GANs for Realistic and Extreme Events

Research: Steve Antony, Konstantinos Gklantzounis, Haris Vossos

Abstract

The analysis and simulation of extreme events (market crashes, credit crises, climate extreme events, etc.) are important in financial risk management.

In this project, we build a pipeline to generate extreme and realistic natural disaster-related samples using GANs and theorems from the Extreme Value Theory.

Project Introduction

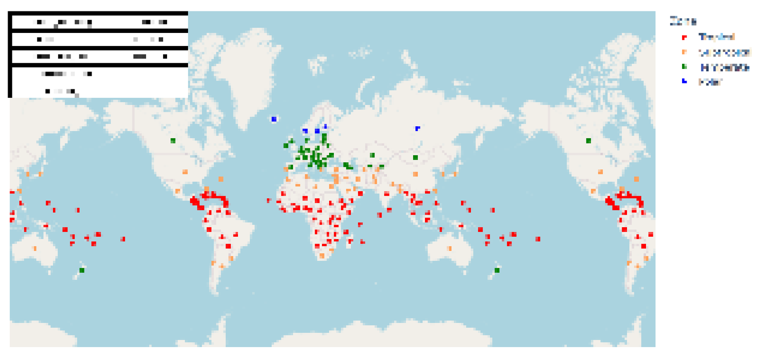
Input

- Financial losses due to extreme natural disaster-related events in various countries around the world.
- Performed pre-processing and post-processing in the EM-DAT dataset to fill in missing values.

Classification into Climate Bands

The available data points per country were not sufficient for training a Generative Adversarial Network. Thus, we classified countries by climate zones and grouped the losses.

Losses Colored by Climate Zones



Normalization: Exceedance-Based Method

Using the Peaks over Threshold (POT) theorem of Extreme Value Analysis that states that the exceedance of the peaks over a threshold follows a Generalized Pareto Distribution (GPD) and setting a user-defined threshold ui, we transition from the country level to the climate band level, creating the regional variable Y:

$Y \sim GPD(1, \gamma, \xi)$, where γ: shape | ξ: scale

GAN Training

GAN Selection

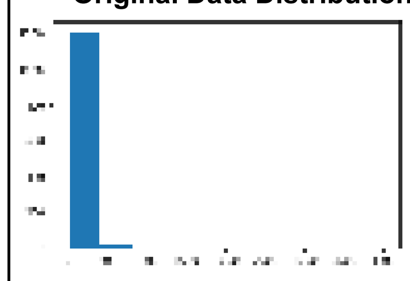
In the process of selecting the optimal GAN, we tested four different GAN architectures, i.e. Quantitative GAN, Deep Convolutional GAN (DCGAN), Temporal Transformer GAN (TTGAN), and Temporal Attention (TAGAN). For each GAN we performed the following steps:

1. Built an architecture for the Generator and Discriminator so that they can generate one-dimensional data because our aim is to generate extreme and realistic losses.
2. Performed batch and epoch optimization.
3. Tested different distribution functions associated with the seed generation (uniform, normal, truncated, exponential, log normal).
4. Applied the 1-Wasserstein distance for performance evaluation.

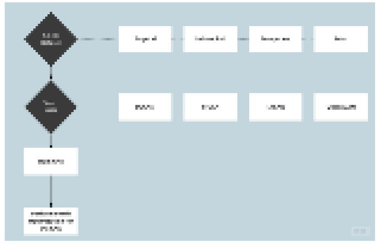
Evaluation

The TAGAN and the QuantGAN did not yield similar results to the original distribution. On the contrary, we managed to produce data that were simulating the original losses on the climate band levels using the architecture of the TTGAN and the DCGAN.

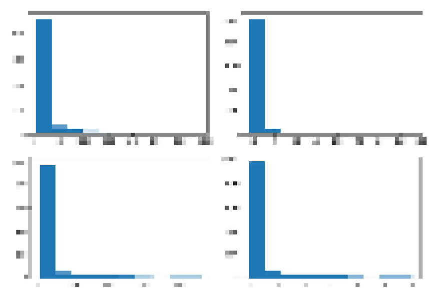
Original Data Distribution



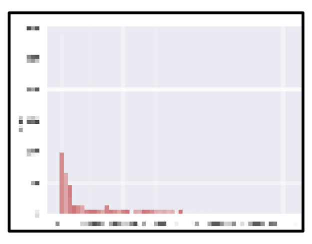
Flowchart



TTGAN Output



DCGAN Output



Distribution Shifting

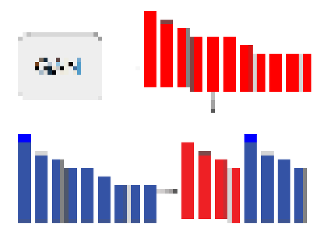
Reason for Implementing Distribution Shifting

To achieve the desired extremeness level and create a dataset that will be used as input for the Extreme Conditional GAN, we perform distribution shifting of our dataset. The distribution shifting works in an iterative manner with two user-defined hyperparameters:

1. c: a threshold representing the quartile to be retained
2. k: number of iterations of distribution shifting

To preserve the size of the original dataset, while removing points with low values, we utilize the previously trained GAN to generate new data points to replace them.

Visual Illustration of Distribution Shifting



ExGAN: Sample Generation

Using the shifted dataset, we create a Generalized Pareto Distribution, which we will use to draw extreme samples for the conditional GAN.

Conditional GAN

- Input: the shifted dataset, transformed to a two-dimensional set: (value, extremeness measure)
- Generator(Z, e): where Z is a standard normal random variable and e is sampled from the GPD of the shifted data set.

Achieving Desired Level of Extremeness

The desired level of extremeness (τ) is user defined. We adjust that to account for the extremeness of the distribution shifting: $\tau' = \tau/c^k$

After training the conditional GAN, we start generating values and keep those that satisfy the extremeness probability (e'). In this step the extremeness probability (e') is drawn from the inverse GPD(1-τ'), where τ' is the adjusted tau.

Revert to Country Level

From Climate Bands Level to Country Level

Using the threshold (ui) of the homogeneous climate bands exceedance-based GPD, we convert the extreme values of the ExGAN back to country level.

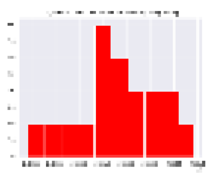
Solving the Issue of Data Scarcity

By performing the inverse transformation at this stage rather than earlier, we are able to leverage the larger dataset at the regional level to identify the desired level of extremeness. It would not have been possible to achieve this with the smaller datasets at the country level, where many countries had fewer than 20 observations.

Conclusion

Results

- Developed a comprehensive pipeline for generating realistic and extreme events.
- Evaluated multiple variants of GANs
- Classified losses per country into losses per climate bands.
- Used GANs and the extreme value theory to train a conditional GAN (ExGAN) that will generate values that comply with the desired extremeness.



Challenges

- Our data sets contained few observations per country which were discontinuous.
- Research on GANs has mostly focused on using them for image generation.
- Finding the optimal parameterization for GANs.

Future Work

- Tuning the hyperparameters of the DCGAN and the TTGAN in an effort to enhance its performance.
- Developing a conditional GAN for use with TTGANs and implementing it to generate extreme and realistic samples.

EIB Advisors: Giuseppe Bonavolontà, Oleg Reichmann
Research Advisors: Dr. Ali Hirsa | **Teaching Assistant:** Miao Wang