Original Article

Default modeling of funds using the Generalized Pareto distribution

Received (in revised form): 22 September 2009

Marco Folpmers
(FRM) works for Capgemini Consulting NL. He heads the financial risk management service line and can be reached at marco.folpmers@capgemini.com.

Correspondence: Marco Folpmers, Capgemini Nederland BV, PO Box 2575, 3500 GN Utrecht, the Netherlands
E-mail: marco.folpmers@capgemini.com

ABSTRACT **In this article, a Generalized Pareto distribution is used for the determination of a fund's Probability of Default (PD) and Exposure at Default (EAD). Once hedge fund returns have been fitted to the Generalized Pareto distribution, a fund's PD is determined by its default point, or the asset return that will just wipe out its equity. The same fitted Generalized Pareto distribution can also be used for the determination of the PD of subordinated tranches and their EADs.**

Journal of Derivatives & Hedge Funds (2010) **16,** 116–122. doi:10.1057/jdhf.2010.9

Keywords: managed portfolios; asset return model; extreme value theory; Generalized Pareto distribution; Probability of Default; Exposure at Default

INTRODUCTION

Credit risk models for managed assets are relevant but difficult to implement. The funds aim at an attractive return through strategies and portfolios, which are not always transparent or are impossible to replicate, such as special purpose vehicles (SPVs) containing positions in many asset-backed securities.[1] However, since the portfolio managers rely on leverage in order to show considerable return-on-equity figures, they need loans and, hence, the credit risk assessment of these managed portfolios is a relevant issue. The subprime crisis points out that returns inflated by leverage were often not sustainable. Not only did leverage become costly and scarce because of increasing funding rates, but large losses were realized as the securities' markets became more illiquid, putting many funds out of business.

In this article, we will apply an asset return model to the credit scoring of a managed portfolio. We will show how a Generalized Pareto distribution can provide a good fit for typical returns and how this fit is used for the determination of PD. Subsequently, we will show how EAD can be easily assessed, once this Generalized Pareto fit has been established.

© 2010 Macmillan Publishers Ltd.

APPLYING AN ASSET RETURN PD MODEL TO SPVS

Hedge funds seeking to acquire a leveraged position use ownership of SPVs, which borrow money to purchase a portfolio of assets (equities, loans, credit derivatives and so on), and pass the economic return of such portfolio, including coupon payments, through to the fund.

The funding of the SPV is structured with the help of an equity tranche and several tranches of loans. Managed assets include, but are not limited to, commercial and residential mortgage loans, other securitizations (for example, credit card loans or small and medium enterprises loans), equity and fixed-income pools or derivatives, Collateralized Debt Obligation (CDO) (bundled to create CDO-squared constructions) and funds that are bundled into funds of funds.

PD of the loan that is provided to the SPV could be based on its rating, but often lenders are warned of relying too much on ratings provided by the rating agencies for structured finance vehicles – and that is even before the subprime crisis.[2]

For an independent, internal PD rating for the loans extended to these SPVs, banks have basically three options:

- a qualitative model;
- a bottom-up cash flow model; and
- an asset return model.

A *qualitative model* rates the portfolio manager based on scores on criteria such as location of the fund, nature of the manager's remuneration, assets under management, complexity of instruments traded, and leverage and management style.[3] The weighted sum of these scores is mapped to an initial PD. The final PD is attained after optional overrides that occur if senior managers attach a higher or lower risk to the portfolio manager than is captured by the initial scores. This procedure has several disadvantages. The risk factors, their scoring and their weights are highly subjective. In addition, only one PD is reached for each portfolio manager, whereas more granularity is needed as soon as the SPV is funded with the help of several tranches of loans that differ in their seniority. Even if one agrees about the asset quality and its downside risk, we need a PD for each tranche, which reflects its position in the subordination structure (lower PDs for the senior tranches).

A *bottom-up cash flow model* determines the incoming cash flows for each single asset or bundle of similar assets of the SPV. These future cash flows can only be predicted with the help of underlying models of the mechanics for each specific instrument and with the help of a simulation engine as the cash flows are not deterministic but stochastic. For each scenario, the model evaluates whether the incoming cash flows provide sufficient cover for interest and amortization. This approach will provide a separate PD for each tranche. Its main drawback is that it requires a detailed cash flow model for each type of instrument, with a specific parameterization for each single instrument in the portfolio (maturity, optionality and so on). The cash flows are stochastic, and therefore many simulation runs are needed in order to gain an understanding of the dispersion of the asset return outcomes for the SPV.

A middle-of-the-road model is the application of the *asset return model* to portfolio managers, which we will describe here.[3] According to this approach, SPV is modeled as a collection of assets that is funded with the help of equity and liabilities. The asset return is modeled with the help of historical asset returns of comparable return indices, that is, monthly published hedge fund returns of funds with the same strategy, or published real estate returns are used for modeling the return of a Commercial Mortgage Backed Securities SPV.

In the next section, we will first outline how PD can be modeled using the return distribution. Subsequently, we will derive the concomitant EAD of the SPV.

QUANTIFICATION OF PD

From the website of Greenwich Alternative Investments, we downloaded monthly hedge fund returns for 23 hedge fund styles.[4] Our data run from January 2004 to June 2009, that is, 66 months or 5.5 years.

First, we standardize the returns:

$$StdR_t = \frac{r_t - \mu}{\sigma} \tag{1}$$

In our example, we standardize the data with the help of an estimated σ equal to 2.39 per cent and an estimated μ equal to 0.33 per cent.

In Figure 1, we show the empirical probability density function (PDF) of these monthly standardized returns x_t.

As can be concluded from the graph, the returns are obviously non-Gaussian. This implies that we cannot treat these returns as normally distributed for the purpose of default modeling.[5]

As we are only interested in the left tail of the distribution, we will adopt an Extreme Value Theory approach and identify peaks over threshold. We define the threshold at the 20th percentile. For the current distribution, the *Threshold* is located at −0.47 standard deviations.

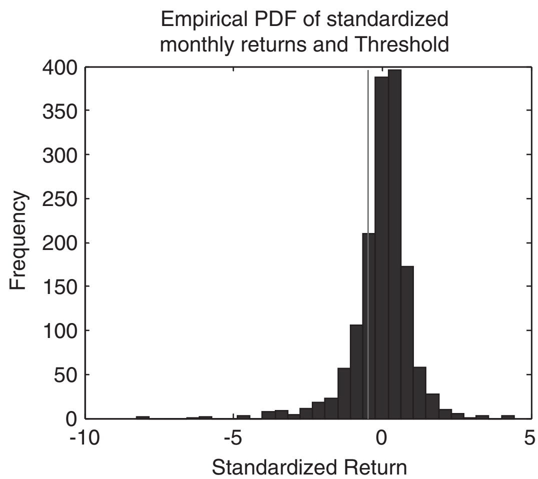

Figure 1: Empirical PDF of standardized monthly returns.

We define x as the absolute value of the peak over this 20th percentile threshold (with $PCT(.)$ for the percentile function):

$$\begin{aligned} x &= -(StdR_t - PCT(StdR_t;20)) \\ &\quad \text{if } StdR_t < PCT(StdR_t;20) \\ x &= .\text{if } StdR_t \geqslant PCT(StdR_t;20) \end{aligned} \tag{2}$$

The distribution of x, that is, the left-tail distribution of *StdR* after mirroring, is depicted in Figure 2. The x vector has 302 elements, which is 20 per cent of the original 1518 observations (=23 hedge fund styles × 66 months).

For the modeling of the exceedances, we use a Generalized Pareto distribution. PDF is equal to

$$f(x;k,\sigma) = \frac{1}{\sigma}\cdot\left(1 + k\cdot\frac{x}{\sigma}\right)^{-(1/k+1)} \tag{3}$$

Its mean equals $\sigma/(1-k)$ with $k<1$.

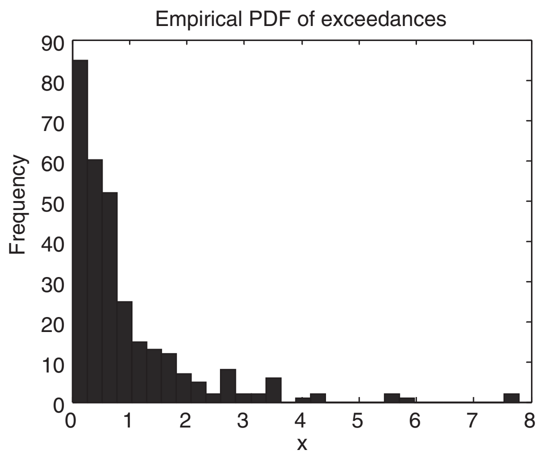

Figure 2: PDF of x (return exceedances).

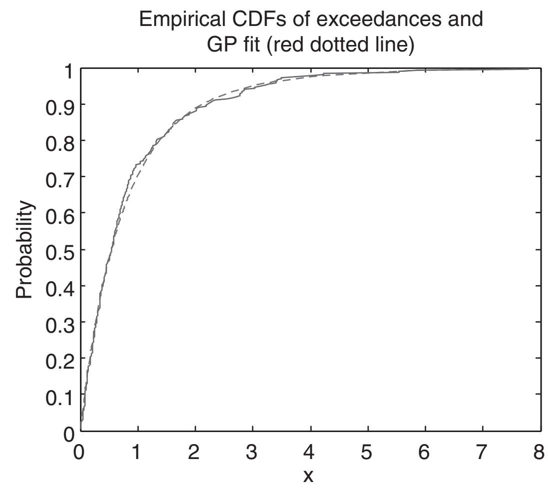

Figure 3: Empirical CDF and fitted Generalized Pareto CDF.

Its Cumulative Distribution Function (CDF) is

$$F(x; k, \sigma) = 1 - \left(1 + k \cdot \frac{x}{\sigma}\right)^{-1/k} \quad (4)$$

In this distribution, k is the tail index and characterizes the shape, and σ is the scaling parameter and is a measure of dispersion. For positive k, the Generalized Pareto distribution is characterized by fat tails.

With the help of Maximum Likelihood estimation,[6,7] we fit a Generalized Pareto distribution to x with parameters: $k = 0.2091$ and $\sigma = 0.7227$.

The empirical CDF is well fitted by this Generalized Pareto distribution, as is shown in Figure 3. This is confirmed with the help of a Kolmogorov-Smirnov test in which we compare the empirical exceedance-vector x with the fitted Generalized Pareto distribution ($P > 10$ per cent).

Suppose now that our specific hedge fund defaults if its standardized return $StdR_i$ is below -3.5. With the help of the threshold of -0.47, the exceedance is in this case equal to 3.03 (the default point $DP_i = -(StdR_i - Threshold)$). The corresponding Generalized Pareto cumulative probability equals 95.06 per cent and the probability that the exceedance exceeds 3.03 equals its complement, 4.94 per cent. As the GP distribution is only applicable in 20 per cent of the cases, the monthly PD_m equals 20 per cent of this 4.94 per cent, and thus 0.99 per cent. This monthly PD is easily converted to a yearly PD (through the survival rate) equal to 11.23 per cent:

$$\begin{aligned} PD_m =& \{1 - F(3.03; k = 0.2091, \\ & \sigma = 0.7227)\} \cdot 0.2 \\ =& (1 - 0.9506) \cdot 0.2 = 0.0494 \cdot 0.2 \\ =& 0.0099 \\ PD_Y =& 1 - (1 - PD_m)^{12} = 11.23\% \end{aligned} \quad (5)$$

We will now turn to the quantification of EAD. This estimate is based on the expected decline in the asset value of SPV once a default occurs.

QUANTIFICATION OF EAD

In order to quantify EAD, we need to derive the expected value of x once we have a default, that is, once the exceedance is larger than the default point (*DP*, in our case 3.03):

$$E(x|x > DP_i) \tag{6}$$

When using the Generalized Pareto distribution, we can prove that this conditional mean is equal to (see the Appendix for the derivation):

$$E(x|x > DP_i) = \frac{DP_i + \sigma}{1 - k} \tag{7}$$

In our example, the conditional mean equals an exceedance equal to 4.74. When including the threshold, we have a standardized return for the default point equal to −5.21. As we standardized our data with a σ equal to 2.39 per cent and a μ equal to 0.33 per cent, we have a conditional expected return equal to −0.1213. This means that our EAD is equal to 12.13 per cent times total assets minus the equity funding.

$$\begin{aligned} StdR_{DP} &= -(E(x|x > DP_i) - Threshold) \\ r_{DP} &= StdR_{DP} \cdot \sigma + \mu \\ EAD_i &= -r_{DP} \cdot Assets - Equity \end{aligned} \tag{8}$$

Otherwise stated, if the asset value declines with 12.13 per cent, this reduction is first accommodated by the equity buffer. Subsequently, the decline will impair the loans to the SPV.

Our example is illustrated in Figure 4, where the probability density and cumulative distribution functions are shown, as well as $g(x; k, \sigma) = f(x; k, \sigma) \cdot x$, which is needed for the calculation of the conditional mean.[8] It shows the outcome of the conditional mean for each value of x as well.

The figure illustrates how an exceedance of 3.03 (the default point) leads to a conditional mean of 4.74 (vertical black dotted line). The 1-month PD can be read from the figure as well: it is the distance between the red line (the CDF) and the $Y = 1$ line depicted as the black horizontal line. As the probability of an exceedance of the threshold equals 20 per cent, this distance has to be multiplied by 0.20 for the 1-month PD.

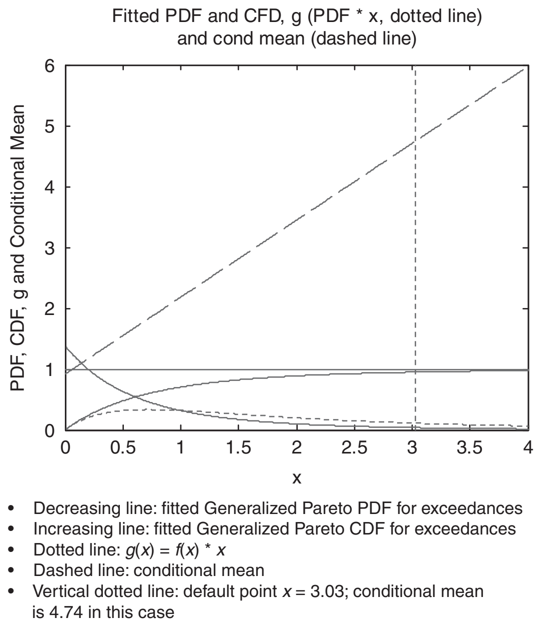

Figure 4: PDF and CDF of x, $g(x)$, conditional mean and default point.

PARTING THOUGHTS

In this article, we have shown how the Generalized Pareto distribution can be applied to return data. Subsequently, the fitted Generalized Pareto distribution is used to determine the PD of a managed portfolio. As the conditional mean of the Generalized Pareto is easily calculated, EAD is established as the asset value that is impaired after subtraction of equity.

The methodology is easily extended to the case in which several tranches of debt need to be rated. For each tranche, a separate default point DP_i should be established. Generally, the senior tranches will have higher default points, and hence lower PDs than the junior tranches. Their EADs equal the asset value that is impaired after subtraction of the equity and the tranches that are subordinated to these senior tranches.

REFERENCES AND NOTES

1 HedgeFundNet. (2009) Strategy Focus Report. August, http://www.hedgefund.net/dailyemailreports/Strategy_Focus_Report_MBS_ABS_Securitized_Credit_Funds.pdf.
2 Fender, I. and Mitchell, J. (2005) *Structured Finance: Complexity, Risk and the Use of Rating*. Basel, Switzerland: BIS Quarterly Review.
3 Christory, C., Daul, S. and Giraud, J.-R. (2007) *Quantification of Hedge Fund Default Risk*. Nice, France: EDHEC.
4 See http://www.greenwichai.com.
5 Escobar, M. and Seco, L. (2008) The mathematics of risk transfer. *International Journal of Services Science* 1(1): 21–35.
6 McNeil, A.J. (1997) Estimating the tails of loss severity distributions using extreme value theory. *ASTIN Bulletin* 27(1997): 117–137, http://www.ma.hw.ac.uk/~mcneil/ftp/astin.pdf.
7 Generalized Pareto distributions are fitted with the help of MLE, see McNeil,[6] p. 8.
8 See the Appendix for a further explanation of $g(x)$ and the derivation of the conditional mean.
9 McNeil, A.J. (2000) Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. *Journal of Empirical Finance* 7: 271–300.
10 McNeil, A.J., Frey, R. and Embrechts, P. (2005) *Quantitative Risk Management: Concepts, Techniques, Tools*. Princeton, NJ: Princeton University Press.

APPENDIX

Derivation of the conditional mean for the Generalized Pareto distribution

The Generalized Pareto's PDF is

$$f(x;k,\sigma) = \frac{1}{\sigma}\cdot\left(1 + k\cdot\frac{x}{\sigma}\right)^{-(1/k+1)} \qquad \text{(A1)}$$

Its mean equals $\sigma/(1-k)$ with $k<1$.

Its CDF is

$$F(x;k,\sigma) = 1 - \left(1 + k\cdot\frac{x}{\sigma}\right)^{-1/k} \qquad \text{(A2)}$$

For the determination of the expected value of the exceedance in case $x > x_1$, we first define $g(x,k,\sigma) = f(x;k,\sigma)\cdot x$. For the area under the curve of g, we need its integral:

$$\int_{x_1}^{\infty} f(x;k,\sigma)\cdot x dx = [G(x;k,\sigma)]_{x_1}^{\infty} \qquad \text{(A3)}$$

Now, we can determine the expected value of the exceedance in case $x > x_1$ (the conditional mean for x if $x > x_1$) as

$$E(x; x > x_1) = \frac{[G(x;k,\sigma)]_{x_1}^{\infty}}{[F(x;k,\sigma)]_{x_1}^{\infty}} = \frac{(1 - G(x_1;k,\sigma))}{(1 - F(x_1;k,\sigma))} \qquad \text{(A4)}$$

By evaluating the indefinite integral $G()$, we arrive at

$$G(x;k,\sigma) = \int f(x;k,\sigma)\cdot x dx = \frac{\sigma\cdot(k-1) + q\cdot(\sigma^2 + k\cdot x^2 + x\cdot\sigma + x\cdot\sigma\cdot k)}{\sigma\cdot(k-1)},$$

with

$$q = \left(\frac{\sigma + k\cdot x}{\sigma}\right)^{-\frac{1+k}{k}} \qquad \text{(A5)}$$

Now, $1-G()$ can be written as

$$1 - G(x;k,\sigma) = \frac{\sigma\cdot(k-1)}{\sigma\cdot(k-1)} - \frac{\sigma\cdot(k-1) + q\cdot\left(\sigma^2 + k\cdot x^2 + x\cdot\sigma + x\cdot\sigma\cdot k\right)}{\sigma\cdot(k-1)} = -\frac{q\cdot\left(\sigma^2 + k\cdot x^2 + x\cdot\sigma + x\cdot\sigma\cdot k\right)}{\sigma\cdot(k-1)} \qquad \text{(A6)}$$

Whereas $1-F()$ can be written as

$$1 - F(x;k,\sigma) = \left(\frac{\sigma + k\cdot x}{\sigma}\right)^{\left(-\frac{1}{k}\right)} \qquad \text{(A7)}$$

By rewriting q as

$$q = \left(\frac{\sigma + k \cdot x}{\sigma}\right)^{-\frac{1+k}{k}} = \left(\frac{\sigma + k \cdot x}{\sigma}\right)^{\left(-\frac{1}{k}\right)} \cdot \left(\frac{\sigma + k \cdot x}{\sigma}\right)^{-1} = \left(\frac{\sigma + k \cdot x}{\sigma}\right)^{\left(-\frac{1}{k}\right)} \cdot \left(\frac{\sigma}{\sigma + k \cdot x}\right) \quad \text{(A8)}$$

We can rewrite $(1-G(x;k,\sigma))/(1-F(x;k,\sigma))$ as

$$\frac{(1 - G(x;k,\sigma))}{(1 - F(x;k,\sigma))} = -\frac{\dfrac{\left(\frac{\sigma + k \cdot x}{\sigma}\right)^{\left(-\frac{1}{k}\right)} \cdot \left(\frac{\sigma}{\sigma + k \cdot x}\right) \cdot \left(\sigma^2 + k \cdot x^2 + x \cdot \sigma + x \cdot \sigma \cdot k\right)}{\sigma \cdot (k-1)}}{\left(\frac{\sigma + k \cdot x}{\sigma}\right)^{\left(-\frac{1}{k}\right)}}$$
$$= -\frac{\left(\frac{\sigma}{\sigma + k \cdot x}\right) \cdot \left(\sigma^2 + k \cdot x^2 + x \cdot \sigma + x \cdot \sigma \cdot k\right)}{\sigma \cdot (k-1)}$$
$$= -\frac{\left(\sigma^2 + k \cdot x^2 + x \cdot \sigma + x \cdot \sigma \cdot k\right)}{(\sigma + k \cdot x) \cdot (k-1)}$$
$$= -\frac{(\sigma + k \cdot x) \cdot (x + \sigma)}{(\sigma + k \cdot x) \cdot (k-1)} = \frac{(x + \sigma)}{(1 - k)} \quad \text{(A9)}$$

Hence, the conditional mean simplifies to

$$E(x|x > x_1) = \frac{x_1 + \sigma}{1 - k} \quad \text{(A10)}$$

This is the same result as the exceedance function of the Generalized Pareto function that gives the expected value of the distribution dependent on the choice of a threshold. The exceedance function is

$$e(u) = E(x - u|x > u) = \frac{\sigma + k \cdot u}{1 - k} \quad \text{(A11)}$$

Therefore,

$$E(x|x > u) = \frac{u + \sigma}{1 - k} \quad \text{(A12)}$$

Notice that exceedances over this higher threshold u are Generalized Pareto distributed themselves with the same shape parameter k but with a different scaling parameter. See, for further analysis, McNeil[9] and McNeil *et al.*[10]