



An Autocorrelated Loss Distribution Approach: back to the time series

Dominique Guegan, Bertrand Hassani

► To cite this version:

Dominique Guegan, Bertrand Hassani. An Autocorrelated Loss Distribution Approach: back to the time series. 2012. halshs-00771387v1

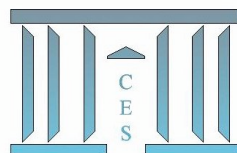
HAL Id: halshs-00771387

<https://shs.hal.science/halshs-00771387v1>

Submitted on 8 Jan 2013 (v1), last revised 9 Feb 2016 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**An Autocorrelated Loss Distribution Approach :
back to the time series**

Dominique GUEGAN, Bertrand K. HASSANI

2012.91



An Autocorrelated Loss Distribution Approach: back to the time series

December 22, 2012

Authors:

- Dominique Guégan: Université Paris 1 Panthéon-Sorbonne, CES UMR 8174, 106 boulevard de l'Hopital 75647 Paris Cedex 13, France, phone: +33144078298, e-mail: dguegan@univ-paris1.fr
- Bertrand K. Hassani: Santander UK and Université Paris 1 Panthéon-Sorbonne CES UMR 8174, 3 Triton Square, NW1 3AN London, United Kingdom, phone: +44 (0)7530167299, e-mail: bertrand.hassani@santander.co.uk

Acknowledgment:

The authors would like to thank the Sloan Foundation for funding their researches.

Abstract

The Advanced Measurement Approach requires financial institutions to develop internal models to evaluate regulatory capital. Traditionally, the Loss Distribution Approach (LDA) is used mixing frequencies and severities to build a loss distribution function (LDF). This distribution represents annual losses, consequently the 99.9th percentile of the distribution providing the capital charge denotes the worst year in a thousand. The traditional approach approved by the regulator implemented by financial institutions assumes the independence of the losses. This paper proposes a solution to address the issues arising when autocorrelations are detected between the losses. Our approach suggests working with the losses considered as time series. Thus, the losses are aggregated periodically and several models are adjusted on the related time series among AR, ARFI and Gegenbauer processes, and a distribution is fitted on the residuals. Finally a Monte Carlo simulation enables constructing the LDF, and the pertaining risk measures are evaluated. In order to show the impact of internal models retained by financial institutions on the capital charges, the paper draws a parallel between the static traditional approach and an appropriate dynamical modelling. If by implementing the traditional LDA, no particular distribution proves its adequacy to the data - as soon as the goodness-of-fit tests reject them - keeping the LDA corresponds to an arbitrary choice. This paper suggests an alternative and robust approach. For instance, for the two data sets explored in this paper, with the introduced time series strategies, the independence assumption is released and the autocorrelations embedded within the losses are captured. The construction of the related LDF enables the computation of the capital charges and therefore permits to comply with the regulation taking into account at the same time the large losses with adequate distributions on the residuals, and the correlations between the losses with the time series processes.

Key words: Operational Risk, Time Series, Gegenbauer Processes, Monte Carlo, Risk Measures.

1 Introduction

In the Advanced Measurement Approach, Basel II/III (BCBS (2001; 2010)) and Solvency II (EP (2009)) accords binds insurance and financial institutions running internal models on loss data sets to evaluate the regulatory capital (Pillar I) pertaining to operational risks. Furthermore, internal models may also be developed by TSA¹ banks to fulfill the capital adequacy exercises.

Operational risk is an overall risk emerging either from internal business failures or external events. Banks are required to track record the losses, and to categorise them following for instance the Basel classification or an internal taxonomy related to the entity risk profile. A loss is characterised at least by a business line, an event type, a amount of money and a date. The loss may take into account recoveries and the date may be the occurrence date, the accounting date, the detection date, etc. To comply with the regulation, financial institutions may opt for different internal model as soon as these are validated by the authorities. To reach this objective, the most interesting point comes from the way institutions use the data sets. In the traditional Loss Distribution Approach (Frachot et al. (2001), Cruz (2004), Chernobai et al. (2007), Shevchenko (2011)), the losses found in a category form a distribution called severity distribution, and the dates enable creating a frequency distribution. The losses are assumed independent and identically distributed (i.i.d.). Then mixing them, an aggregated loss distribution providing the capital requirement as the 99.9th percentile may be constructed. This approach does not take into account correlations between two similar events for two different dates, phenomenon that may naturally arise from the loss generating process (e.g. a hacker who discovered how to steal money from random bank accounts will potentially do it several times until the bank changes the security system). In order to take into account this kind of related events, some banks consider them as single events (by summing them), thus the complete dependence scheme is not taken into account.

Another point of view modelling existing dependences between the losses using time series processes is introduced in this paper. Then the underlying generating process enables creating annually aggregated loss distribution functions for which risk measures are associated. Applying

¹Standard Approach

this approach, our objective is to capture the risks associated to the loss intensity which may increase during crises or turmoil, taking into account correlations, dynamics inside the events, and large events thanks to adequate residual distributions.

Therefore, in order to build a model closer to reality, the assumption of independence between the losses has been released. Thus, a general representation of the losses $(X_t)_t$ is $\forall t$,

$$X_t = f(X_{t-1}, \dots) + \varepsilon_t. \quad (1.1)$$

The function $f(\cdot)$ can take various expressions to model the serial correlations between the losses, and $(\varepsilon_t)_t$ is a strong white noise with a distribution $h(\cdot)$. In the following we focus on two different classes of models. The first captures short term dependences, i.e. AR processes. The second enables modelling long term dependences, i.e. Gegenbauer processes. Denoting B the lag operator, these models may be represented as follows:

1. the AR(p) processes (Brockwell and Davis (1991)):

$$f(X_{t-1}, \dots) = \phi_1 B + \dots + \phi_p B^p \quad (1.2)$$

where $\phi_i, i = 1, \dots, p$ are real parameters to estimate.

2. the Gegenbauer process (Gray et al. (1989), Appendix A):

$$f(X_{t-1}, \dots) = \sum_{j=1}^{\infty} \psi_j \varepsilon_{t-j} \quad (1.3)$$

where ψ_j are the Gegenbauer polynomials which may be represented as follows:

$$\psi_j = \sum_{k=0}^{[j/2]} \frac{(-1)^k \Gamma(d+j-k)(2\nu)^{j-2k}}{\Gamma(d)\Gamma(k+1)\Gamma(j-2k+1)},$$

Γ represents the Gamma function, d and ν are real numbers to be estimated, such that $0 < d < 1/2$ and $|\nu| < 1$ to ensure stationarity. When $\nu = 1$, we obtain the ARFI model, (Guégan (2003), Palma (2007)) or FI(d) model without autoregressive terms.

In the following our approach is not restricted to Gaussian noises $(\varepsilon_t)_t$. More general distributions such as the lognormal, the GEV (Embrechts et al. (1997)), the Weibull, the hyperbolic and the logistic distributions are considered. A dynamic approach has also been suggested by

Baroschia and Bellotti (2012), but in this paper, different models have been proposed to capture autocorrelation phenomena using both short and long memory processes. This paper also details our approach to obtain the regulatory capital.

When the model is fitted a Monte Carlo simulation based on the residuals' distributions enables building the related loss distribution functions used to evaluate the risks measures, i.e. the Capital requirement (VaR) and the Expected Shortfall (Appendix B). These are compared with those obtained implementing the traditional LDA based on the same residuals' distributions. The alternative proposed in this paper to compute the capital requirement of a bank has the interest to give an important place to the existence of dependence between the losses which is not considered with the classical LDA approach usually accepted by regulators. This proposal is important for two reasons. First, in most cases, the goodness-of-fit tests reject fitted distributions on the severities and therefore the LDA cannot be carried out in a robust way. Then, the dynamical approach avoids this uncertainty while capturing the autocorrelation.

In the next section, our methodology is illustrated considering two data sets on which time series processes are fitted with their adequate residuals' distributions. In section three, the capital charges are computed for all the approaches and compared. Section four concludes.

2 Time Series Analysis and Modelling

The previous methodologies have been applied to the following data sets²:

- Execution Delivery and Process Management / Commercial Banking (EDPM/CB), monthly (M) and weekly (W) aggregated
- Client, Product and Business Practices / Retail Banking (CPBP/RB), weekly (W) aggregated.

²The data sets contain losses since 1986, however according to the bank, the collection system was only reliable from 2004. Consequently, all the losses reported before are "remembered" incidents and not the result of a proper detection system.

The data³ have been taken from 1st January 2004 to 31st December 2010. In order to bypass biases engendered by the reporting lag, i.e. the fact that a loss which occurred on a date is reported a few days later, the series have been studied on weekly and monthly horizons⁴. The preliminary statistics presented in Table 1 show that the data sets are right skewed (Skewness > 0) and thick tailed (Kurtosis > 3). The presence of some seasonality has been observed on the data sets, consequently these are filtered to remove the seasonal components by applying a Lowess procedure (Cleveland (1979)). Both AR and ARFI processes have been adjusted on the filtered data. However, because of their higher flexibility, Gegenbauer processes have been adjusted on unfiltered data⁵.

Distribution	Mean	Variance	Skewness	Kurtosis
EDPM / CB (M)	268604.2	308191157365	4.011863	21.32839
EDPM / CB (W)	65387.75	46872497889	7.166842	66.25279
CPBP / RB (W)	126033.2	57442688236	6.183258	56.86517

Table 1: Preliminary statistics for the three data sets. Their distributions are right skewed (Skewness > 0) and fat-tailed (Kurtosis > 3).

The augmented Dickey-Fuller tests (Said and Dickey (1984)) presented in Table 2 provides sufficient evidence to support the stationarity assumption, consequently no transformation has been done on the data to remove a trend.

In a first step, a linear short memory AR model is adjusted on the data sets. The parameters are estimated using least squares approach. According to the AIC (Akaike (1974)), a "perforated" AR(6) for EDPM/CB (W), a "perforated" AR(9) for CPBP/RB (W), and an AR(1) for EDPM/CB (M) should be retained. The estimated values of the parameters denoted ϕ_i are presented in the first line of Table 3, along their standard error highlighting their performance.

³The data has been provided by a European first Tier Bank.

⁴For the data set CPBP/RB, weekly data have only been considered as adjusting the previous models using monthly data was not satisfactory.

⁵The transformation $(I - B^{12})X_t$ and $(I - B^{52})X_t$ are respectively considered for monthly and weekly data. B denotes the lag operator.

Distribution	Test Results
EDPM / CB (M)	Dickey-Fuller = -3.6122, Lag order = 4, p-value = 0.03734
EDPM / CB (W)	Dickey-Fuller = -5.0498, Lag order = 7, p-value = 0.01
CPBP / RB (W)	Dickey-Fuller = -5.4453, Lag order = 7, p-value = 0.01

Table 2: Augmented Dickey-Fuller applied to the three data sets The p-values are inferior to 5%, as a result the series are stationary.

For instance, for the AR(1) the absolute value of the parameter should be higher than 0.2139, for the AR(6), and for the AR(9) these should be higher than 0.1027 at the 5% level. This approach enables capturing the time dependence embedded within the series. The order of the calibrated models on the weekly aggregated series is very high. This fact highlights the presence of long memory in the data and justifies the fitting of models such as Gegenbauer processes or ARFI models.

The second block of Table 3 provides the estimated parameters (with their standard deviation) of the ARFI models adjusted on EDPM/CB (W) and CPBP/RB (W). No reliable fitting has been obtained for this class of models on the monthly data set EDPM/CB (M). On EDPM / CB (W), the long memory component is coupled with a "perforated" AR(5). In this case, the negative long memory parameter d indicates an antipersistent behavior, i.e. some stochastic volatility is embedded inside the data. This phenomenon is difficult to analyse. On CPBP/RB (W) the long memory component is associated to a "perforated" AR(2) model.

An ARFI approach enables capturing both short and long memory behaviors, nevertheless the number of parameters to be estimated and the restrictions on these parameters may prevent a reliable capture of the different phenomena, while Gegenbauer approaches enable capturing both a long memory behavior and a hidden periodicity reducing the number of required parameters. As a result, the related adjustments are presented in the last block of the Table 3. Interesting results have been obtained for the two weekly data sets. The estimated parameter are respectively $d = 0.7735605$ and $u = -0.9242256$ (EDPM/CB) and, $d = 0.8215687$ and $u = -0.7225097$ (CPBP/RB). The results obtained from this parametrisation can be compared to those obtained

from both the AR and the ARFI models considering the AIC⁶. We observe that the latter models correspond to the best fittings in the sense of that criterion. To confirm the adjustment adequacy, the Portmanteau test⁷ is performed to confirm the whiteness of the residuals. The results presented in the third line of each block of Table 3 provide sufficient evidence to retain the three models for each data set. However, the different degrees of whiteness may be used to justify the selection of a model over another. The performance of the associated models in terms of risk measure is discussed in the next section. Combining the results of the AIC and the Portmanteau tests (Table 3 second and third line of each block) the models can be sorted as follows: the AR model for EDPM/CB (M), the ARFI model for EDPM/CB (W) and the Gegenbauer Process for CPBP/RB(W). However, for EDPM / Commercial Banking (W), the use of an ARFI model to characterise the series is questionable as the detected embedded antipersistence parameter ($d < 0$) which may be translated by a stochastic variance of the losses, prevents us from using it. As a result, the second best adjustment will be selected, i.e. the Gegenbauer process.

To use these models for operational risk management purposes, the distributions characterising the residuals have to be identified. It is important to note that the parameters of the previous stochastic processes have been estimated adopting the least square method. Consequently, the residuals are not assumed following any particular distribution. The Jarque-Bera tests (Jarque and Bera (1980)) (Table 3, fourth line of each block) rejecting the Gaussian distribution, various distributions have to be considered to model the residuals.

The philosophy behind time series approaches is different from the traditional LDA, leading to different results. In the LDA the losses are considered independently while using time series, the losses follow a time dependent process and are aggregated accordingly. Consequently, the distributions used to fit the losses (severities) in the LDA should be one-sided, like the lognormal

⁶Associating an ordinary least square estimation procedure in order not to assume a Gaussian distribution for the residual is quite antinomic with the AIC which supposes to have the best likelihood value. However this approach enables limiting the order of the AR processes.

⁷The Portmanteau test enables validating the whiteness of the residuals, and even if in our approach different lags from 5 to 30 with a path of 5 have been tested, it is possible to find non stationary processes by testing some other lags. However, addressing the problem engendered by some assumptions opens the door to a whole new class of approach and risk measurement.

Model		EDPM / CB (M)	EDPM / CB (W)	CPBP / RB (W)
AR	Parameterisation	$\phi_1 = 0.3743$ (0.1023)	$\phi_1 = 0.1207$ (0.0514) $\phi_2 = 0.1074$ (0.0521) $\phi_5 = 0.2386$ (0.0519)	$\phi_1 = 0.1821$ (0.0552) $\phi_9 = 0.1892$ (0.0549)
	AIC	2444.73	9889	9964.2
	Portemanteau	lag/df = 5 Statistic = 1.820355 p-value = 0.8734018	lag/df = 30 Statistic = 13.7347201 p-value = 0.9951633	lag/df = 30 Statistic = 25.4906064 p-value = 0.7008565
	Jarque-Bera (df = 2)	$\chi^2 = 263.5033$ p-value < 2.2e-16	$\chi^2 = 34109.1$ p-value < 2.2e-16	$\chi^2 = 26517.27$ p-value < 2.2e-16
ARFI	Parameterisation	NA	$d = -0.167144$ (0.086078), p-value = 0.0521644 $\phi_1 = 0.292714$ (0.074330), p-value = 8.215e-05 $\phi_2 = 0.156173$ (0.074330), p-value = 0.0356332 $\phi_5 = 0.260554$ (0.074330), p-value = 0.0004559	$d = 0.184673$ (0.086078), p-value = 0.03192 $\phi_2 = -0.089740$ (0.052857), p-value = 0.08955
	AIC	NA	-139.6855	-144.7204
	Portemanteau	NA NA NA	lag/df = 30 Statistic = 12.002588 p-value = 0.9985968	lag/df = 30 Statistic = 31.320582 p-value = 0.3997717
	Jarque-Bera (df = 2)	NA NA	$\chi^2 = 34801.85$ p-value < 2.2e-16	$\chi^2 = 23875.25$ p-value < 2.2e-16
Gegenbauer	Parameterisation	NA	$d = 0.774$ (0.043) $u = -0.924$ (0.092)	$d = 0.822$ (0.067) $u = -0.723$ (0.045)
	AIC	NA	-4 897.046	-6 466.381
	Portemanteau	NA NA NA	lag/df = 30 Statistic = 19.953573 p-value = 0.9177448	lag/df = 30 Statistic = 12.011896 p-value = 0.9985863
	Jarque-Bera (df = 2)	NA NA	$\chi^2 = 4021.289$ p-value < 2.2e-16	$\chi^2 = 14639.36$ p-value < 2.2e-16

Table 3: The table presents the estimated values of the parameters for the different models adjusted on the data sets, with their standard deviation in brackets, and also the results of the AIC criteria, the Portmanteau test and the Jarque-Bera test. The Portemanteau test has been applied considering various lags, and no serial correlation has been found after the different filterings. However, the "whiteness" of the results may be discussed using the p-values. Regarding the p-values of the Jarque-Bera test it appears that the residual distributions do not follow a Gaussian distribution.

or the Weibull distribution as these are defined on $[0, \infty[$, while the residuals' distributions for the time series models are defined on $] - \infty, +\infty[$, and may be chosen among a larger class of

distributions such as the Gaussian, the logistic, the GEV, the Weibull and the hyperbolic⁸. As a result even if the traditional LDA may be interesting to model the severities, assuming the same distribution to characterise the residuals may lead to unreliable risk measures. However, the LDA will be used as a benchmark.

In order to compare the parameters and the adjustment quality, all the distributions previously selected have been fitted on both the losses in case of the LDA approach and on the residuals for the stochastic processes. First, it has been highlighted that no distribution is adequate according to the Kolmogorov-Smirnov test (K-S) for the traditional severity fittings (first column of Tables 4, 5 and 6). This fact has already been highlighted in Guégan and Hassani (2012).

Adopting a time series approach and considering independent monthly and weekly losses, the following distributions have been selected. For EDPM / CB (M), the lognormal distribution should be retained while for EDPM / CB (W) it should be a Weibull distribution and for CPBP / RB (W), a GEV distribution should be selected. Introducing dynamics inside the data sets, even if they correspond to a white noise permits to improve the fitting quality of the models to the data sets⁹. Nevertheless this approach is not sufficient because it does not take into account the existence of correlation between the losses. Thus, more sophisticated models have been considered.

On the three data sets, the previous processes i.e. the AR, the ARFI and the Gegenbauer have been adjusted with a logistic distribution for the residuals (which is the best distribution according to the K-S test), except for the AR process adjusted on the EDPM / CB (M) for which the Gaussian distribution seemed better. The parameters for both the lognormal and the Weibull distributions have also been estimated. However, only the positive data have been considered, consequently the parameters obtained are higher than for the other models. Fitting the GEV distribution on the different data sets led to three kinds of outcomes: for the traditional LDF, the parameters could not be fitted by maximum likelihood (mle) (first column of Tables 4, 5

⁸The estimation of the hyperbolic distribution parameters by maximum likelihood may not be reliable as estimating the four parameters at the same time may lead to convexity problems.

⁹In this particular case, the non-parametric structure of the Kolmogorov-Smirnov test may be favorable to this situation as the number of data points is significantly lower. The lower the number of data points, the larger the chance to reach a statistical adequacy.

and 6); fitting a white noise, in most cases workable parameters have been obtained, but in two cases an infinite mean model has been obtained¹⁰ ($\xi > 1$, Guégan et al. (2011)) (second column of Tables 4, 5 and 6). On the other hand even if the logistic distribution is consistent for all the models (i.e. the AR, the ARFI and the Gegenbauer processes), it does not present a real interest in the traditional LDA. The hyperbolic distribution is theoretically interesting, either the estimation of the parameters did not converge or the goodness-of-fit could not be tested, as a result this solution has not been retained in the next section¹¹.

Theoretically the following models should be retained:

1. For EDPM / Commercial Banking (M): an AR(1) associated to a Gaussian distribution to fit the residuals seems the most appropriate.
2. For EDPM / Commercial Banking (W): as the ARFI model cannot be selected, a Gegenbauer process associated to a logistic distribution has been selected. As no satisfactory fitting has been found, the risk measures obtained from the same process but considering alternative distributions will be compared.
3. For CPBP / Retail Banking (W): a Gegenbauer process associated to a logistic distribution.

The previous selections may be refined taking into account some risk management considerations, and this will be discussed in the following.

3 Risk Measurement and Management based on the Loss Distribution Functions outcomes.

This section describes how risks are measured considering three different approaches: the first one corresponds to the traditional Loss Distribution Approach (Guégan and Hassani (2009; 2012)), the second assumes that the losses are strong white noises (they evolve in time but inde-

¹⁰Another approach such as the method of moments could have been used but this would no be consistent with our model selection strategy as the shape parameter (ξ) estimates would have been constrained to be in $[0, 1]$.

¹¹The Fisher information matrix has been computed for the hyperbolic distributions. However as this solution has not been selected, they are not displayed in the tables in order not to overload them. Nevertheless, they are available on demand.

pendently)¹², and the third filters the data sets using the time series processes developed in the previous sections. In the next paragraphs, the methodologies are detailed in order to associate to each of them the corresponding capital requirement through a specific risk measure.

To build the traditional loss distribution function we proceed as follows. Let $p(k, \lambda)$ be the frequency distribution associated to each data set, $F(x; \theta)$, the severity distribution, then the loss distribution function is given by $G = \sum_{k=1}^{\infty} p(k; \lambda) F^{\otimes k}(x; \theta)$, $x > 0$, with $G = 0$, $x = 0$. The notation \otimes denotes the convolution operator between distribution functions and therefore $F^{\otimes n}$ the n -fold convolution of F with itself. Our objective is to obtain annually aggregated losses by randomly generating the losses. A distribution selected among the Gaussian, the lognormal, the logistic, the GEV and the Weibull is fitted on the severities. A Poisson distribution is used to model the frequencies. As losses are assumed i.i.d., the parameters are estimated by mle.

For the second approach, in a first step, the aggregation of the observed losses provides the time series $(X_t)_t$. These weekly losses are assumed to be i.i.d. and the following distributions have been fitted on the severities: the Gaussian, the lognormal, the logistic, the GEV and the Weibull distributions. Their parameters have been estimated by mle. Then 52 data points have been generated accordingly by Monte Carlo simulations and aggregated to create an annual loss. This procedure is repeated a million times to create a new loss distribution function.

With the third approach the weekly data sets are modelled using an AR, an ARFI and a Gegenbauer process when it is possible. Table 3 provides the estimates of the parameters, and for the residuals a distribution is selected among the Gaussian, the lognormal, the logistic, the GEV and the Weibull distributions. To obtain annual losses, 52 points are randomly generated from the residuals' distributions $(\varepsilon_t)_t$ from which the sample mean have been subtracted, proceeding as follows: if $\varepsilon_0 = X_0$ corresponds to the initialisation of the process, X_1 is obtained applying one of the adjusted stochastic processes (1.2) or (1.3) to X_0 and ε_1 , and so on, and so forth until X_{52} . The 52 weeks of losses are aggregated to provide the annual loss. Repeating this procedure a million of times enables creating another loss distribution function.

¹²This section presents the methodologies applied to weekly time series, as presented in the result section. They have also been applied to monthly time series.

To assess the risks associated to the considered types of incidents and to evaluate the pertaining capital charges, the VaR and the Expected shortfall measures (Appendix B) are used. The results obtained from these different approaches are presented in Table 7 for the cell EDPM / Commercial Banking (M), in Table 8 for the cell CPBP / Retail Banking (W) and in Table 9 the cell EDPM / Commercial Banking (W).

The first remark is that, focusing on the distributions selected before, the adequacy tests may be misleading as the values are not conservative at all. The distributions have been adjusted on the residuals arising from the adjustment of the AR, the ARFI and the Gegenbauer processes. However, to conserve the white noise properties, the mean of the samples has been subtracted from the generated value, therefore, the distribution which should be the best according to the K-S test may not be in reality the most appropriate. As highlighted in Tables 7, 8 and 9, the use of two sided distributions lead to lower risk measures, while one sided distributions lead to more conservative risk measures. Besides these are closer to those obtained from the traditional LDA, meanwhile the autocorrelation embedded within the data has been captured.

It is also interesting to note that there is not an approach always more or less conservative than the others. The capital charge depends on the strategy adopted and the couple - time series process / distribution to generate the residuals - selected. For example a Gegenbauer process associated to a lognormal distribution on CPBP / RB (W) will be slightly more conservative than the traditional approach and enables the capture of time dependency, long memory, embedded seasonality and larger tail. As a result, this may be a viable alternative approach to model the risks. The distribution generating the white noise has a tremendous impact on the risk measures. From Tables 7, 8 and 9, we observe that even if the residuals have an infinite two-sided support, they have some larger tails and an emphasised skewness. Therefore, even if the residuals have been generated using one sided distribution, as the mean of the sample has been subtracted from the values to ensure they remain white noises, the pertaining distributions have only been shifted from a $[0, +\infty[$ support to a $] - \infty, +\infty[$ support. As a result the large positive skewness and kurtosis characteristics of the data have been kept.

Now, comparing Tables 7 and 8, both representing EDPM / CB on different horizons, we observe that the horizon of aggregation may tremendously impact the risk measures, and consequently the capital charges. Adopting either the traditional approach or an AR process (the only available strategy) offsets this problem. For the first approach, it is offset by construction as the approach is not sensitive to the selection of a particular path (weekly or monthly), while for the AR process the autocorrelation has been captured and therefore the values obtained are of the same order.

4 Conclusion

So far, the traditional LDA was the recommended approach to model internal loss data in banks and insurances. However, this approach does neither capture the different cycles embedded in the loss processes nor their autocorrelation. Besides, by modeling the frequency, they may bias the risk measure by double counting some incidents which should be aggregated as they belong to either the same incident or related ones.

As a result, this paper focuses on the independence assumption release of the losses that is used as soon as the traditional LDA is considered, and provides an alternative way ("Taylor Made") of modeling by using time series processes, for instance an AR, an ARFI, or a general Gegenbauer. The paper shows that depending on the data set considered, different models could be considered as they do not capture the same kind of information, and one of them is not always better than the others. Besides, as the processes adjusted should be associated to a distribution characterising the residuals, the differences between the risk measures may be quite significant.

The capital charges are most of the time not comparable, which implies that the capital charge is extremely related to the way the losses are modeled. This result suggests a modeling bias which may in extreme cases threaten the financial institution; if studying the time series, the independence assumption cannot be rejected, the enforcement of an i.i.d. model may be extremely fallacious and in the worst case, far from being conservative enough. Besides, the Pillar 1 capital charge being different, this impacts the Tier 1 and Core Tier 1 ratio, and as a consequence

the capability of the financial institution to generate some more business. Therefore, the way operational risks are modeled may impact the bank strategy.

We understand that for the first Pillar, the result may not lead to conservative capital charges, even if they are legitimate, and this may engender a systemic risk. If while determining the risk taxonomy, the risk identification process fails to capture the loss series behavior described in this paper, the corresponding risk taxonomy may be misleading and the capital charge preposterous. However for the second Pillar, and in terms of management actions, as the measures are different to traditional ones, the key actions to undertake to prevent, mitigate or face these risks may be completely different, and we believe such alternative approach should be considered in the decision making process.

In our point of view, a financial institution which would consider alternative approaches to improve the robustness and the reliability of their risk measures in order to improve the efficiency of their risk management, even if the computed capital charges are relatively lower than those obtained from the traditional approaches, may prove their desire to understand and manage their risks, and may have strong arguments to deal with their regulator.

References

- Akaike, H. (1974), ‘A new look at the statistical model identification.’, *IEEE Transactions on Automatic Control* **19**, 716–723.
- Baroscia, M. and Bellotti, R. (2012), ‘A dynamical approach to operational risk measurement’, *Working paper, Università degli Studi di Bari*.
- BCBS (2001), ‘Working paper on the regulatory treatment of operational risk’, *Bank for International Settlements, Basel*.
- BCBS (2010), ‘Basel iii: A global regulatory framework for more resilient banks and banking systems’, *Bank for International Settlements, Basel*.
- Brockwell, P. J. and Davis, R. D. (1991), *Time Series: Theory and Methods*, Springer, New York.
- Chernobai, A., Rachev, S. T. and Fabozzi, F. J. (2007), *Operational Risk: A Guide to Basel II Capital Requirements, Models, and Analysis*, John Wiley & Sons, New York.
- Cleveland, W. (1979), ‘Robust locally weighted regression and smoothing scatterplots’, *Journal of the American Statistical Association* **74(368)**, 829–836.
- Cruz, M. (2004), *Operational Risk Modelling and Analysis*, Risk Books, London.
- Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997), *Modelling Extremal Events: for Insurance and Finance*, Springer, Berlin.
- EP (2009), ‘Directive2009/138/ecoftheeuropeanparliamentandofthecouncilof25november 2009 on the taking-up and pursuit of the business of insurance and reinsurance (solvency ii) text with eea relevance.’, *Official Journal* **L335**, 1–155.
- Frachot, A., Georges, P. and Roncalli, T. (2001), ‘Loss distribution approach for operational risk’, *Working Paper, GRO, Crédit Lyonnais, Paris*.
- Gray, H., Zhang, N. and Woodward, W. (1989), ‘On generalized fractional processes’, *JTSA* **10**, 233–257.
- Guégan, D. (2003), *Les chaos en finance. Approche statistique*, Economica, Paris.

- Guégan, D. and Hassani, B. (2012), ‘Operational risk: A basel ii++ step before basel iii.’, *Journal of Risk Management in Financial Institutions* **6(1)**.
- Guégan, D. and Hassani, B. K. (2009), ‘A modified panjer algorithm for operational risk capital computation’, *The Journal of Operational Risk* **4**, 53 – 72.
- Guégan, D., Hassani, B. and Naud, C. (2011), ‘An efficient threshold choice for the computation of operational risk capital.’, *The Journal of Operational Risk* **6(4)**, 3–19.
- Jarque, C. and Bera, A. (1980), ‘Efficient tests for normality, homoscedasticity and serial independence of regression residuals’, *Economics Letters* **6(3)**, 255–259.
- Palma, W. (2007), *Long-Memory Time Series: Theory and Methods*, Wiley, New York.
- Said, S. and Dickey, D. (1984), ‘Testing for unit roots in autoregressive moving average models of unknown order’, *Biometrika* **71**, 599–607.
- Shevchenko, P. (2011), *Modelling Operational Risk Using Bayesian Inference*, Springer-Verlag, Berlin.

A Appendix: Gegenbauer process

Gegenbauer polynomials $C_{(n)}^d(u)$ generalise Legendre polynomials and Chebyshev polynomials, and are special cases of Jacobi polynomials.

Gegenbauer polynomials are characterised as follows:

$$\frac{1}{(1 - 2ux + x^2)^\alpha} = \sum_{n=0}^{\infty} C_n^{(\alpha)}(u)x^n, \quad (\text{A.1})$$

and satisfy the following recurrence relation:

$$C_0^\alpha(x) = 1 \quad (\text{A.2})$$

$$C_1^\alpha(x) = 2\alpha x \quad (\text{A.3})$$

$$C_n^\alpha(x) = \frac{1}{n}[2x(n + \alpha - 1)C_{n-1}^\alpha(x) - (n + 2\alpha - 2)C_{n-2}^\alpha(x)]. \quad (\text{A.4})$$

Considering these polynomials as functions of the lag operator, we can write

$$\frac{1}{(1 - 2uB + B^2)^\alpha} = \sum_{n=0}^{\infty} C_n^{(\alpha)}(u)B^n. \quad (\text{A.5})$$

Remark A.1. Note that if $u = 0$ then we have a $FI(d)$ process.

B Appendix: Risk Measure evaluation

For financial institutions, the capital requirement pertaining to operational risks are related to a VaR at 99.9%. The definition of the VaR is recalled in the following:

Given a confidence level $\alpha \in [0, 1]$, the VaR associated to a random variable X is given by the smallest number x such that the probability that X exceeds x is not larger than $(1 - \alpha)$

$$VaR_{(1-\alpha)\%} = \inf(x \in \mathbb{R} : P(X > x) \leq (1 - \alpha)). \quad (\text{B.1})$$

And we compare these results to those obtained based on the Expected shortfall defined as follows:

For a given α in $[0, 1]$, η the $VaR_{(1-\alpha)\%}$, and X a random variable which represents losses during a prespecified period (such as a day, a week, or some other chosen time period) then,

$$ES_{(1-\alpha)\%} = E(X|X > \eta). \quad (\text{B.2})$$

Distribution	Traditional LDF		Time Series LDF		AR(1)	
	Severity	GoF	Severity	GoF	Residuals	GoF
Gaussian	$\mu = 2998.38, \sigma = 41136.97$ (458.64), (8729.09)	$< 2.2e - 16$	$\mu = 208604.20, \sigma = 555149.70$ (58697.91), (136079.3)	$2.82e - 08$	$\mu = \mathbf{32411.16}, \sigma = \mathbf{538246.7}$ (64730.38), (112277.9)	$\mathbf{0.0004998}$
Lognormal	$\mu = 4.50, \sigma = 2.71$ (0.035), (0.032)	$< 2.2e - 16$	$\mu = \mathbf{11.03}, \sigma = \mathbf{2.09}$ (0.22), (0.20)	$\mathbf{0.1793}$	$\mu = 11.80, \sigma = 1.60$ (0.299), (0.19)	$< 2.2e - 16$
Logistic	$\mu = 477.09, \beta = 2807.19$ (45.91), (30.75)	$< 2.2e - 16$	$\mu = 92569.39, \beta = 88498.89$ (4194.30), (4194.30)	$2.312e - 05$	$\mu = -38184.27, \beta = 146530.70$ (6849.27), (6849.27)	$8.835e - 05$
GEV	NA	NA	$\xi = 2.478, \beta = 329588, \mu = 130660$ (0.134), (2647.96), (6075)	\emptyset	$\xi = 0.112, \beta = 422405, \mu = -165672$ (0.1096), (5931.69), (5932.33)	$1.14e - 10$
Weibull	$\xi = 0.37, \beta = 319.05$ (3.41e - 03), (12.74)	$< 2.2e - 16$	$\xi = 5.26e - 01, \beta = 555250$ (4.896e - 02), (4.195e + 03)	$1.562e - 08$	$\xi = 6.78e - 01, \beta = 538247$ (1.05e - 01), (8.389e + 03)	$< 2.2e - 16$
Hyperbolic	$\alpha = 0.17, \delta = 36.40,$ $\beta = 0.167, \mu = -1361.43$	$< 2.2e - 16$	NA	NA	$\alpha = 3.295e - 06, \delta = 1180.20,$ $\beta = 2.798e - 10, \mu = -442.0196$	NA
	OD		NA		OD	

Table 4: The table presents the parameters of the distributions fitted on either the i.i.d. losses or the residuals characterising the EDPM / Commercial Banking monthly aggregated. Traditional LDF denotes Frequency \otimes Severity, Time Series LDF characterises the second approach considering i.i.d. monthly losses and AR denotes the autoregressive process. The standard deviations are provided in brackets. The Goodness-of-Fit (GoF) is considered to be satisfactory if the value is $> 5\%$. If it is not, then the distribution associated to the largest p-value is retained. The best fit per column are presented in bold characters. Note: NA denotes a model "Not Applicable", and OD denotes results which may be provided "On Demand".

Distribution	Traditional LDF		Time Series LDF		AR		Gegenbauer	
	Severity	GoF	Severity	GoF	Residuals	GoF	Residuals	GoF
Gaussian	$\mu = 2998.38, \sigma = 41136.97$ (458.64), (8729.09)	$< 2.2e - 16$	$\mu = 62001.79, \sigma = 213429.81$ (10628.2), (49148.27)	$< 2.2e - 16$	$\mu = -1.45e - 12, \sigma = 194850.90$ (10147.4), (35651.06)	$< 2.2e - 16$	$\mu = -16548.24, \sigma = 120414.5$ (8604.64), (22011.79)	$3.09e - 11$
Lognormal	$\mu = 4.72, \sigma = 2.21$ (0.035), (0.032)	$< 2.2e - 16$	$\mu = 8.97, \sigma = 2.25$ (0.12), (0.09)	$< 2.2e - 16$	$\mu = 10.30, \sigma = 1.54$ (0.13), (0.09)	$< 2.2e - 16$	$\mu = 8.63, \sigma = 2.50$ (0.31), (0.17)	$< 2.2e - 16$
Logistic	$\mu = 477.09, \beta = 2807.19$ (45.91), (30.75)	$< 2.2e - 16$	$\mu = 51948.66, \sigma = 55350.80$ (0.001), (2097.15)	$< 2.2e - 16$	$\mu = -10882.70, \beta = 42638.36$ (2568.48), (1482.91)	0.00244	$\mu = -10882.70, \sigma = 42638.36$ (2568.48), (1482.91)	2.03e - 09
GEV	NA	NA	$\xi = 2.24, \beta = 4254.05,$ $\mu = 1808.99$ (0.17), (625.59), (284.34)	\emptyset	$\xi = -1.16e - 02, \beta = 1.55043e + 05,$ $\mu = -83216$ (1.51e - 02), (42.5), (2098)	$1.31e - 14$	$\xi = -0.18, \beta = 93890,$ $\mu = -71000$ (2.003e - 06), (374.7), (891.8)	$1.965e - 14$
Weibull	$\xi = 0.37, \beta = 319.05$ (3.41e - 03), (12.74)	$< 2.2e - 16$	$\xi = 4.63e - 01, \beta = 2.1343e + 04$ (1.78e - 02), (1779)	0.03658	$\xi = 0.591, \beta = 194851$ (3.94e - 02), (4207)	$< 2.2e - 16$	$\xi = 0.386, \beta = 120414$ (4.11e - 02), (5959)	$< 2.2e - 16$
Hyperbolic	$\alpha = 0.17, \delta = 36.40,$ $\beta = 0.167, \mu = -1361.43$ OD	$< 2.2e - 16$	NA	NA	$\alpha = 1.163515e - 05, \delta = 892.57,$ $\beta = 1.368229e - 06, \mu = -20950.96$ OD	NA	NA	NA

Table 5: The table presents the parameters of the distributions fitted on either the i.i.d. losses or the residuals characterising the EDPM / Commercial Banking weekly aggregated. Traditional LDF denotes Frequency \otimes Severity, Time Series LDF characterises the second approach considering i.i.d. weekly losses, AR denotes the autoregressive process, and Gegenbauer denotes the related model. The standard deviations are provided in brackets. The Goodness-of-Fit (GoF) is considered to be satisfactory if the value is $> 5\%$. If it is not, then the distribution associated to the largest p-value is retained. The best fit per column are presented in bold characters. Note: NA denotes a model "Not Applicable", and OD denotes results which may be provided "On Demand".

Distribution	Traditional LDF		Time Series LDF		AR(9)		ARFI(d,2)		Gegenbauer	
	Severity	GoF	Severity	GoF	Residuals	GoF	Residuals	GoF	Residuals	GoF
Gaussian	$\mu = 775.60, \sigma = 13440.39$ (55.77), (1557.07)	$< 2.2e - 16$	$\mu = 125676.6, \sigma = 240290.7$ (12396.71), (49772.55)	$< 2.2e - 16$	$\mu = -3.73e - 12, \sigma = 103951.9$ (10321.18), (31866.62)	$2.716e - 13$	$\mu = 213.66, \sigma = 210162.3$ (10540.58), (33898.01)	$5.551e - 16$	$\mu = -33271.55, \sigma = 187160.6$ (13697.62), (46220.53)	$2.677e - 09$
Lognormal	$\mu = 4.12, \sigma = 2.01$ (0.0147), (0.00998)	$< 2.2e - 16$	$\mu = 10.85, \sigma = 1.55$ (0.081), (0.092)	0.0008751	$\mu = 10.86, \sigma = 1.39$ (0.115), (0.0856)	$< 2.2e - 16$	$\mu = 10.90, \sigma = 1.45$ (0.118), (0.099)	$< 2.2e - 16$	$\mu = 10.28, \sigma = 1.43$ (0.171), (0.138)	$< 2.2e - 16$
Logistic	$\mu = 146.16, \beta = 702.67$ (4.199), (2.813)	$< 2.2e - 16$	$\mu = 117515.05, \beta = 72633.31$ (2097.15), (2097.15)	$< 2.2e - 16$	$\mu = -14762.13, \beta = 52516.15$ (2097.15), (1482.91)	0.2351	$\mu = -22832.232, \beta = 51028.33$ (2965.82), (2097.15)	0.1166	$\mu = -11419.42, \sigma = 38697.341$ (2965.82), (1712.32)	0.1077
GEV	NA	NA	$\xi = 7.37e - 01, \beta = 41154,$ $\mu = 34506$ (6.88e - 02), (1537), (1387)	0.03361	NA	NA	$\xi = -1.74e - 02, \beta = 173140,$ $\mu = -87169$ (1.57e - 02), (1483), (2097)	$< 2.2e - 16$	$\xi = -2.04e - 01, \beta = 111700,$ $\mu = -89320$ (2.003e - 06), (595.2), (2760)	$2.932e - 10$
Weibull	$\xi = 0.441, \beta = 172.3$ (1.459e - 03), (2.066)	$< 2.2e - 16$	$\xi = 0.62, \beta = 51430$ (2.202e - 02), (2326)	$< 2.2e - 16$	$\xi = 0.752, \beta = 199529$ (5.05e - 02), (4201)	$< 2.2e - 16$	$\xi = 0.725, \beta = 210262$ (4.94e - 02), (4199)	$< 2.2e - 16$	$\xi = 0.662, \beta = 186161$ (6.94e - 02), (5974)	$< 2.2e - 16$
Hyperbolic	$\alpha = 569.74, \mu = 9.37e - 04$ $\delta = 3.40e - 06, \beta = 569.74$ OD	$< 2.2e - 16$	NA	NA	$\alpha = 9.8845e - 06, \delta = 1031.03,$ $\beta = -3.088e - 11, \mu = 6.003$ OD	NA	NA	NA	$\alpha = 1.33394e - 05, \delta = 875.97,$ $\beta = -2.746983e - 06, \mu = -1451.36$	NA

Table 6: The table presents the parameters of the distributions fitted on either the i.i.d. losses or the residuals characterising the CPBP / Retail Banking weekly aggregated. Traditional LDF denotes Frequency \otimes Severity, Time Series LDF characterises the second approach considering i.i.d. weekly losses, AR denotes the autoregressive process and both ARFI and Gegenbauer denote their related processes. The standard deviations are provided in brackets. The Goodness-of-Fit (GoF) is considered to be satisfactory if the value is $> 5\%$. If it is not, then the distribution associated to the largest p-value is retained. The best fit per column are presented in bold characters. Note: NA denotes a model "Not Applicable", and OD denotes results which may be provided "On Demand".

Distribution	Traditional LDF		Time Series LDF		AR(1)	
	VaR	ES	VaR	ES	VaR	ES
Gaussian	7 756 475	8 129 947	9 277 896	9 789 675	965 307	1 058 404
Lognormal	15 274 797	28 749 884	173 662 663	349 741 524	3 581 232	5 700 192
Logistic	1 078 214	1 124 771	2 852 054	3 020 592	557 618	634 773
GEV	NA	NA	NA	NA	851 146	1 111 389
Weibull	3 471 022	3 496 938	53 782 255	61 925 147	1 430 910	1 605 585

Table 7: The table presents the Capital charge (VaR) and the Expected Shortfall of the cell EDPM / Commercial Banking monthly aggregated using different distribution either to model i.i.d. losses or the residuals. Traditional LDF denotes Frequency \otimes Severity, Time Series LDF characterises the second approach considering i.i.d. monthly losses and AR denotes the autoregressive process. Note: NA denotes a model not applicable.

Distribution	Traditional LDF		Time Series LDF		AR(5)		Gegenbauer	
	VaR	ES	VaR	ES	VaR	ES	VaR	ES
Gaussian	7 756 475	8 129 947	8 024 234	8 491 073	911 863	990 236	1 418 736	1 583 040
Lognormal	15 274 797	28 749 884	82 762 607	140 772 623	4 481 548	6 575 072	74 316 491	169 973 881
Logistic	1 078 214	1 124 771	4 915 791	5 133 650	366 622	406 527	899 892	973 593
GEV	NA	NA	NA	NA	807 027	865 336	1 235 108	1 340 894
Weibull	3 471 022	3 496 938	7 024 009	7 681 144	1 868 536	2 069 461	38 461 359	52 293 351

Table 8: The table presents the Capital charge (VaR) and Expected Shortfall of the cell EDPM / Commercial Banking weekly aggregated using different distribution either to model i.i.d. losses or the residuals. Traditional LDF denotes Frequency \otimes Severity, Time Series LDF characterises the second approach considering i.i.d. monthly losses, AR denotes the autoregressive process and Gegenbauer is self explanatory. Note: NA denotes a model not applicable.

Distribution	Traditional LDF		Time Series LDF		AR(9)		ARFI(d,2)		Gegenbauer	
	VaR	ES	VaR	ES	VaR	ES	VaR	ES	VaR	ES
Gaussian	10 417 929	10 677 551	11 798 726	12 219 110	641 575	694 887	1 421 140	1 553 661	1 997 072	2 199 513
Lognormal	6 250 108	7 234 355	39 712 897	52 972 896	1 241 192	1 679 228	4 840 470	7 609 198	6 262 572	9 779 669
Logistic	1 606 725	1 637 926	9 057 659	9 307 824	317 715	350 396	628 637	692 862	750 756	830 100
GEV	NA	NA	182 588 314	519 616 925	NA	NA	1 425 307	1 541 832	1 258 450	1 393 567
Weibull	4 168 547	4 198 009	7 497 567	8 046 971	882 146	955 434	2 587 094	2 967 185	5 892 381	6 992 815

Table 9: The table presents the Capital charge (VaR) and Expected Shortfall of the cell CPBP / Retail Banking weekly aggregated using different distribution either to model i.i.d. losses or the residuals. Traditional LDF denotes Frequency \otimes Severity, Time Series LDF characterises the second approach considering i.i.d. monthly losses, AR denotes the autoregressive process and both the ARFI and the Gegenbauer labels are self explanatory. Note: NA denotes a model not applicable.