to the scaling procedure.

In the models used to describe discrete variables in the literature (Cox and Lewis, 1966; El Sayyad, 1973; Frome, Kutner, and Beauchamp, 1973; Hausman, Hall, and Griliches, 1984; Gouriéroux, Monfort, and Trognon, 1984), the endogenous variables are supposed to have a Poisson regression distribution. The parameter of this distribution is a function of the values of the exogenous variables. The choice of this model is justified when the dependent variable counts the occurrence of a given event over a specific period and when the usual hypotheses of the Poisson distribution are satisfied.

Several applications of this model appear in the literature. It has been used to model such risks as: the number of patents received by a firm (Hausman, Hall, and Griliches, 1984), the number of visits to a doctor (Cameron, Trivedi, Milne, and Piggott, 1988) or the number of automobile or plane accidents (Dionne and Vanasse, 1989 and 1992; Dionne et al. 1997). The Dionne-and-Vanasse application (1989) is the first to introduce a regression component in the insurance field, a field showing many similarities with operational risk. The number of accidents per individual is supposed to follow a Poisson distribution whose parameter will vary from one exposure unit to the next. This parameter actually depends on the characteristics of the units exposed. As discussed by Maddala (1983) and Cameron, and Trivedi (1986), the coefficients of these variables are estimated using the maximum likelihood method.

The Poisson regression model supposes equidispersion (equality between the conditional average and variance). This restriction may not be compatible with operational loss data. Recourse to a negative binomial distribution[2] compensates for this problem, since it allows overdispersion. The studies done by Dionne and Vanasse (1989 and 1992) and by Boyer, Dionne, and Vanasse (1991) have shown the superiority of the negative binomial regression model over the Poisson regression model when treating automobile accidents. The negative binomial regression model is now frequently used in the insurance literature.

Once certain conditions have been met, it is possible to use maximum likelihood in estimating distribution parameters. But if the density is poorly specified, the estimators found with maximum likelihood will not be good. Gouriéroux, Monfort, and Trognon (1984a and 1984b) have proposed other methods to counteract this problem, such as the pseudo maximum

likelihood (PML) and the quasi-generalized pseudo maximum likelihood (QGPML). They have specified the conditions under which these PML and QCPML estimators from the linear exponential family of models will behave coherently when applied to non-truncated models. However, if the density of the negative binomial is correctly specified, the maximum-likelihood estimators will be more efficient than the PML and QGPML (Dionne and Vanasse, 1992).

In modeling the number of operational losses, we shall apply these models, which have gained great popularity in the literature. These models make it possible to introduce information on the financial institution where the loss occurred. Exogenous variables reflecting the firm's location and geographical distribution will help account for scaling. To our knowledge, this is the first time these models are being applied to operational risk and, more precisely, being used to scale the frequency of operational losses. However, we observe that the frequencies do exceed zero. We thus develop truncated Poisson and negative binomial regression models at point zero. The truncated densities of these models have been presented by Cameron and Trivedi (1998) and Gurmu (1991). Gurmu and Trivedi (1992) have developed overdispersion tests for the same models.

3. Description of external data

Fitch's OpVaR database is made up of operational losses of US $1 million and over. This database contains losses from all industries. Since our only target is banks, the database was first screened to select only operational losses connected with financial institutions.

The database contains the following types of information.

1. Type of event, level 1: Types of risk defined by regulatory authorities. Under this heading we find:
    - External fraud
    - Internal fraud
    - Clients, products, and business practices
    - Employment, practices, and workplace safety
    - Execution, delivery and process management

[2] This is a Poisson distribution whose random parameter follows the gamma distribution.

- Damage to physical assets
- Business disruption and system failures

Also available are types of events at levels 2 and 3, which offer greater precision and granularity. For example: discrimination and diversity as a sub event of the *employment, practices, and workplace safety* risk type. As type-3 events under the diversity and discrimination subtype we have, for example, discrimination due to age, sex, race, sexual orientation, and sexual harassment.

2. The name of the parent company and that of the subsidiary;
3. A detailed description of the loss event;
4. The loss amount in local currency, in American dollars, and its real value (counting inflation);
5. Date of the event. Since we are not always sure of the exact date (day and month); we use only the year of the event;
6. Industry: Either financial services or public administration;
7. Business unit, levels 1, 2, and 3: The first level makes the distinction between financial and non-financial institutions. In our case, we are only interested in the financial sector. Level 2 is concerned with financial institutions and makes the distinction between banks, insurance firms, investment banks, and other institutions. In level-3 business units, we find the lines of business defined by the Basel Committee.
8. The country where the loss occurred.
9. An identification code for each loss.
10. Information on the institution where the loss occurred: total assets, total equity, total deposits, total revenues, and number of employees.

It is worth mentioning that the firm-related information needed for some observations is missing. Since this information will be of key use in the scaling model later on, we are obliged to select only the loss data for which specific information on the firm is available. Moreover, 1.8% of losses occurred between 1981 and 1994 and averaged $130.31 M per event, whereas the losses,

which occurred later, averaged $67.15 M per event. We thus remove the events having occurred before 1994 from the external base because of a collection bias. Hence, 1,056 observations of losses of more than a million American dollars remain in the database.

4. Hypotheses of the model

We apply a theoretical model designed to scale severities and frequencies. The empirical application of scaling models to external data is of great help in showing the model's simplicity and in producing results. It is clear that the database used is open to criticism for the reasons mentioned above. Since, for the moment, no better data sources exist, we will be guided by the following hypotheses in carrying out our analysis. The methodology used would admittedly be applicable to other bases, provided they contain the information required.

- We suppose that the loss amounts recorded in the base as reported in the media are exact and factorial. The evaluation of losses is thus based neither on rumours nor predictions.
- We suppose that all types of losses are as likely to be recorded in the base; there is thus no media effect related to certain types of risk.
- We suppose that the external base provides all the losses of more than a million dollars for the financial institutions contained in it.
- We suppose that there is no correlation between the amount of the loss and the probability of its being reported. The severity and frequency distributions are thus supposed to be independent.

5. Scaling model for external-loss amounts

5.1 Theoretical scaling model

The scaling mechanism depends on three fundamental hypotheses. The first is that the monetary loss can be broken down into two components: common and idiosyncratic or specific. The second stipulates a non-linear relation between the idiosyncratic component and the different factors composing it. The third and last hypothesis states that, aside from the factors controlled for the purpose of scaling, all the other non-observable factors (quality of control environment, etc.) are supposed to remain the same for all banks.

Concerning the first hypothesis, we can suppose that the operational loss can be broken down into two components (Na, 2004, Na and al., 2006): a component common to all banks and an idiosyncratic component specific to each loss. The common component contains all the factors which, being independent of any specific bank's activities, can have the same impact on all banks—thus making it a constant component for all loss events. It refers to the macroeconomic, geopolitical or cultural environment or even to human nature in general.

The idiosyncratic component refers to the specific risk facing the financial institution or line of business. Some elements of this component are observable: bank size, type of risk, line of business or location of loss event. These could therefore be quantified or measured. But there are non-observable elements related to the control environment, which are difficult to quantify. These elements are not studied in this article.

We can thus identify a loss amount as a function of these two components:

$$Loss_i = f\left((Comp_{common}), (Comp_{idiosyncratic})_i\right). \quad (1)$$

The second hypothesis stipulates that the function f is non-linear. Na (2004) supposes that the f function is the product of a function of the common component and of a function of the idiosyncratic component. Now, since the common component is constant, we can model it with the parameter:

$$Loss_i = Comp_{common} \times g(Comp_{idiosyncratic})_i. \quad (2)$$

As for function *g*, we draw on the study by Shih, Samad-Khan, and Medapa (2000) which supposes a power relationship between the loss amount and firm size. However, size (estimated by total assets) is not the only factor we use to determine the severity of losses. We add to it other factors expressed in the function *h* which follows:

$$g(Comp_{idiosyncratic}) = Assets^{a} \times h(factors).$$

We can thus rewrite (2) as follows:

$$Loss_i = Comp_{common} \times (Assets^{a} \times h(factors)).$$

To simplify the analysis, we suppose that:

$$h(factors) = exp(\sum_j b_j \ factors_j).$$

Thus,

$$Log(Loss_i) = Log(Comp_{common}) + a\ Log(assets_i) + (\sum_j b_j \ factors_{ij}). \quad (3)$$

In order to explain the variability of the losses and to construct the scaling model, the different elements of the idiosyncratic component must be identified, since they play a role as factors explaining the severity of losses.

5.2 Description of the variables

The dependent variable is the logarithm of the operational losses. The statistics in Table 1a show that the average by loss event is evaluated at $67 million, with a standard deviation of $521 million. The maximum of the losses is $16 billion. The loss amounts thus vary widely from quite substantial to catastrophic.

The explanatory variables to be included in the model designed to explain the variation in the logarithm of losses are described below. Table 1 presents descriptive statistics of the losses in terms of these variables.

- Size: The base contains variables characterizing firm size. According to the results of the study by Shih, Samad-Khan, and Medapa (2000), size is weakly lined to the loss amount. Other variables must explain this variability.

  Many information on size are available, such as: total revenues, total assets, total deposits, number of employees, and total equity. However, since all these variables are correlated, we have chosen total assets (the variable most correlated with losses) as the estimator for size. Financial institutions having sustained losses reported in the database used differ greatly in size, varying from the smallest bank (with total assets of $43 million) to the largest institution (with assets of $1,533,036 million). The average total in assets is evaluated at $270,681M.

  In Table 1a, we present the number of events, the average, and the standard deviation for losses according to size. We have thus classified the banks into three size categories: those with assets under $400 billion (smaller and medium size); those with assets between $400 and $800 billion (large size); and very large banks whose assets excess $800 billion (very large size). The results in the table show that average losses are much higher in the very large banks than in the others. But the average loss for large financial institutions is lower than that for small and medium size banks.

  We expect losses to increase with the size of the financial institution. So a natural catastrophe could, for example, cause more serious damage (losses) to a bank whose total assets are higher than those of another bank. Size could thus have a positive impact on the severity of losses.

- Location: As losses do not all occur in the same country, a variable capturing the effect of location must be incorporated. Seeing differences in environment, legislation, etc., we expect this variable to be significantly linked to loss amounts. It is worth noting that 60% of the losses occurred in the United State, as compared to only 4% in Canada. This variation can be explained by the fact that the number of banks in the United States greatly exceeds that in Canada. The remaining proportion of losses has been divided between the countries of Europe and the rest of the world.

Table 1b presents statistics for losses according to location. We note that the average for operational losses differs according to the location where they occur, thus reflecting their different environments. It is worth mentioning that the average for losses in the United States is higher than that in Canada ($38 M Vs $9 M). And the environment designated *Other* (countries other than Canada, the United States and those in Europe) reports the highest average for losses ($163 M), thus making it the most risky.

Seeking to define the link between size of financial institution and location, we present, in table 1c, statistics on the size of institutions according to the location where the losses occurred. We notice that the average for total assets is lower in Canada than in the United States and Europe. However, though the environment in countries designated *Other* is riskier (highest average losses), the institutions having suffered those losses are on average smaller than those located in the three other environments. Thus, there is no direct link between the financial institution's size and the location where the losses occurred. Dummy variables (United States, Canada, Europe, and Others) capture the effect of the location of losses in one of the countries or continents mentioned above.

- Line of business: We expect the business line to have an impact on the severity of losses. Certain units register higher losses than others, on average. Highlighting the line of business where the loss occurred can explain the severity of extreme losses. According to the statistics presented for seven business lines in table 1d, we see that two of them—commercial *banking* (25%) and *retail banking* (33%)—account for 58% of the losses. And the average loss is much higher for *commercial banking* than for the other units.

  Based on the information collected by LDCE and QIS-4[3] from 27 financial institutions, the line of business *retail banking* accounts for 44% of the operational losses in the 177 data on losses of over $1 M collected over the 2001-2004 period, whereas the line of business *commercial banking* accounts for only 9%. This gap can be explained either by the collection period (the LDCE study covers 4 years of losses, whereas the external base covers 11 years of reporting) or by the number of different financial institutions where

[3] Loss Data Collection Exercise (LDCE) and Quantitative Impact Study-4 (QIS-4): two studies conducted by the U.S. federal bank and thrift regulatory agencies to evaluate the impact of Basel II on the minimum regulatory capital required.

the losses occurred. The dichotomic variables for each line of business will thus capture the effect of the nature of its activity, when determining loss amounts.

- Types of risk: Certain risk types are infrequent but extremely severe, whereas others are very frequent but of relatively weak severity. Table 1e shows that 44% of losses are of the *client, products, and business practices* type and that more than 40% of losses are divided between *internal* and *external fraud*. However, less than 0.5% of losses are of the *damage to physical assets* type and 0.5% are of the *business disruptions and* system *failures* type. The average loss is highest for the *damage to physical assets* risk type ($115 M), whereas it is lowest for *business disruptions and system failures* ($5 M)

  The results of the LDCE and QIS-4 studies show a very great difference in relation to this distribution. In effect, 49% of losses are of the *execution, delivery and process management type*; 31% are of the *clients, products, and business practices* type; 7% of the *external fraud* type; and 3% of the *internal fraud* type. As for the types of risk *damage to physical assets* and *business disruption and system failures*, the proportion of losses is just as low as in the external base.

  Introducing dichotomic variables to capture this aspect of the risk type can be relevant in explaining the variability of the loss amounts. Thus, 7 variables will capture the "risk type" effect in our model.

5.3 Linear regression

To explain the degree of variability of external losses, we shall estimate the coefficients of the regression below. This will allow us to evaluate the common and specific components for each loss amount. The following regression follows from equation (3):

$$Y_i = a_0 + a_1\ Size_i + a_2\ US_i + a_3\ Canada_i + a_4\ Europe_i + \sum_{j=5}^{11} a_j BL_{ij} + \sum_{j=12}^{17} a_j RT_{ij} + e_i \qquad (4)$$

with:

Y_i: Log($losses_i$);

a_0: Common component;

$Size_i$: Log(assets$_i$);

US_i: Binary variable assuming the value 1 if the loss occurred in the United States, otherwise 0;

$Canada_i$: Binary variable assuming the value 1 if the loss occurred in Canada, otherwise 0;

$Europe_i$: Binary variable assuming the value 1 if the loss occurred in Europe, otherwise 0;

The category omitted is *Others*;

BL_{ij}: Binary variable assuming the value 1 if the loss occurred in the business unit j, otherwise 0;

The category omitted is *payment and settlement*;

RT_{ij}: Binary variable assuming the value 1 if the loss is of the risk type j, otherwise 0;

The category omitted is *business disruptions and system failures* type of risk;

e_i: Deviation variable representing the non-observable specific component which is supposed to follow a normal distribution with parameters $(0, \sigma^2)$.

5.4 Results of the regression

The Ordinary Least Squares (OLS) method is used to estimate the parameters. The results of this estimation are presented in Table 2. The adjusted R^2_{adj} is 10.63%. Though the value is low, it is better than the 5% found in the literature to date (Shih et al., 2000). Remember that it is difficult to capture certain non-observable factors, which are not present in the external base.

Estimated by the logarithm of total assets, the size variable is significantly different than 0. The coefficient is positive, confirming the fact that the larger the firm, the higher its level of losses. The binary coefficients US and Canada are significantly different then 0. The negative sign must be interpreted in relation to the category *Others* (variable omitted). Comparing the coefficients US and Canada, we note that the United States' environment is riskier than Canada's. It is also

worth noting the even higher losses having occurred in the rest of the world where financial environments are less regulated than those in the United States, Canada, and Europe.

The *commercial banking variable* is the only one linked to business line, which shows a significantly non-null impact at a 99% confidence level. The coefficient is positive, showing that losses for this type of risk are higher than for others. Finally, the *clients, products, and business practices* variable has significant explanatory power. This shows that the losses associated with this type of risk are higher compared to the others.

5.5 Robustness tests for the size variable

We start with a simple regression including only the size variable— a model similar to the one used by Shih et al. (2001). The other categories of variables are then added to this same model one by one in order to capture any possible additional effects and to test the stability of the parameters.

The results show that size plays a very small part in explaining the level of losses. Model 1 in Table 3 actually shows an R^2_{adj} of 0.6%. This statistic is sharply improved (to 4.32%) once the variables associated with location are introduced. It should be mentioned that the values of the coefficients estimated remain stable and significantly different from 0 when compared to the basic model. Model 3 adds variables estimating the impact of the line of business where the loss occurred. The adjusted determination coefficient jumps from 4.32% (model 2) to 7.16% (model 3). The *commercial banking variable* still remains significantly different from 0. Each category of variables thus has significant power in explaining the severity of operational losses. The coefficients of the variables are relatively stable.

We next select only those variables, which are statistically non-null at the 90% confidence level. We then regress the log of losses on the 5 remaining variables. The model thus constructed allows us to test whether the variables shown to be significant in the basic model will keep their explanatory power when tested alone. The results presented in model 4 of table 3 show that all the variables remain significantly non-null at the 90% confidence level. The adjusted coefficient of determination is on the order of 9.38%. The signs and scope of the variables do not change. These 5 significant variables will be used in developing the normalization formula.

5.6 Normalization Formula

We are investigating a certain bank *A* and we want to find the equivalent value of a loss occurring in another financial institution *B*. A normalization formula will allow us to put a loss having occurred in bank *B* on the same scale as one in bank *A.*

According to equation (2), a loss *i* is the product of a common component and of a function of the specific component. The regression analysis performed above allowed us to identify these two components.

$$Log(loss_i) = \underbrace{a_0}_{\log(\text{Comp.Comm})} + \underbrace{a_1 Size_i + a_2 US_i + a_3 Canada_i + a_4 CB_i + a_5 CPBP_i + e_i}_{\log\ g(Comp\ \text{Idiosyncratic})}$$

where:

CB: refers to the *commercial banking* business line;

CPBP: refers to the *clients, products, and business practices* risk type.

As the common component is constant for all loss amounts, it is possible to re-write equation (2) as follows:

$$Comp_{comm} = \frac{Loss_A}{g(Comp_{idio})_A} = \frac{Loss_B}{g(Comp_{idio})_B} = ... = \frac{Loss_N}{g(Comp_{idio})_N}. \quad (5)$$

Suppose that we have a loss which occurred in bank *B* and that we want to know its equivalent value if it occurred in bank *A.* Based on the analysis above, we can determine the idiosyncratic components of loss *B* as well as that of *A*. We multiply the coefficients already estimated by the corresponding value of the different variables to find the idiosyncratic or specific component.

$$Loss_A = \frac{g(Comp_{idio})_A}{g(Comp_{idio})_B} \times Loss_B \quad (6)$$

with:

$$g(Comp_{idio})_A = \exp(\hat{a}_1 Size_A + \hat{a}_2 US_A + \hat{a}_3 Canada_A + \hat{a}_4 CB_A + \hat{a}_5 CPBP_A).$$

Equation (6) supposes that, in addition to the variables selected to perform the scaling, the unexplained part of the regression model (attributable to unobservable qualitative factors such as management quality, control environment, etc.) is supposed to be the same between $Loss_A$ and $Loss_B$ (third hypothesis).

So, to calculate a loss sustained by a given bank in the banking industry, the idiosyncratic components of the two losses must first be calculated with the preceding equation. Next, we apply formula (6) to find the equivalent loss for bank A. By applying this same method to the whole external base, we obtain a base of extreme losses having occurred in other banking institutions but scaled to a given bank. The severity of losses has thus been adjusted by taking into account several factors such as size, location, business line, and risk type.

5.7 Validation of the scaling severity model

In order to concretize the scaling model, we have chosen the American bank Merrill Lynch[4] from the external base. This bank shows 52 loss events over the 1994-2004 period. We shall first scale the operational losses in the external base to this bank. We shall next compare the statistics on the losses actually observed to those found after the scaling procedure.

Our first step is to determine the equivalent of the 1,056 loss events in the external base for the Merrill Lynch bank. We shall thus calculate the loss amount, which could occur at Merrill Lynch for the same type of risk, in the same line of business, and in the same year as the one in the external base. On the other hand, we shall take Merrill Lynch's total assets in the year of the event and apply them to all the losses. And we shall take the United States as the place where all the external losses occur, since all the losses observed for Merrill Lynch did occur in the United States. Once the explanatory variables for regression (4) have been identified, we shall calculate the equivalent idiosyncratic component for each loss recorded in the external base (as shown in Appendix 1) and also include the coefficients of the variables previously estimated. The normalization formula presented in the preceding section allows us to scale the loss for the bank in question.

[4] We have chosen this bank from the external base because it is the one with the maximum number of losses of more than $1 million over the 1994-2004 period.

We next compare the statistics calculated on the sample of the 52 losses actually observed at Merrill Lynch to the statistics calculated on the 1, 050[5] losses equivalent to those in the external base which could occur at the same bank. These statistics are presented in Table 4. They show that the averages of the two samples are quite close. A hypothesis test confirms this and shows that the two averages are statistically identical at a 95% threshold. It should also be noted that the standard deviations of the two samples are close. (83.1 vs. 84.3).

In the second step of the analysis, we look to see what impact the scaling variables have on the loss amounts obtained after normalization. Table 5 presents the example of a loss-event at the Bank of New York drawn from the external base, along with loss amounts scaled to fictional banks. First of all, we modify one characteristic of the event at a time in order to see its monetary impact on the loss. We next analyze the aggregate effect of several variables on the loss amount.

We note that, if the event took place in one of the larger banks (all other factors being equal), the loss amount would be slightly higher (rising from $8.26 M to $9.27 M, while the asset total has more than quadrupled). However, if the same event took place in Canada rather than the United States, the loss would be smaller in scope (it would go from $8.26 M to $4.97 M). As we have already shown, Canada's environment is less risky than that of the United States.

We also note that the *commercial banking line* of business where the loss-event occurred is more likely to produce heavy losses than the other lines of business. And the loss will move from $8.26 M to $16.06 M if it takes place in the *commercial banking* business line rather than the *retail banking* business line. Type of risk has also a big impact on the scope of losses. If the loss is of the *clients, products, and business practices* type, the amount will move to $15.56 M, given that this type of risk has a significant impact on the severity of losses, as already concluded in section 5.4.

Finally, in the three last lines of table 5, we present the aggregated impact of two or more variables. We scale by modifying the size of the bank and the location of the event. Note that even if the bank is a larger one, the impact of location (Canada) wins out and the loss amount produced by scaling is lower than that of the original event. When the line of business and type

[5] We have excluded 6 losses from the analysis because they represent outliers.

of risk are also modified, we find that the loss more than doubles. It is worth noting that the resulting loss is very different from the one found when size alone is modified. This analysis thus shows us that the size effect is quite weak compared to the other scaling factors. Our model is thus an improvement over models existing in the literature, which are based solely on size.

6. Scaling model for frequency of external losses

Remember that our objective is to correct the scaling bias so that a combination of internal and external data can be used to measure operational risk capital. In the preceding section, we worked out the scaling for loss amounts. It is thus possible to find several extreme losses likely to occur in our reference bank *A*. The question still to be asked is: How frequently will a bank sustain these losses?

The scaling of frequencies is a notion, which rarely surfaces in the literature. Some researchers have developed models to scale severity, but the number of external losses, which should be combined with internal data, has not yet been modeled. In what follows, we propose a model, which allows us to adjust the number of external losses per bank and to scale it down to a given bank *A*.

6.1 Description of the model

With the model developed in this section, it is possible to scale the number of external losses and to determine what theoretical distribution fits to the frequencies. We expect that, on a given horizon, the number of losses per financial institution will depend on certain factors describing the characteristics of the financial institution. The institution's size can indeed play an important role in determining the number of losses. It should be noted that the larger the bank, the more exposed it is to operational risks. If a bank does more transactions and has more assets, employees, and revenues than another, it will probably have more operational losses of various types (fraud, damage to assets…). And the geographical distribution of the institution's activities can give us an idea of the effectiveness of its controls. The more widely dispersed a bank's activities and, consequently, its measures of control, the less effective these measures will be.

It is thus possible to explain the number of losses by a regression over the different variables mentioned above. However, since we are dealing with frequencies and thus discrete variables, these numbers can be more suitably modeled with count data distributions such as the Poisson and negative binomial. So the count data regression model can be appropriately applied in this context. The advantage of these models is that they can both find the theoretical distribution adjustable to the frequency data and also provide flexible parameters suitable to each observation. In other terms, the distribution's parameters depend on the variables identifying the characteristics of the financial institution where the loss occurred. Once the parameters have been estimated, it is possible to calculate those belonging to a given bank.

Since the only institutions to which we have access are those, which have sustained losses, the frequencies are non-null. This bias must be corrected by using distributions truncated at zero. In what follows, we shall first describe the variables, which will be included in the model and then present and test each of the two models: the truncated Poisson regression model and the truncated negative binomial regression model.

6.2 Description of the variables

We create a variable describing the number of over $1 M losses per financial institution over the 1994-2004 period.[6] This gives us a sample of 323 financial institutions having sustained losses of over $1 M which have been reported in the external base. Frequency will be explained based on a bank's size and on the geographical distribution of its activities. We expect to find that the number of losses will increase with bank size and that control costs will grow and decrease in their effectiveness, as a financial institution expands the geographical distribution of its activities.

Size will be estimated by the logarithm of the average total of the firm's assets over the 1994-2004 period. The geographical distribution will be estimated by binary variables such as:

- *US:* Binary variable assuming the value 1 if the institution has had losses in the United States over the 1994-2004 period, otherwise 0.

[6] We select this period for which the collection of losses is most exhaustive.

- *Canada:* Binary variable assuming the value 1 if the institution has had losses in Canada over the 1994-2004 period, otherwise 0.
- *Europe:* Binary variable assuming the value 1 if the institution has had losses in Europe over the 1994-2004 period, otherwise 0.
- *Others:* Binary variable assuming the value 1 if the institution has had losses in another country over the 1994-2004 period, otherwise 0.

Unlike those in the model for severity, these variables are not mutually exclusive. Table 6a presents the descriptive statistics for the number of losses per bank as well as total assets per financial institution over the 1994-2004 period. The average number of losses of over $1 M is 3.3 events per institution on an 11-year horizon, with a maximum of 52. The financial institutions vary greatly in size, the average total in assets being $123,174 M. With regard to the geographical distribution of the banks' activities, we find that losses are more concentrated in the United States and in other countries (other than Canada, the United States, and Europe) Table 6b shows that banks of small and medium size (average assets under $400,000 M) suffer fewer losses over the 1994-2004 period than do large banks (average assets of between $400,000 and $800,000 M). And very large banks, with assets of over $800,000 M, have a higher number of losses (19) than the other banks. These statistics show a link between the financial institution's size and the number of losses of over $1 M.

Table 6c presents statistics on the number of losses per bank according to the geographical distribution of the activities of the institution in question. The results show that if activities are concentrated in the same country, the average number of losses is 2 per bank. But when activities are spread over two or three countries, the average varies between 8 and 10. The average number of losses per banking institution jumps to 28 when activities are very widely dispersed geographically. In modeling frequency, it is thus interesting to take the geographical distribution of activities into account.

6.3 Truncated Poisson regression model

If Y_i — the number of losses sustained by company *i* over the 1994-2004 period—follows a Poisson distribution, then the probability of having *y* losses will be:

$$P(Y_i = y) = \frac{e^{-\lambda}\lambda^y}{y!} \qquad y = 0, 1, 2... \ and \quad \lambda > 0$$

where λ is the Poisson parameter. The main characteristic of this distribution is $E(Y_i) = Var(Y_i) = \lambda$.

However, as we are in the presence of zero truncated data, the probability of the number of conditional losses must be estimated on the fact that the frequencies observed are strictly superior to zero. The conditional probability is:

$$P(Y_i = y \mid Y_i > 0) = \frac{e^{-\lambda}\lambda^y}{y!\left(1 - e^{-\lambda}\right)} \qquad y = 1, 2... \ and \quad \lambda > 0 .$$

We can, moreover, allow parameter λ to vary from one observation to the next. Let $\lambda_i = \exp(X_i\beta)$ where X_i is a vector of (1 × *m*) exogenous variables (characteristics of the firms where the loss occurred) and β a vector of (*m* × 1) coefficients. The exponential function allows the non-negativity of parameter λ_i.

In our context, the parameter λ_i takes the form:

$$\lambda_i = \exp(\beta_0 + \beta_1 \ln(Assets)_i + \beta_2 US_i + \beta_3 Canada_i + \beta_4 Europe_i + \beta_5 Others_i). \qquad (7)$$

Therefore, the probability that a financial institution *i* would have *y* losses over an 11-year horizon (when its specific characteristics are known) is:

$$P(Y_i = y \mid Y_i > 0 \ ; X_i) = \frac{e^{-\exp(X_i\beta)} \exp(X_i\beta)^y}{y!\left(1 - e^{-\exp(X_i\beta)}\right)}. \qquad (8)$$

It should be noted that, unlike the Poisson, the truncated Poisson distribution does not present an equidispersion. In fact, the first and second moments are:

$$E(y_i \mid y_i > 0;\ X_i) = u_i = \lambda_i + \gamma_i$$
$$E(y_i \mid y_i > 0;\ X_i) = \frac{\lambda_i}{1 - e^{-\lambda_i}}$$

and

$$V(y_i \mid y_i > 0;\ X_i) = \sigma_i^2 = \lambda_i - \gamma_i\ (u_i - 1)$$
$$V(y_i \mid y_i > 0;\ X_i) = u_i(1 - \gamma_i)$$

with

$$\gamma_i = \frac{\lambda_i}{e^{\lambda_i} - 1}.$$

γ_i is called an adjustment factor. It is interpreted as being the difference between the averages of the truncated and non-truncated distributions. The average of the truncated distribution is in fact higher than that of the non-truncated distribution. However, the variance of the truncated distribution is smaller, as the preceding expressions show.

The vector of the β parameters can be estimated using the maximum likelihood method (see Maddala, 1983 and Cameron and Trivedi, 1986, for a detailed discussion on the estimation methods of the Poisson regression models). It is thus a matter of estimating the six coefficients of equation (7) and of calculating parameter λ_i for each exposure unit *i* (a firm in our case). It is thus possible to calculate a given bank's λ parameter by retaining the estimated coefficients and taking the size and geographical distribution variables of the bank in question.

6.4 Truncated negative binomial regression model

In the preceding model, the exogenous variables may not contain all the information to explain the number of conditional losses. To compensate for this restriction, we introduce an error term in the definition of the Poisson parameter that can be written:

$$\lambda_i^* = \exp(\beta_0 + \beta_1 \ln(Assets)_i + \beta_2 US_i + \beta_3 Canada_i + \beta_4 Europe_i + \beta_5 Others_i + \varepsilon_i). \qquad (9)$$

The term ε_i constitutes the error specification arising from the omitted (unobserved) explanatory variables, which are independent of the model's exogenous X_i variables. Contrary to the preceding model, λ_i^* is a random variable. The conditional distributions of Y_i becomes:

$$\Pr\left(Y_i = y \mid Y_i > 0\,;\, X_i\right) = \int_{-\infty}^{+\infty} \frac{e^{-\exp\left(X_i\beta+\varepsilon_i\right)} \exp\left(X_i\beta + \varepsilon_i\right)^{y_i}}{y_i!\left(1 - e^{-\exp\left(X_i\beta+\varepsilon_i\right)}\right)} g\left(\varepsilon_i\right) d\varepsilon_i \quad \text{y} = 1,2,...$$

with $g(\varepsilon_i)$ as the density function of ε_i.

For convenience sake, we suppose, in what follows, that $\mu_i = \exp(\varepsilon_i)$ follows a gamma distribution of parameter α such that:

$$f(\mu_i) = \frac{\mu_i^{\frac{1}{\alpha}-1} \exp\left(-\frac{\mu_i}{\alpha}\right)}{\alpha^{\frac{1}{\alpha}}\Gamma\left(\frac{1}{\alpha}\right)}, \quad \alpha > 0,$$

with $E(\mu_i) = 1$ and $V(\mu_i) = \alpha$.

The conditional distribution of Y_i is a truncated negative binomial whose density is:

$$P\left(Y_i = y \mid Y_i > 0;\; X_i\right) = \frac{\Gamma\left(y + \frac{1}{\alpha}\right)}{\Gamma(y+1)\Gamma\left(\frac{1}{\alpha}\right)} \times \frac{1}{\left(1 + \alpha \exp\left(X_i\beta\right)\right)^{\frac{1}{\alpha}} - 1} \times \left(\frac{\alpha \exp\left(X_i\beta\right)}{1 + \alpha \exp\left(X_i\beta\right)}\right)^{y}. \qquad (10)$$

The mean and the variance of the truncated negative binomial regression are the following:

$$E\left(y_i \mid y_i > 0\,;\; X_i\right) = u_i^* = \lambda_i + \gamma_i^*$$

and

$$\mathrm{V}\left(y_i \mid y_i > 0\;;\; X_i\right) = \sigma_i^{*2} = \lambda_i + \alpha\lambda_i - \gamma_i^*\left(u_i^* - 1\right)$$

with

$$\gamma_i^* = \frac{\lambda_i}{\left[(1+\alpha)^{\alpha^{-1}\lambda_i} - 1\right]}.$$

As for the Poisson distribution, the average of the truncated negative binomial distribution is higher than that of the non-truncated one. And the point-zero truncation has reduced the variance by the adjustment factor γ_i^*. Though the truncated Poisson distribution no longer shows the equidispersion characteristic $\left(E(y_i \mid y_i > 0 \ ; \ X_i) \neq V(y_i \mid y_i > 0 \ ; \ X_i)\right)$, the truncated negative binomial model introduces overdispersion, in the sense that its variance is higher than that of the Poisson.

Under the hypothesis that the density is correctly specified and that the other usual conditions are met, a maximum likelihood estimation of the parameters β and α can be performed. It is thus a matter of estimating the six coefficients of equation (9) as well as the coefficient α. It is thus possible to calculate the parameter λ_i^* for a given bank. We thus retain the estimated coefficients and take the size and geographical distribution variables of the bank in question. However, if the density is poorly specified, the maximum likelihood estimators of the negative binomial regression model will not be consistent.

6.5 Presentation and comparison of results

The results obtained from estimating the parameters of the two models are summed up in table 7. Note that the estimated coefficients have the same signs for both models. The signs confirm our expectations. In fact, the positive size and geographical distribution coefficients show that the number of losses do increase with the bank's size and the distribution of its activities across different environments.

In the truncated Poisson regression model, all the explanatory variables introduced in the regression are statistically significant at confidence levels ranging between 90% and 99%. They thus have an explanatory impact on the number of losses. As to the truncated negative binomial regression model, the coefficients of the variables *Europe* and *Others* are not statistically different from zero.

In order to compare the two models and to test for the existence of overdispersion in data, we run an overdispersion test specifically designed for truncated regression models by Gurmu (1991).

The test consists in testing the null hypothesis $H_{0:}\ \alpha = 0$ against the alternative hypothesis $H_{1:}\ \alpha > 0$.

Let $\rho = \begin{pmatrix} \beta \\ \alpha \end{pmatrix}$ be the vector of the coefficients and $\hat{\rho} = \begin{pmatrix} \hat{\beta} \\ 0 \end{pmatrix}$ the maximum-likelihood estimator of ρ under the null hypothesis. In what follows, the symbol $\hat{}$ indicates that the parameters are evaluated at point $\hat{\rho}$ under the hypothesis stipulating that $\alpha = 0$.

The statistic of the Score test or the Lagarange multiplier used to test for overdispersion in the truncated Poisson regression model is calculated as follows:

$$\tau = \frac{\sum_{i=1}^{N} \hat{\lambda}_i^{-1}\left[\hat{\upsilon}_i^2 - y_i + (\hat{\upsilon}_i + y_i)\hat{\gamma}_i\right]}{2\left[\Im_{\alpha\alpha}(\hat{\rho}) - \Im_{\alpha\beta}(\hat{\rho})\Im_{\beta\beta}^{-1}(\hat{\rho})\Im_{\beta\alpha}(\hat{\rho})\right]^{\frac{1}{2}}} \overset{a}{\sim} N(0,1)$$

where

$\hat{\upsilon}_i = y_i - \hat{u}_i$;

$\overset{a}{\sim}$ indicates that this statistic is asymptotically distributed as the standard normal distribution

$\Im(\rho) = -E\left(\frac{\delta^2 L(\rho)}{\delta\rho\delta\rho'}\right)\Big|_{H_0} = \begin{bmatrix} \Im_{\beta\beta}(\rho) & \Im_{\beta\alpha}(\rho) \\ \Im_{\alpha\beta}(\rho) & \Im_{\alpha\alpha}(\rho) \end{bmatrix}$ is the information matrix.[7]

Under the hypothesis H_0: $\alpha = 0$. The value of statistic τ based on the truncated negative binomial distribution at point zero is 34, 637. The test thus reject of the null hypothesis. The result shows the existence of significant overdispersion in the data, indicating that it is suitably modeled with the negative binomial distribution.

To find the parameters of the distribution selected for a given bank, it is simply a matter of calculating parameter λ_i^* using equation (9) along with the coefficients already calculated and

[7] Refer to the Gurmu study (1991) for more details on the evaluation of information matrices at point $\hat{\rho}$.

the value of the variables for this bank over the period in question. It is thus possible to generate a number of $1 M+ losses a bank might sustain, based on its own frequency distribution.

To concretize our frequency scaling, we use the example of a bank whose assets total $100, 000 M and whose activities are divided between the United States and Canada over the 1994-2004 period. We show how we calculate the parameter of the Poisson distribution using equation (7) in the table 8a and the parameters of the negative binomial distribution using equation (9) in the table 8b, along with the coefficients already estimated and the appropriate values of the variables.

From the description in table 8b, we see that it is possible to calculate the parameters of the negative binomial distribution. Equation (9) does make it possible to calculate the average number of losses on an 11-year horizon for the bank used as an example.

Using the frequency distribution selected for a given bank, we can generate a number of $1 M+ losses, which could occur in that bank over an 11-year period. We can then make a random draw of over $1 M loss amounts from the base scaled to the bank in question, according to the frequency generated. By calculating the total loss over the period in question and by repeating these steps several times, we thus generate a distribution of the extreme losses a bank could sustain over 11 years. It is thus possible to use the aggregated distribution to calculate statistics such as the average and different quintiles.

7. Conclusion and discussion

The use of external data is a regulatory requirement for banks seeking to develop an advanced method of calculating operational risk capital. This use must, however, be relevant and unbiased. Since we believe that all banks are exposed to infrequent but potentially heavy operational losses, it is imperative that they supplement their internal loss data with data on extreme external losses. This will allow them to make better estimations of distribution tails.

The objective of this study was to construct a scaling model for the severity and frequency of external losses in order to correct the scaling bias and to make better use of external losses. The results of the OLS estimation have shown that size, location (United States and Canada) as well

as business line (*commercial banking)* and risk type (*clients, products and business practices*) should be considered in explaining a part of the external loss amounts. A normalization formula allows us to take a loss observed in the industry and scale it to a reference bank. We have validated our model by comparing the actual losses observed in a bank with those found after scaling. The results show that the two samples have statistically equal averages and very similar standard deviations.

A model for scaling and adjusting frequencies has also been developed in response to the need to scale the number of losses. This model's originality (previously overlooked in the literature) is its ability to scale external frequencies to a given bank over a determined horizon. This same model can also generate a number of random $1 M and over losses, which could possibly occur in a bank. Two models have been tested: the truncated Poisson regression model and the truncated negative binomial regression model. The results show that the latter outperforms the former. Frequencies are scaled by calculating the distribution parameters of the bank in question. These parameters depend on the financial institution's characteristics, such as the size and geographical distribution of its activities.

It would be interesting to extend this model by finding the distribution of severity most suitably to fit to the loss amounts found after scaling. This would make it possible to generate losses of over $1 M randomly. The number of losses generated would be given by the negative binomial model used in this paper. It is thus possible to calculate operational capital based on scaled external losses. Other promising areas of research would be finding the best ways to combine internal and external data or determining what weight should be assigned to capital calculated on internal data as opposed to capital calculated on external data.

References

Basel Committee on Banking Supervision (2001), *Operational Risk-Consultative Document*, Supporting Document to the New Basel Capital Accord.

Basel Committee on Banking Supervision (2001), *Working Paper on the Regulatory Treatment of Operational Risk*.

Basel Committee on Banking Supervision (2003), *Third Consultative Paper*, The New Basel Capital Accord.

Federal Reserve System, Office of the Comptroller of the Currency, Office of Thrift Supervision and Federal Deposit Insurance Corporation (2005), *Results of the 2004 Loss Data Collection Exercise for Operational Risk*.

Baud. N, A. Frachot and T. Roncalli (2002), "How to Avoid Over-Estimating Capital Charge for Operational Risk?" Working paper, Groupe de Recherche Opérationnelle, Crédit Lyonnais.

Boyer, M., G. Dionne and C. Vanasse (1991), "Econometric Models of Accident Distributions," in G. Dionne (ed.): *Contributions to Insurance Economics*, Norwell, MA: Kluwer Academic Publishers.

Cameron, A. C. and P. K. Trivedi (1986) "Econometric Models Based on Count Data: Comparison and Applications of Some Estimators and Tests," *Journal of Applied Econometrics* 1, 29-53.

Cameron, A. C. and P. K. Trivedi (1998), "Regression Analysis of Count Data." *Econometric Society Monographs*, 30.

Cameron, A.C., P.K. Trivedi, F. Milne, and J. Piggott (1988), "A Microeconometric Model of the Demand for Health Care and Health Insurance in Australia," *Review of Economic Studies* 55, 85-106.

Chavez-Demoulin, V., P. Embrechts, and J. Nešlehová (2006), "Quantitative Models for Operational Risk: Extremes, Dependance and Aggregation," Working Paper, Department of Mathematics ETH-Zürich.

Cox, D. R. and P. A. W. Lewis (1966), *The Statistical Analysis of Series of Events*, Chapman and Hall, London.

Cruz, M. G. (2002), *Modeling, Measuring and Hedging Operational Risk*, John Wiley & Sons, LTD, Chichester.

Dionne, G. and C. Vanasse (1989), "A Generalization of Actuarial Automobile Insurance Rating Models: the Negative Binomial Distribution with a Regression Component," *Astin Bulletin* 19, 199-212.

Dionne, G. and C. Vanasse (1992), "Automobile Insurance Ratemaking in the Presence of Asymmetrical Information," *Journal of Applied Econometrics* 7, 2, 149-165.

Dionne, G., R. Gagné, F. Gagnon, and C. Vanasse (1997), "Debt, Moral Hazard and Airline Safety: An Empirical Evidence," *Journal of Econometrics* 79, 379-402.

El Sayyad, G. M. (1973), "Bayesian and Classical Analysis of Poisson Regression," *Journal of the Royal Statistical Society, Series B*, 35, 445-451.

Frachot, A and T. Roncalli (2002), "Mixing Internal and External Data for Managing Operational Risk," Groupe de Recherche Opérationnelle, Crédit Lyonnais.

Frome, E., M. Kutner, and J. Beauchamp. (1973), "Regression Analysis of Poisson Distributed Data," *Journal of the American Statistical Association*, 68, 935-940.

Gouriéroux, C., A. Monfort, and A. Trognon. (1984), "Pseudo Maximum Likelihood Methods: Applications to Poisson Models," *Econometrica* 52, 3, 701-720.

Gurmu, S. (1991), "Tests for Detecting Overdispersion in the Positive Poisson Regression Model," *Journal of Business and Economic Statistics* 9, 2, 215-222.

Gurmu, S., and P. K. Trivedi. (1992), "Overdispersion Tests for Truncated Poisson Regression Models," *Journal of Econometrics* 54, 347-370.

Hartung, T, (2004), "Operational Risks: Modelling and Quantifying the Impact of Insurance Solutions," Working Paper, Institute of Risk Management and Insurance Industry, Ludwig-Maximilians-University Munich, Germany.

Hausman, J., B. Hall, and Z. Griliches. (1984), "Econometrics Models for Count Data with an Application to the Patents R. & D. Relationship," *Econometrica* 52, 4, 909-938.

Klugman, S. A, H. H. Panjer, and G. E. Willmot (1998), *Loss Models, from Data to Decisions*, Wiley Series in Probability and Statistics, New York.

Maddala, G. S. (1983), *Limited-dependent and Qualitative Variables in Econometrics*, Cambridge University Press, Cambridge.

Na, H.S. (2004), *Analysing and Scaling Operational Risk*, Master Thesis, Erasmus University Rotterdam, Netherlands.

Na, H.S., J. Van Den Berg, L.C. Miranda, and M. Leipoldt (2006), "An Econometric Model to Scale Operational Losses" *The Journal of Operational Risk* 1, 2, 11-31.

Shih. J, A. Samad-Khan, and P. Medapa (2000), "Is the Size of an Operational Loss Related to Firm Size?" *Operational Risk Magazine* 2, 1.

Table 1: Descriptive Statistics on the Data

The following tables present descriptive statistics on losses according to size of financial institution, location of event, risk type and business line in which the event occurred.

Table 1a: Loss statistics according to size of financial institution, estimated by total assets in millions of dollars. The banks are classified in three categories: those with assets under $400, 000 M; those with assets between $400,000 M and $800, 000 M; and those of greater size with total assets of over $800 billion.

	Total assets (Millions of dollars)			**Whole sample**
	Assets < 400 000	**400 000 ≤ Assets <800 000**	**Assets ≥ 800 000**	
Average of losses (Millions of dollars)	69.079	51.422	115.570	67.154
Number of losses	781	231	44	1,056
Standard deviation of losses (Millions of dollars)	594.895	114.625	408.972	521.044

Table 1b: Statistics on operational losses according to location of event. Locations have been classified in three categories: United States, Canada, Europe and Other countries.

	Location			
	United States	**Canada**	**Europe**	**Other countries**
Average of losses (Millions of dollars)	38.424	8.964	75.204	162.512
Number of losses	635	42	177	202
Standard deviation of losses (Millions of dollars)	147.215	14.987	168.605	1,148.548

Table 1c: Statistics of the total assets of financial institutions according to location of event. Locations have been classified in four categories: United States, Canada, Europe, and Other countries.

	Location			
	United States	**Canada**	**Europe**	**Other countries**
Average of assets (Millions of dollars)	272,673.036	220,107.561	357,057.401	199,247.953
Standard deviation of assets (Millions of dollars)	277,926.188	288,945.739	283,704.460	225,372.722

Table 1d: Statistics of operational losses according to business lines in which the losses occurred. We have selected the classification proposed by Basel II, including 8 lines of business: RB, PS, CF, AM, TS, AS, CB, RB*.

	Business lines							
	RB	**PS**	**CF**	**AM**	**TS**	**AS**	**CB**	**RB**
Average of losses (Millions of dollars)	23.225	35.078	87.504	50.549	89.287	29.429	133.333	39.550
Number of losses	174	52	55	56	82	24	268	345
Standard deviation (Millions of dollars)	63.164	79.545	361.277	90.866	280.857	41.791	993.640	131.195

* RB: Retail brokerage; PS: Payment and settlement; CF: Corporate finance; AM: Asset management; TS: Trading and sales; AS: Agency services; CB: Commercial banking; RB: Retail banking.

Table 1e: Statistics of operational losses according to risk types. We use the classification proposed by Basel II, which includes 7 types of risk: DPA, CPBP, EPWS, EF, IF, EDPM, BDSF**.

	Risk types						
	DPA	**CPBP**	**EPWS**	**EF**	**IF**	**EDPM**	**BDSF**
Average of losses (Millions of dollars)	114.597	61.014	11.59	97.856	64.196	67.528	4.993
Number of losses	4	460	57	222	227	80	6
Standard deviation of losses (Millions of dollars)	168.831	179.294	18.502	1074.485	194.352	306.819	4.022

DPA: Damage to physical assets; CPBP: Clients, products, and business practices;** EPWS: Employment practices and workplace safety; EF: External fraud; IF: Internal fraud; EDPM: Execution, delivery, and process management; BDSF: Business disruption and system failures.

Table 2: Results from the Estimation of the Linear Regression Parameters

The following table presents the results obtained from estimating the linear regression coefficients with the ordinary least squares method. The figures in italics are the P-value statistics

Variable	Basic model
Constant	0.799 *(0.273)*
Log (assets)	0.077*** *(0.005)*
US	-0.541*** *(0.000)*
Canada	-1.087*** *(0.000)*
Europe	-0.053 *(0.755)*
Retail brokerage	0.115 *(0.641)*
Corporate finance	0.386 *(0.199)*
Asset management	0.398 *(0.184)*
Trading and sales	0.417 *(0.140)*
Commercial banking	0.797*** *(0.001)*
Retail banking	0.043 *(0.858)*
Agency services	0.478 *(0.213)*
Damages to physical assets	1.300 *(0.196)*
Clients, products, and business practices	1.123* *(0.079)*
Employment, practices and workplace safety	0.088 *(0.895)*
External fraud	0.360 *(0.578)*
Internal fraud	0.731 *(0.257)*
Execution, delivery, and process management	0.577 *(0.381)*
F(18, 1037)	**8.38** ***(0.000)***
R^2	**12.07%**
R^2_{adj}	**10.63%**

*** Coefficient significant at the 99% confidence level.
** Coefficient significant at the 95%à confidence level.
* Coefficient significant at the 90% confidence level.

Table 3: Robustness Tests

This table presents robustness tests. Models (1) to (3) are used to test the stability of each category of variables in the basic model. Model (4) contains only the non-null significant variables in the scaling model. The figures in italics are P-value statistics.

	Model 1	**Model 2**	**Model 3**	**Model 4**
Constant	1.495*** *(0.000)*	1.797*** *(0.000)*	1.713*** *(0.000)*	1.379*** *(0.000)*
Log (assets)	0.076** *(0.007)*	0.084*** *(0.003)*	0.077*** *(0.006)*	0.082*** *(0.001)*
US		-0.583*** *(0.000)*	-0.446** *(0.001)*	-0.595*** *(0.000)*
Canada		-1.219*** *(0.000)*	-1.055*** *(0.000)*	-1.102*** *(0.000)*
Europe		0.020 *(0.907)*	0.083 (0.624)	
Retail brokerage			-0.275 *(0.271)*	
Corporate finance			0.423 *(0.166)*	
Assets management			0.369 *(0.225)*	
Trading and sales			0.148 *(0.599)*	
Commercial banking			0.455* *(0.062)*	0.665*** *(0.000)*
Retail banking			-0.200 *(0.398)*	
Agency services			0.383 *(0.327)*	
Clients, products and business practices				0.633*** *(0.000)*
F	**7.6** ***(0.000)***	**12.91** ***(0.000)***	**8.39** ***(0.000)***	**22.84** ***(0.000)***
R^2	**0.69%**	**4.68%**	**8.12%**	**9.81%**
R^2_{adj}	**0.60%**	**4.32%**	**7.16%**	**9.38%**

*** Coefficient significant at the 99% confidence level.
** Coefficient significant at the 95% confidence level.
* Coefficient significant at the 90% confidence level.

Table 4: Statistics on Scaled Losses

The second column of this table presents statistics pertaining to losses actually observed in the Merrill Lynch bank. In the third column, we also present statistics calculated on the amounts of external losses scaled to the bank in question. Appendix 1 shows in detail how the scaling was done.

	Observed losses at Merrill Lynch	Scaled external losses to Merrill Lynch
Average ($M)	38.868	35.359
Median ($M)	11.053	7.941
Standard deviation ($M)	83.106	84.298
Kurtosis coefficient	21.112	35.733
Skewness coefficient	4.282	5.272
Minimum ($M)	1.081	0.591
Maximum ($M)	506.154	902.126
Number of losses	52	1,050

Table 5: Impact of Scaled Variables on Loss amounts

In this table, we present a loss event extracted from the external base, along with equivalent scaled losses. In each case, we modify a scaling variable to observe its impact on the loss amount.

Bank	Loss amount	Total assets	Location	Business line	Risk type
Event extracted from external base					
Bank of New York	**$8.26 M**	$48,879 M	United States	Retail banking	External fraud
Scaled event					
Fictional bank 1	**$9.27 M**	***$200,000 M***	United States	Retail banking	External fraud
Fictional bank 2	**$4.97 M**	$48,879 M	***Canada***	Retail banking	External fraud
Fictional bank 3	**$16.06 M**	$48,879 M	United States	***Commercial banking***	External fraud
Fictional bank 4	**$15.56 M**	$48,879 M	United States	Retail banking	***Clients, products and business practices***
Fictional bank 5	**$5.58 M**	***$200,000 M***	***Canada***	Retail banking	External fraud
Fictional bank 6	**$10.86 M**	***$200,000 M***	***Canada***	***Commercial banking***	External fraud
Fictional bank 7	**$20.46 M**	***$200,000 M***	***Canada***	***Commercial banking***	***Clients, products and business practices***

Table 6: Descriptive Statistics on Variables Introduced in the Frequency Model

The following tables present statistics on the distribution of the number of losses per bank, according to the financial institution's size and the geographical distribution of its activities over the 1994-2004 period.

Table 6a: Statistics calculated on the number of losses and the total assets per bank over the 1994-2004 period.

	Average	Standard deviation	Median	Minimum	Maximum
Number of losses per bank	3.27	6.13	1	1	52
Total assets per bank	$123,174 M	$181,820 M	$34,003 M	$43 M	$887,515 M

Table 6b: Statistics for the number of losses per bank according to the financial institution's size estimated based on its total assets in millions of dollars. The banks have been classified in three categories: those with under $400,000 M in average assets over the 1994-2004 period; those with between $400,000 M and $800,000 M in average assets; and those of great size with an average of more than $800,000 M over the period in question.

	Total assets (millions of dollars)		
	Assets < 400 000	**400 000 ≤ Assets <800 000**	**Assets ≥ 800 000**
Average number of losses per bank	2.587	8.538	19
Number of banks	293	26	4
Standard deviation of the number of losses per bank	4.647	10.879	18.166

Table 6c: Statistics on the number of operational losses, according to level of geographical distribution of activities. We consider 4 different environments: the United States, Canada, Europe and Other countries. The columns of the following table show whether the activities are concentrated in the same environment or spread over 2, 3 or 4 environments.

	Geographical Distribution			
	One environment	**Two environments**	**Three environments**	**Four environments**
Average number of losses per bank	2	10	7.667	28
Number of banks	275	33	12	3
Standard deviation of the number of losses per bank	2.957	12.661	4.559	13.077

Table 7: Results of the Estimation of the Coefficients in the Frequency Model

The following table presents the results of the estimation of the coefficients in the frequency scaling model. The estimation is made using the maximum likelihood method. The last line gives the result of the Score test or the Lagrange multiplier test, making it possible to compare the two models. The figures in italics represent *t*-Student.

	Truncated Poisson regression model	**Truncated negative binomial regression model**
Constant	-5.876*** *(-15.28)*	-10.439*** *(-8.35)*
Log (assets)	1.176*** *(15.30)*	1.783*** *(8.45)*
United States	1.432*** *(14.20)*	2.000*** *(6.60)*
Canada	0.559*** *(5.51)*	1.721*** *(2.83)*
Europe	0.141* *(1.73)*	-0.111 *(-0.34)*
Other countries	0.191** *(2.43)*	0.457 *(1.52)*
α		4.347 *(1.53)*
Log (likelihood function)	-695.314	-439.414
Score test	34.637	

*** Coefficient significant at the 99% confidence level.
** Coefficient significant at the 95% confidence level
* Coefficient significant at the 90% confidence level

Table 8: Frequency Scaling Results

This table presents an application of the frequency scaling model. We consider a bank whose average in total assets is evaluated at $100, 000 M over the 1994-2004 period. Its activities are divided mainly between the United States and Canada. Below, we determine the parameters of the Poisson distribution and the negative binomial for each bank.

Table 8a: Poisson regression model.

	Constant β_0	Ln (assets) β_1	United States β_2	Canada β_3	Europe β_4	Other countries β_5	**Poisson parameter***
Coefficients	-5.876	1.176	1.432	0.559	0.141	0.191	**7.352***
Variables		Ln (100000)	1	1	0	0	

* $\lambda_i = \exp(\beta_0 + \beta_1 \ln(Assets)_i + \beta_2 US_i + \beta_3 Canada_i + \beta_4 Europe_i + \beta_5 Others_i)$

Table 8b: Negative binomial regression model

	Constant β_0	Ln (assets) β_1	United States β_2	Canada β_3	Europe β_4	Others β_5	Gamma parameter α	**Negative binomial parameters** (r,p)**
Coefficients	-10.439	1.783	2	1.721	-0.111	0.457	4.347	**(0.23; 0.025)****
Variables		Ln (100000)	1	1	0	0		

** $r = \frac{1}{\alpha}$

$$p = \frac{1}{1 + \alpha\, exp(\beta_0 + \beta_1\, ln(Assets)_i + \beta_2 US_i + \beta_3 Canada_i + \beta_4 Europe_i + \beta_5 Others_i)}$$

Appendix 1

In the tables below, we present three examples of losses extracted from the external base and we show in detail how the scaling is done.

Bank	Observed loss ($M)	Year	Total assets ($M)	Location	Business line	Risk type
BankAmerica Corp.	4.70	1994	169,604	United States	Trading and sales	Clients, products and business practices
Crédit Lyonnais	117.95	1997	250,279	Europe	Retail banking	Internal fraud
Bank of Montreal	10.83	2002	157,780	Canada	Commercial banking	Internal fraud

We scale these losses to the Merrill Lynch bank. We retain the business line, the risk type, and the year of loss recorded in the database.

Bank	Scaled loss ($M)	Year	Total assets ($M)	Location	Business line	Risk type
Merrill Lynch	To be determined	1994	**163,749**	**United States**	Trading and sales	Clients, products and business practices
Merrill Lynch	To be determined	1997	**292,819**	**United States**	Retail banking	Internal fraud
Merrill Lynch	To be determined	2002	**447,928**	**United States**	Commercial banking	Internal fraud

To find the equivalent loss, we must use regression (4) to determine the idiosyncratic components of the losses in the external database as well as those at Merrill Lynch.

Bank	Observed loss (external loss) (A)	Idiosyncratic component (external loss) (B)	Idiosyncratic component (Merrill Lynch) (C)	Loss scaled to Merrill Lynch (D) = A × C/B
BankAmerica Corp.	*4.70*	2.78[a]	2.79[d]	*4.69*
Crédit Lyonnais	*117.95*	2.77[b]	1.55[e]	*66.00*
Bank of Montreal	*10.83*	2.73[c]	3.13[f]	*19.59*

[a] 2.79 = exp (0.082 × ln (Assets) – 0.595 × US – 1.102 × Canada + 0.665 × CB + 0.633 × CPBP)
= exp (0.082 × ln (169,604) – 0.595 × 1 – 1.102 × 0 + 0.665 × 0 + 0.633 × 1)

[b] 2.77 = exp (0.082 × ln (Assets) – 0.595 × US – 1.102 × Canada + 0.665 × CB + 0.633 × CPBP)
= exp (0.082 × ln (250,279) – 0.595 × 0 – 1.102 × 0 + 0.665 × 0 + 0.633 × 0)

[c] 1.73 = exp (0.082 × ln (Assets) – 0.595 × US – 1.102 × Canada + 0.665 × CB + 0.633 × CPBP)
= exp (0.082 × ln (157,780) × 0.595 × 0 – 1.102 × 1 + 0.665 × 1 + 0.633 × 0)

[d] 2.79 = exp (0.082 × ln (Assets) – 0.595 × US – 1.102 × Canada + 0.665 × CB + 0.633 × CPBP)
= exp (0.082 × ln (163,749) – 0.595 × 1 – 1.102 × 0 + 0.665 × 0 + 0.633 × 1)

[e] 1.55= exp (0.082 × ln (Assets) – 0.595 × US – 1.102 × Canada + 0.665 × CB + 0.633 × CPBP)
= exp (0.082 × ln (292,819) – 0.595 × 1 – 1.102 × 0 + 0.665 × 0 + 0.633 × 0)

[f] 3.13 = exp (0.082 × ln (Assets) – 0.595 × US – 1.102 × Canada + 0.665 × CB + 0.633 × CPBP)
= exp (0.082 × ln (447,928) – 0.595 × 1 – 1.102 × 0 + 0.665 × 1 + 0.633 × 0)