MPDD WORKING PAPERS

Social and Economic Impact of Disasters: Estimating the Threshold between Low and High Levels of Risk

Clovis Freire

UNITED NATIONS
ESCAP
Economic and Social Commission for Asia and the Pacific

Social and Economic Impact of Disasters: Estimating the Threshold between Low and High Levels of Risk

Clovis Freire

UNITED NATIONS
ESCAP
Economic and Social Commission for Asia and the Pacific

WP/11/15

MPDD Working Papers
Macroeconomic Policy and Development Division

Social and Economic Impact of Disasters: Estimating the Threshold between Low and High Levels of Risk

by Clovis Freire[1]

April 2011

Abstract

Catastrophes caused by natural hazards that hit "without warning" serve as grim reminders of the challenge that governments and civil society face in identifying and protecting the areas that are at risk of extreme events. This paper presents a methodology to estimate the threshold of social and economic impact of disasters that indicate events that were the manifestation of high levels of risk. It shows the result of the application of the methodology to Desinventar dataset, which covers 20 countries/regions, and the change in the level of the threshold in the past forty years. The methodology is expected to contribute to the international effort to identify, assess and monitor disaster risks to allow the effective integration of risk reduction into development strategies.

---

[1] Clovis Freire is Economic Affairs Officer at the Macroeconomic Policy and Development Division (MPDD) of ESCAP.

Social and Economic Impact of Disasters: Estimating the Threshold between Low and High Levels of Risk

Clovis Freire

1. Introduction

Throughout the world, governments and civil society are faced with the challenge of identifying the areas that are at risk of extreme disaster events. Catastrophes such as the earthquake and tsunami that hit Japan in March 2011 or the floods in Pakistan that have submerged almost one-fifth of the country's total land area in 2010 - two recent natural disasters in Asia-Pacific that resulted in large social and economic impact – are reminders for Governments in developing and developed countries alike of the unavoidable need to be prepared by reducing the risk of disasters.

Disaster risk is a function of the hazard (e.g. earthquakes, tropical cyclones, floods, etc) and the vulnerability of people and economic activities to hazards. Many of the root causes of vulnerability to natural hazards can be found in aspects of poverty. In that connection, a major challenge for reducing the risk of disasters is to mainstream risk reduction into development strategies in an effective manner. That is the core element of the Hyogo Framework for Action (HFA), a plan adopted by 168 Governments to pursuit the reduction of our collective vulnerability to natural hazards.

One of the five HFA priorities for action is to identify, assess and monitor disaster risks and enhance early warning. For which one of the key activities is to:

> "*Develop systems of indicators of disaster risk and vulnerability at national and sub-national scales that will enable decision-makers to assess the impact of disasters on social, economic and environmental conditions and disseminate the results to decision-makers, the public and the populations at risk.*" (United Nations, 2005)

Some international initiatives have resulted in the development of global risk indexes for measuring disaster risk and risk management performance. The Disaster Risk Index (DRI) of the United Nations Development Programme (UNDP), in partnership with UNEP-GRID, models the factors influencing levels of human losses from disasters caused by natural hazards for the period 1980-2000. It ranks countries according to the level of risk based on a multi-hazard risk index that was constructed by adding the estimated number of deaths by disasters from different hazard types obtained from the model (Peduzzi et al, 2009). The Hotspot project implemented by Columbia University and the World Bank under the ProVention Consortium aimed at identify places were the risk of disaster-related mortality and economic losses are highest. It assessed the risk of mortality and economic losses on a 5x5 km grid (Dilley, 2006). The Americas Indexing Programme of the Instituto de Estudios Ambientales of the Universidad Nacional de Colombia in partnership with the InterAmerican Development Bank produced a set of composite indexes that represent different aspects of disaster risk or disaster risk management (Cardona, 2006).

A useful categorization of risk of impact of disasters is the distinction between intensive and extensive risk (United Nations, 2009). Intensive risks are associated with major concentrations of vulnerable population and economic assets exposed to extreme hazards and facing high risk, while extensive risks are associated with geographically dispersed exposure of vulnerable people and economic assets to low or moderate intensity hazards and facing low risk of impact of disasters. Such categorization can be used in identifying geographic areas that are subject to higher levels of impact of disaster. The distinction between these two categories based on empirical data, however, has proved to be very difficult. While information on impact caused by disasters, in terms of causalities and economic damage and loss, provide information regarding the realized risk, it does not provide information regarding the actual level of risk. For example, before the catastrophic destruction caused by the cyclone Nargis in Myanmar in 2008, analysis of damage and loss caused by past disasters would not provide any indication of such high intensity of risk to the impact of cyclones in the country.

Statistical analysis of sample of disaster data has been used to estimate the threshold between intensive and extensive risks. One of such methodologies has suggested that disasters that resulted on more than 50 deaths or 500 houses destroyed are manifestations of intensive risk; disasters which impact is below these thresholds are, thus, manifestations of extensive risk (United Nations, 2009).

Using a probabilistic definition of risk, this study presents a methodology that utilizes the concept of Value at Risk (VaR) - a risk measure widely used in financial mathematics and financial risk management - to estimate the threshold between intensive and extensive risk. The application of the methodology to a dataset of covering a large range of impact of disasters suggests that the thresholds in terms of deaths and houses destroyed should be about of half of the threshold mentioned above - 17 deaths and 281 houses destroyed. Analysis of the risk of impact covering the period from 1970 to 2009 and four 10-year periods covering the 70s, 80s, 90s and the last decade suggested that the threshold for the number of deaths has not changed much over the years but the threshold considering houses destroyed has decreased significantly. That may indicate improvements in the disasters risk prevention and mitigation.

This paper proposes the use of VaR as measure of risk instead of the well-known annual expected value of loss (e.g. average killed/year) because, as detailed in the following section, the expected value of loss can jump back and forth depending on the number of events with high impact during the period considered.

This paper is divided in four sections. Section 2 presents the definition of concepts and presents the measure of risk used in the study. Section 3 presents the data sources and the transformations used in the analysis. Section 4 presents the methodology. Section 5 presents the results and Section 6 lists conclusions.

2. Measuring risk

This study adopts the definition of risk proposed by Cardona (2003):

"Risk: the expected number of lives lost, persons injured, damage to property and disruption of economic activity due to a particular natural phenomenon, and consequently the product of specific risk and elements at risk…Thus, risk is the potential loss to the exposed subject or system, resulting from the `convolution' of hazard and vulnerability. In this sense, risk may be expressed in mathematical form as the probability of surpassing a determined level of economic, social or environmental consequences at a certain place and during a certain period of time".

Figure 1 illustrates this concept. It shows the risk as an exceedance probability curve of deaths caused by geological hazards in 20 countries/regions for with data is available in the Desinventar database (hereinafter called Desinventar) covering the period of 1988 to 2007.[2] The horizontal axis presents levels of impact per day, measured as the absolute number of deaths, as the result of disasters caused by geological hazards (e.g. earthquakes, eruption). For each empirical value in the x axis we can find a correspondent value in the vertical axis that indicates the probability of that value being exceeded, thus the risk. The Figure shows that the higher the loss the lower the risk. However, although low, the risk of very high loss is not negligible.

Figure 1. Risk of deaths caused by geological hazards, Desinventar - period 1988-2007
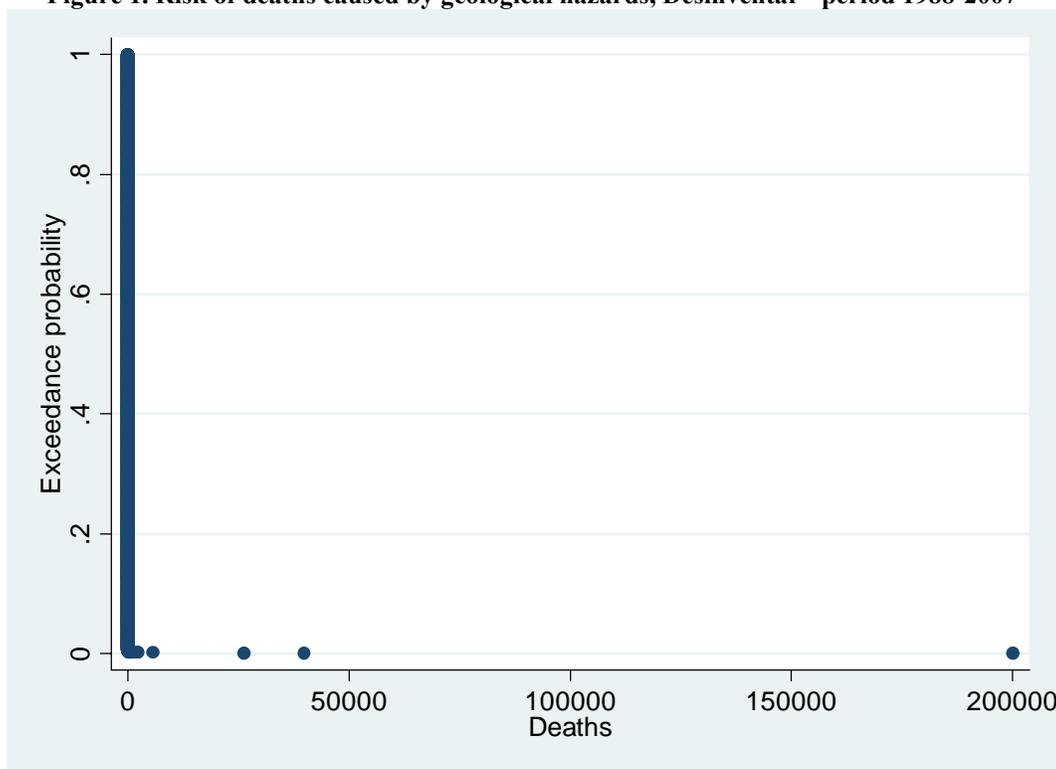
Source: Author's calculations based on data on impact of disasters from Desinventar dataset.

Given a probability distribution of risk such as the one presented in Figure 1, how to express the risk faced by a country or a region? One way would be to present the levels of loss that would be exceeded with certain probability during a certain period of time. For example, one could

---

[2] Countries/regions are: Argentina, Bolivia, Chile, Colombia, Costa Rica, Ecuador, Guatemala, Indonesia, Iran, Jordan, Mexico, Mozambique, Nepal, Orissa/India, Panama, Peru, Salvador, Sri Lanka, Venezuela, and Tamil Nadu/India.

measure the value of loss per day that has 5% of probability of being exceeded. A well-know measure of risk based on this idea is the Value-at-risk (VAR), which could be defined as the maximum loss during a certain period of time within a confidence interval.

Mathematically, given some confidence level α in (0,1) the VaR at the confidence level α is given by the smallest number l such that the probability that the loss L exceeds l is not larger than $(1 - \alpha)$ (McNeil et al, 2005).

$$VaR_\alpha = \inf\{l \in \Re : P(L > l) \le 1 - \alpha\} \qquad (1)$$

Given an empirical distribution function of loss, one can try to fit a theoretical probability distribution function and from that estimate the VaR for any confidence interval α in (0,1). Usually, one is interested in estimating the VaR with a high confidence interval, 95% or more. Invariable, that would require the estimate of the exceedance probability at the fat tail of the density probability distribution.

Figure 2. Probability distribution of deaths caused by geological hazards, Desinventar - period 1988-2007

kernel = epanechnikov, bandwidth = 365.75
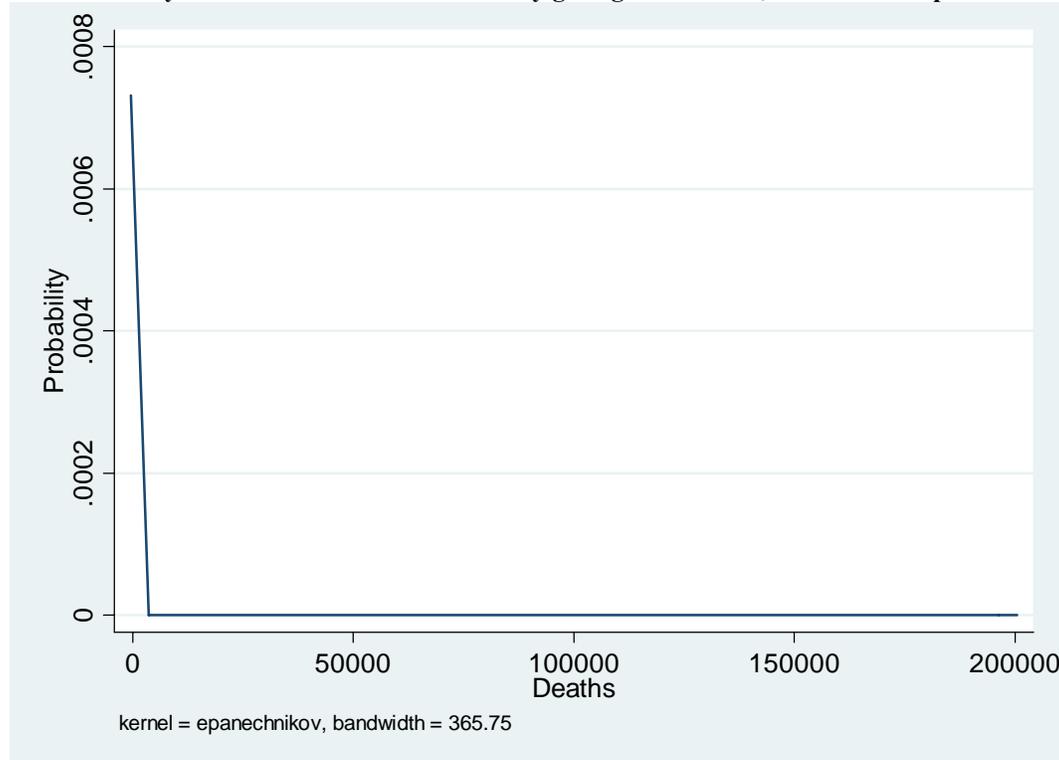
Source: Author's calculations based on data on impact of disasters from Desinventar dataset.

Figure 2 illustrates the fat tail in the case of loss owing to disasters. It shows, for the period from 1988 to 2007, the probability distribution of deaths caused by geological hazards reported in Desinventar. The number in the horizontal axis at the far right side of the graphic represents the highest loss suffered during that period. When the distribution has the characteristic of a fat tail, the expected size of an event larger than any event yet seen is much larger than the largest event experienced to date (Kousky, C, Cooke, R, 2010).

4

One way to observe the fat tail is using mean excess plots. If variable X has cumulative distribution function F then the mean excess curve for X is defined as (Kousky, C, Cooke, R, 2009):

$$G(x) = E(X - x_0 \mid X > x_0) \qquad (2)$$

For a finite ordered sample, $x_1 < x_2, \ldots < x_n$ the sample mean excess plot gives the values:

$$\{x_i, g(x_i)\}; g(x_i) = \sum_{j=i,\ldots n-1} (x_{j+1} - x_i)/(n-i) \quad \text{for } i=1,\ldots n-1 \;\; ; g(x_n)=0 \qquad (3)$$

Figure 3 presents the mean excess plot of deaths caused by geological hazards based on Desinventar data. The positive slopes of these curves indicate that the next extreme event could be much worst than the last one.

Figure 3. Mean excess plots deaths caused by geological hazards, Desinventar - period 1988-2007
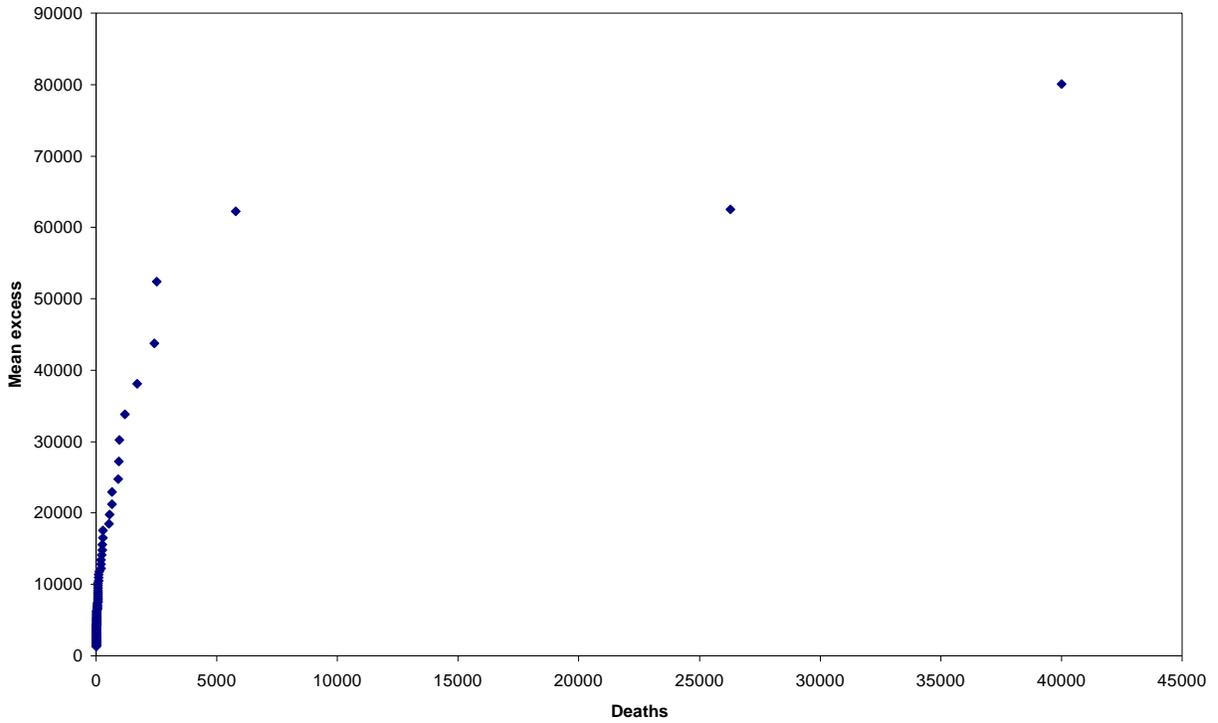
Source: Author's calculations based on data on impact of disasters from Desinventar dataset.

The fat tail behavior can be described by a power-law distribution. A damage variable X follows a power-law distribution if, above a certain lower bound x_{min} the probability that X exceeds x is given by

$$P(X > x) = k\, x^{-\alpha}, \;\; k, \, \alpha > 0, \text{ and } x > x_{min} \qquad (4)$$

Where α is a constant parameter of the distribution known as the exponent or scaling parameter. An important characteristic of the power-law is that if $\alpha \leq 1$, the mean or first moment is infinite. Of course, on N samples from such distribution, the average of the N sample is finite, but it increases with N. If $1 < \alpha \leq 2$, the variance is infinite. The sample mean also has infinite variance no matter how many samples we draw.

Therefore, if the probability distribution of loss owing to a disaster follows a power-law, the average loss (in terms of number of deaths or economic damage and loss) caused by disasters in a given period can not be used as a reliable measure of risk, since the average loss can jump back and forth depending of the period considered.

Figure 4 illustrates this result. Again, using disaster data from Desinventar related to Geological hazards, it shows the number of people killed in from January 1988 to December 2007. If we calculate the average deaths per month in this period we will find it equal to 39 deaths per day. On the other hand, if we were making the same calculation before December 2004, the result would be 3 times lower – 13 deaths per day.
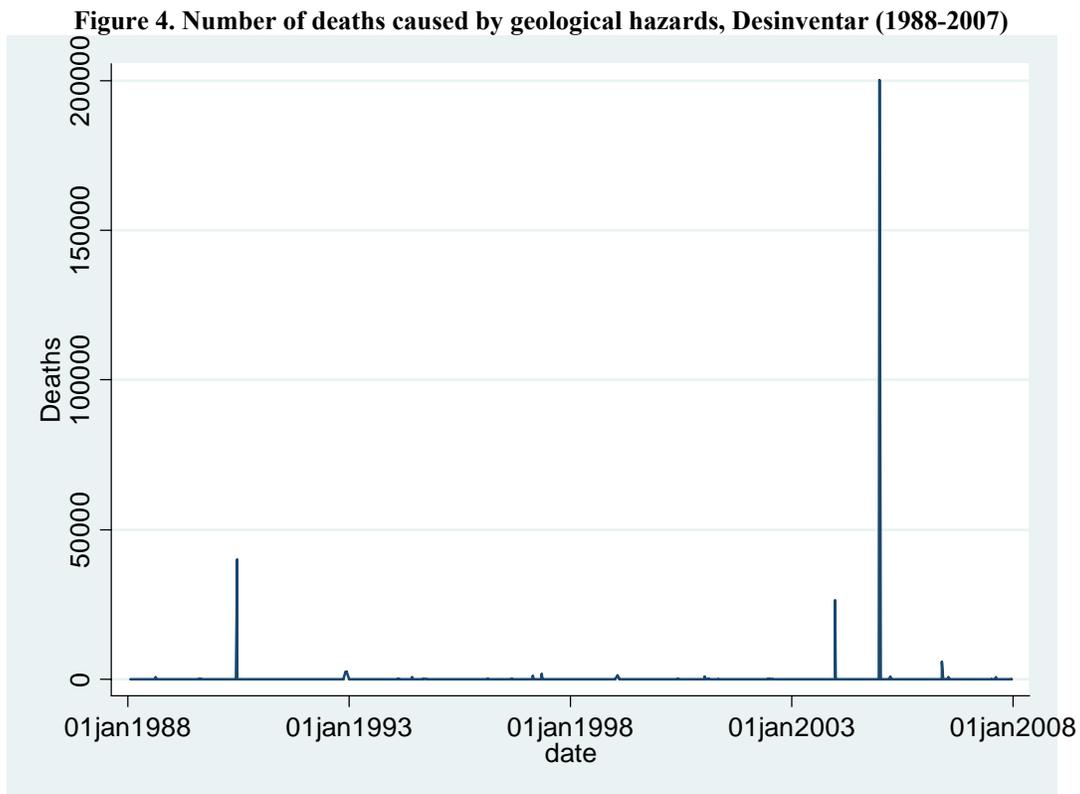
Figure 4. Number of deaths caused by geological hazards, Desinventar (1988-2007)

Source: Author's calculations based on data on impact of disasters from Desinventar dataset.

One may argue that the solution would be to collect more historical data – considering a 100 year period would improve the estimation of the mean, as the argument goes. The problem is that, since the distribution of loss is fat tailed and the analysis indicate that it follows a power-law distribution with scaling parameter with a value lower than 2, when the sample size increases, the average square deviation from the true average keeps growing. In other words, we can collect 100

6

year data but the next large extreme event may be even larger than the last one and it may just happen next year, or tomorrow.

Another argument against the use of more data to estimate the risk is that the process that generate the risk most likely changes with time – owing to socio-economic development, changes in demographics, etc… Therefore, the underlining process that generated the risk 100 years ago is not the same as the one that generated the risk 50 years ago, or 20 years ago.

The VaR could still be used a measure of risk even if the distribution follows a power-law behaviour, by measuring VaR at a confidence interval α such as that the expected maximum loss falls below the x_{min} threshold. How the concept of VaR can be used to estimate the threshold between intensive and extensive risk? The idea is that such threshold is in fact the maximum loss in a country or region during a certain period of time if no extreme events occur (i.e. severe, infrequent hazard events such as major earthquakes, volcanic eruptions and tsunamis as well as severe, cyclical droughts, floods and cyclones). Such threshold, therefore, can be calculated as the VaR in the case that the distribution of risk of impact of disaster follows a power law. The following sections present the source of data and methodology used in the analysis.

3. Data

The source of the data is the DesInventar database accessed on 2 November 2010. It was used the data related to the following countries/regions: Argentina, Bolivia, Chile, Colombia, Costa Rica, Ecuador, Guatemala, Indonesia, Iran, Jordan, Mexico, Mozambique, Nepal, Orissa/India, Panama, Peru, Salvador, Sri Lanka, Venezuela, and Tamil Nadu/India. The information used was date of the disaster, the type of event (i.e. geological or hydro-met) and the number of deaths and houses destroyed.

The analysis has used the whole DesInventar dataset (1970-2009) and subsets covering the period from January 1988 to December 2007, in which data from all countries were available, and four 10-years period covering each decade from 1970 to 2009. The data was aggregated by day and the analysis of the smaller dataset was conducted considering the different types of event separately and combined.

The summary statistics of the data used in the analysis for the period 1988-2007 is:

a) Geological

Variable	Obs	Mean	Std. Dev.	Min	Max
deaths	3154	91.83164	1898.751	0	77804
houses destroyed	3154	376.7676	3103.863	0	96576

b) Hydro-met

Variable	Obs	Mean	Std. Dev.	Min	Max
deaths	110021	.6692086	24.31822	0	5880
houses destroyed	110021	26.54638	940.8488	0	228800

c) Geological & Hydro-met

Variable	Obs	Mean	Std. Dev.	Min	Max
deaths	113175	3.209755	318.1848	0	77804

```
houses destroyed  |    113175    36.30646     1064.072         0     228800
```

4. Methodology

The methodology used to assess the risk of impact of disasters followed to great extend the methodology adopted by Push (2004) and World Bank, ISDR and CAREC (2009). That methodology has five steps:
1) Assemble the disaster's impact data by categories;
2) Sort the data in descending order;
3) Determine by non-parametric methods the probability of each event;
4) Fit a probability curve to determine the probability density curve; and
5) Determine the annual damage by integrating under the curve.

The difference between the methodology used in this study and the one used by Push (2004) and World Bank, ISDR and CAREC (2009) is the following:
- It tests if a power-law distribution fits the empirical distribution. The idea is to refrain to estimate the exceedance probability of impacts in the part of the probability exceedance function that may follow a power-law.
- It uses of maximum likelihood method for fitting the theoretical probability function instead of regression. The cumulative distribution function must take the maximum value of 1. Ordinary linear regression, however, does not incorporate such constraints and hence, in general, the regression line does not respect them.
- It uses of a goodness-of- fit tests such as Kolmogorov-Smirnov (KS) statistic or Bayesian information criterion (BIC) to select the best-fit distribution instead of averaging the estimates obtained by different distributions using weights based on the R2 of the regression analysis.
- It estimates the confidence interval of the measure of risk.

The methodology adopted in this study has seven steps:
1) Assemble the disaster's impact data by categories;
2) Try to identify the power-law behavior by estimating the scaling parameter (alpha) and the lower bound (xmin);
3) Test the goodness of fit of the power-law (Pareto distribution) when compared with other candidate distributions **for x ≥ xmin;**
4) Estimate the confidence level of the VaR calculated **for 0 ≤ x < xmin;**
5) Find the distribution that better fit the data **for 0 ≤ x < xmin;**
6) Calculate the VaR using the distribution that better fit the data **for 0 ≤ x < xmin; and**
7) Estimate the upper and lower confidence limits of the estimated VaR with 90% confidence.

Step 1. Assemble the disaster's impact data by categories

Data is assembled in absolute values of deaths or houses destroyed per disaster per day and per type of event (i.e. geological and hydro-met).

8

Step 2. Identify the power-law behavior

Try to fit a power-law distribution by estimating the scaling parameter and the lower bound.

If the value of x_{min} is known, the scaling parameter α can be estimated by direct numerical maximization of the logarithm of the likelihood function:

$$L = -n \ln(\alpha - 1) - n \ln x_{min} - \alpha \sum_{i=1}^{n} \ln \frac{x_i}{x_{min}}$$

(5)

To estimate the value of x_{min}, this study uses the method proposed by Clauset et all (2009): to choose the value of x_{min} that makes the probability distributions of the empirical data and the best-fit power law model as similar as possible above x_{min}. The measure used to quantify how similar the distributions are is the Kolmogorov-Smirnov or KS statistic, which is the maximum distance between the cumulative distribution functions of the data $F_n(x)$ and the fitted model $F(x)$:

$$D = \max | F_n(x) - F(x) | \quad , \quad x \geq x_{min}$$

(6)

Step 3. Compare with competing distributions

To verify if the best fit power-law provides a better fit for $x \geq x_{min}$ than other distributions, the analysis uses the Schwarz information criterion, also known as Bayesian information criterion (BIC), as a measure of goodness of fit (Vose, 2010):

$$BIC = k \ln(n) - 2\ln(L_{max})$$

(7)

Where n is the number of observations, k is the number of parameters to be estimated, and L_{max} is the maximized value of the log-Likelihood for the estimated model.

Other candidate distributions considered in the analysis are:

- Generalized extreme value
- Lognormal
- Log-logistic
- Inverse Gamma (Pearson Type V)
- Gumbel
- Weibull
- Inverse Gaussian

The description of the cumulative distribution functions of these distributions is presented in Annex 1. The analysis uses the fitted distribution only to interpolate the data and do not use it to

extrapolate to extreme values that were not apparent during the relative short period of observation.

Step 4. Estimate the confidence level of VaR

The maximum confidence level considered in the analysis is 95% - meaning that we try to find the minimum value of loss (Value at Risk) in a period of 20 years such that the probability that of occurrence of loss higher than that VaR is less than 5%.

If from the analysis of the previous step one can not discard the possibility that the distribution follows a power-law for $x \geq x_{min}$, the next step is to find the new confidence level. We use frequency analysis to estimate the frequency of non-exceedance of xmin as if it were drawn from the sample – calculated using the following frequency analysis method:

1) Rank the total number of data (n) in descending order according to their value (x), the highest value first and the lowest value last, including xmin;
2) Assign a serial number (i) to each value $x(x_i, r=1,2,3,...,7305)$, the highest value being x_1 and the lowest being x_n;
3) Divide the rank (i_{xmin}) of xmin by the total number of observations plus 1 to obtain the frequency of non-exceedance:

$$F(x \leq xmin) = 1 - (i_{xmin} /(7305+1)) \tag{8}$$

Step 5. Fit a probability curve for $0 \leq x < xmin$

Use maximum likelihood to try to fit seven candidate distributions and then use BIC to test for the better fit. The seven candidate distributions are the distributions listed under step 3.

Step 6. Estimate VaR for $0 \leq x < xmin$

To calculate the VaR, we calculate the minimum of **xmin** and the estimate of the minimum value of loss for which the cumulative distribution that better fit the data **for $0 \leq x < xmin$** is equal to 1.

Step 7. Estimate the confidence limits

To calculate the confidence limits of the confidence interval of the frequency distributions, the binomial distribution was used given that in case of the cumulative frequency (and the exceedance probability) there are only two possibilities - a certain reference value *X* is exceeded or it is not exceeded.

The standard deviation was calculated from the formula:

$$S_d = \sqrt{ (F * (1-F) / n } \tag{9}$$

Where F is the cumulative probability and n is the number of observations, in this case n=60.

To determine the confidence interval of F we make use of Students' t-test (*t*). For a confidence level of 90% and n=7305, *t*= 1.671.

The lower (L) and upper (U) confidence limits of F in an asymmetrical distribution can be approximated by using F and (1-F) as weight factors:

$$L = F - 2 \ F \ t \ S_d \qquad\qquad (10)$$

$$U = F + 2 \ (1\text{-}F) \ t \ S_d \qquad\qquad (11)$$

We use (9), (10) and (11) to calculate the lower and upper limit of the fitted distribution. The threshold between intensive and extensive risk is taken as the lower limit of the fitted distribution.

5. Results

The analysis resulted in estimates of the thresholds for identification of intensive and extensive risk of Geological and Hydro-meteorological disasters (Table 1). The threshold in terms of numbers of deaths for Geological disasters is 22 and for Hydro-meteorological disasters is 18 causalities. On the other hand, the threshold in terms of houses destroyed is higher for Hydro-meteorological disasters (286 houses destroyed) than for Geological disasters (251 houses destroyed). Applying the methodology to the entire dataset results in thresholds of 17 deaths and 281 houses destroyed.

Table 1. Thresholds of intensive/extensive events

Type of event	Threshold	
	Deaths	Houses destroyed
Geological	22	251
Hydro-met	18	286
Geological & Hydro-met	17	281

Source: Author's calculations based on data on impact of disasters from Desinventar dataset.

Table 2 summarizes the loss reports by hazard type, by disaster impact and by risk manifestation in the period from 1988 to 2007 using the thresholds related to geological and hydro-met events combined – 17 deaths and 281 houses destroyed. The table shows that the majority of the loss reports in the dataset were the manifestation of extensive risk to hydro-meteorological disasters, totaling 108,561 reports or 95.9% of the total. These disasters, however, were not the responsible for the majority of the deaths or damage and economic loss. Geological disasters identified as the manifestation of intensive risk, although accounting for only 0.4% of the reported disasters, accounted for 79.6% of the total deaths in the period. Intensive-risk hydro-meteorological disasters, which accounted for 1.3% of the total disasters, were responsible for 60.2% of total

houses destroyed. These numbers confirms the pattern of higher impact of disasters that were the manifestation of intensive risk.

Table 2. Extensive and intensive loss reports associated with weather-related and geological hazards (Desinventar data – 1988 to 2007)

Risk type	Hazard type	Loss reports	%	Deaths	%	Houses destroyed	%
Extensive	Hydro-met	108,561	95.9	31,587	8.7	445,283	10.8
Extensive	Geological	2,735	2.4	795	0.2	36,640	0.9
Intensive	Hydro-met	1,442	1.3	41,768	11.5	2,474,727	60.2
Intensive	Geological	417	0.4	288,808	79.6	1,151,685	28.0
Total		113,155		362,958		4,108,335	

Source: Author's calculations based on data on impact of disasters from Desinventar dataset.

Figure 5 shows the distribution of mortality across the sample between 1988 and 2007. The horizontal axis is in logarithm scale to make possible to visualize the distribution of mortality associated with extensive risk, which maximum is 10,000 times lower than the maximum deaths associated with intensive risk events in the period considered.

Figure 5. Distribution of mortality associated with extensive and intensive risk across dataset (1988-2007)
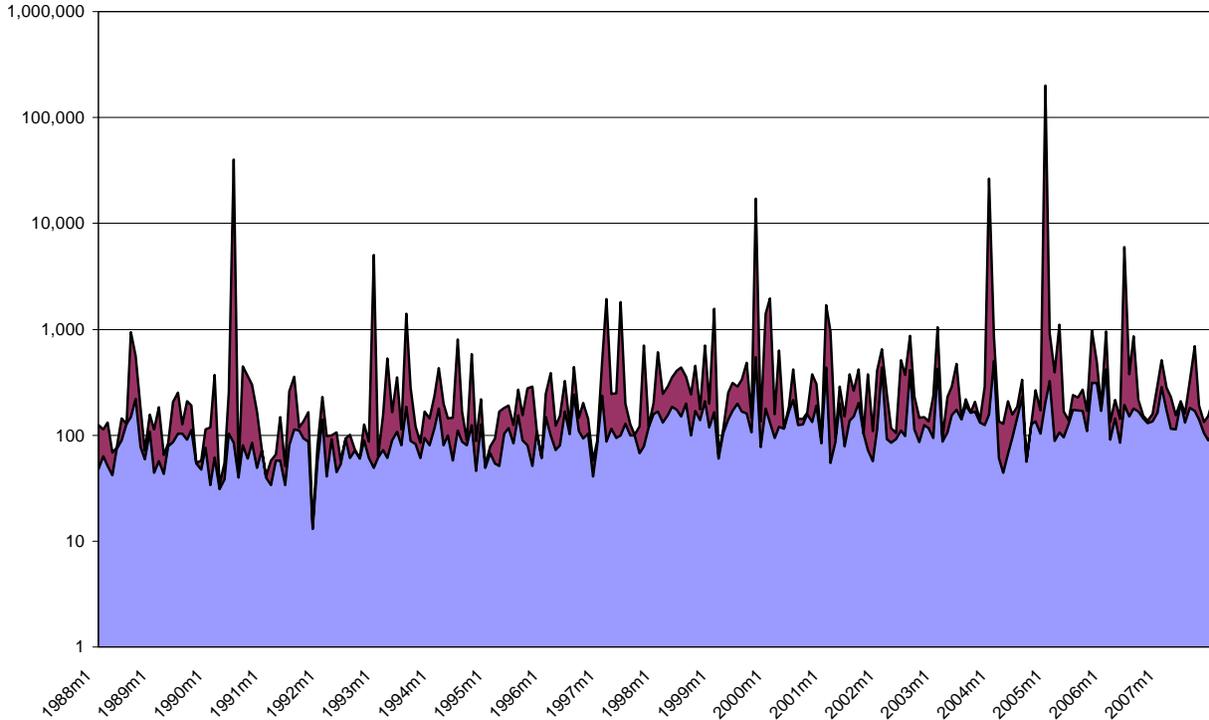
Source: Author's calculations based on data on impact of disasters from Desinventar dataset.

12

 Additional analysis of the risk of impact was conducted using the entire Desinvetar dataset covering the period from 1970 to 2009 and four 10-year periods covering the 70s, 80s, 90s and the last decade. The result of this analysis, which is shown in the table 3, indicates that the threshold for the number of deaths has not changed much over the years – from 17 to 21 deaths depending of the period considered. The threshold considering houses destroyed, however, has decreased significantly from 474 in the 1970s to 94 in the 200s. That may indicate improvements in the disasters risk prevention and mitigation that have contributed to the reduction of the maximum damage and loss in disasters caused by more frequent and less intense hazards.

Table 3. Thresholds of intensive/extensive events for different periods (Geological & Hydro-met)

Period	Threshold	
	Deaths	Houses destroyed
1970-2009	19	397
1988-2007	17	281
1970-1979	19	474
1980-1989	19	393
1990-1999	21	280
2000-2009	18	94

Source: Author's calculations based on data on impact of disasters from Desinventar dataset.

6. Conclusions

This paper used the concept of Value at Risk to estimate the threshold between intensive and extensive risk. Such threshold can be used to identify the areas that have been at risk of extreme disaster events. The application of the methodology to Desinventar dataset, which covers 20 countries/regions in the period of 1988 to 2007, suggests thresholds of 17 deaths and 281 houses destroyed – or about half of the value used as threshold in previous assessments. Analysis of the risk of impact covering the period from 1970 to 2009 and four 10-year periods covering the 70s, 80s, 90s and the last decade suggested that the threshold for the number of deaths has not changed much over the years but the threshold considering houses destroyed has decreased significantly. That may indicate improvements in the disasters risk prevention and mitigation.

Assessment of trends in disaster's impact risks should be taken with caution for many reasons. For example, the assessment is based on historical data that is not complete and it is not homogenous across the years. Future work using this methodology may consider the estimated risk as a function of hazard, exposure and vulnerability (or other combination of these factors since in some models of risk the vulnerability is modeled as exposure and capacity to cope) and based on the estimated level of risk, frequency/intensity of hazard, and exposure, try to estimate the vulnerability. Or one can, based on a model of risk that was constructed, using as dependent variable the estimates of risk from the empirical data, assess what would be the modeled risk.

References

Clauset, C.R. Shalizi, and M.E.J. Newman, "Power-law distributions in empirical data" SIAM Review 51(4), 661-703 (2009). (arXiv:0706.1062)

Cardona, Omar D., 2003. The Notion of Disaster Risk: Conceptual Framework for Integrated Risk Management,IADB/IDEA Program on Indicators for Disaster Risk Management, Universidad Nacional de Colombia, Manizales.

Cardona, Omar D., 2006. A System of Indicators for Disaster Risk Management in the Americas. Measuring Vulnerability to Natural Hazards: Towards Disaster Resilient Societies. Edited by Jorn Birkmann. United Nations University.

Dilley, Max, 2006. Disaster risk hotspots: A project summary. Measuring Vulnerability to Natural Hazards: Towards Disaster Resilient Societies. Edited by Jorn Birkmann. United Nations University.

Kousky, C, and Cooke, R., 2009. The Unholy Trinity: Fat Tails, tail Dependence, and Micro-Correlations. RFF DP 09-36-REV, Resources for the Future. Washington, DC.

Kousky, C, and Cooke, R., 2010. Adapting to Extreme Events: Managing Fat Tails. Issue Brief 10-12, Resources for the Future. Washington, DC.

McNeil, A, Frey, R and Embrechts, P, 2005. Quantitative Risk Management: Concepts Techniques and Tools, Princeton University Press.

Peduzzi, P., Dao, H., Herold, C., Mouton, F., 2009. Assessing global exposure and vulnerability towards natural hazards: the Disaster Risk Index. Natural Hazards and Earth System Sciences.

Pusch, C., 2004. Preventable Lossess: Saving Lives and Property through Hazard Risk Management. Disaster Risk Management Working Paper Series No.9. The World Bank.

United Nations, 2009. 2009 Global Assessment Report on Disaster Risk Reduction: Risk and Poverty in Changing Climate. 2009

Vose, D., 2010. Fitting distributions to data: and why you are probably doing it wrong.

World Bank, the United Nations International Strategy for Disaster Reduction (UNISDR) and CAREC, 2o009. "Central Asia and Caucasus Disaster Risk Management Initiative (CAC DRMI)", available at:
http://www.unisdr.org/preventionweb/files/11641_CentralAsiaCaucasusDRManagementInit.pdf

Annex 1 – Probability Distribution Functions used in the analysis

1) Generalized Extreme Value Distribution
Parameters
k - continuous shape parameter
σ - continuous scale parameter (σ>0)
μ - continuous location parameter

Domain
$$1 + k\frac{(x-\mu)}{\sigma} > 0 \qquad\qquad \text{for } k \neq 0$$
$$-\infty \leq x \leq +\infty \qquad\qquad \text{for } k = 0$$

Cumulative Distribution Function
$$F(x) = \exp(-(1 + kz)^{-1/k}) \quad k \neq 0$$
$$\exp(-\exp(-z)) \qquad k = 0 \text{ (same as Gumbel Maximum Extreme Value Type 1)}$$
where $z \equiv \dfrac{x-\mu}{\sigma}$

2) Log-Logistic Distribution (two parameter)
Parameters
α - continuous shape parameter ($\alpha > 0$)
β - continuous scale parameter ($\beta > 0$)

Domain
$$0 \leq x \leq +\infty$$

Cumulative Distribution Function
$$F(x) = \left(1 + \left(\frac{\beta}{x}\right)^{\alpha}\right)^{-1}$$

3) Lognormal Distribution (two parameter)
Parameters
σ - continuous parameter (σ>0)
μ - continuous parameter

Domain
$$0 < x < +\infty$$

Cumulative Distribution Function
$$F(x) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right)$$

15

where Φ is the Laplace Integral.

4) Pearson Type 5 Distribution (two parameter)
Parameters
α - continuous shape parameter ($\alpha > 0$)
β - continuous scale parameter ($\beta > 0$)

Domain
$0 < x < +\infty$

Cumulative Distribution Function
$$F(x) = 1 - \frac{\Gamma_{\beta/x}(\alpha)}{\Gamma(\alpha)}$$
where Γ is the Gamma Function.

5) Weibull Distribution (two parameter)
Parameters
α - continuous shape parameter ($\alpha > 0$)
β - continuous scale parameter ($\beta > 0$)

Domain
$0 \leq x < +\infty$

Cumulative Distribution Function
$$F(x) = 1 - \exp\left(-\left(\frac{\beta}{x}\right)^{\alpha}\right)$$

6) Inverse Gaussian Distribution (two parameter)
Parameters
λ - continuous parameter ($\lambda > 0$)
μ - continuous parameter ($\mu > 0$)

Domain
$0 < x < +\infty$

Cumulative Distribution Function
$$F(x) = \Phi\left(\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu} - 1\right)\right) + \Phi\left(-\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu} + 1\right)\right)\exp(2\lambda/\mu)$$

where Φ is the Laplace Integral.