Unleashing the Unpredictable: Generating Context-Driven Synthetic Black Swans

Completed Research Paper

Lotfy Abdel Khaliq
German Research Centre for Artificial Intelligence (DFKI), Germany
lotfy.abdel_khaliq@dfki.de

Sabine Janzen
German Research Centre for Artificial Intelligence (DFKI), Germany
sabine.janzen@dfki.de

Wolfgang Maass
German Research Centre for Artificial Intelligence (DFKI), Germany
wolfgang.maass@dfki.de

Abstract

Black swans are rare high-impact crisis events that disrupt societies and financial markets, e.g., the 2008 global financial crisis. Due to their rarity, it is challenging to predict them with certainty using traditional statistical methods. Research on modeling black swans explored the use of deep learning techniques, such as LSTMs, Autoencoders, and self-supervised methods. However, data scarcity remains a challenge for both supervised and unsupervised methods. We present SwanSynthetiX, an approach for generating context-driven synthetic black swan events in open domains that closely resemble real-world extreme event data. It combines Extreme Value Theory (EVT) with conditional GANs (cGANs) and addresses data scarcity through EVT-based Monte Carlo-sampling. The approach uses scenarios to capture unique extreme event circumstances, enabling cGANs to generate context-driven black swans with distinct characteristics. Experiments demonstrate SwanSynthetiX outperforming recent approaches in synthetic time-series generation, empowering signal detection in crisis management for early identification of real-world black swan events.

Keywords: Black Swan Event; Extreme Value Theory; GANs

Introduction

Black swan events are high-impact crisis events that can have significant consequences and disrupt the status quo (Taleb 2007a), e.g., the 9/11 terrorist attacks, the global financial crisis of 2008, the Fukushima nuclear disaster, and the COVID-19 pandemic. Those events are characterized by their extreme rarity, their unexpectedness, and their out-sized impact on financial markets, economies, and societies as a whole (Taleb 2007b). Considering black swan events in risk and crisis management is an open issue for companies, health organizations, and civil defense (Aven 2014; Paté-Cornell 2012). But, up til now it is impossible to predict black swan events with certainty due to their rarity as well as their uniqueness (Swango 2020).

Traditional methods for predicting black swan events, like Extreme Value Theory (EVT) primarily depend on statistical and probabilistic approaches quantifying tail risks and estimating probabilities for extreme events (Allen et al. 2011). One statistical method often used in EVT is the Generalized Extreme Value (GEV) distribution for analyzing extreme values in datasets, e.g., black swan events (Allen et al. 2011; Smith 2002). However, these methods are based on assumptions like linearity and stationarity not representing the complex and nonlinear nature of black swan events (Devarajan et al. 2021). Recent research has explored the use of deep learning techniques to model black swan events. Here, LSTMs and GRUs are applied to predict extreme occurrences in financial time series data (Bhanja and Das 2022) or in healthcare to forecast COVID-19 active and new cases (Devarajan et al. 2021). Further related work treats the problem as anomaly detection issue for identifying atypical samples deviating from expected patterns, such as Autoencoders (Gong et al. 2019), GANs (Perera et al. 2019; Schlegl et al. 2019), and self-supervised methods (Cho et al. 2021; Golan and El-Yaniv 2018). However, the rarity of black swan events presents a significant challenge for such methods as it is difficult to distinguish anomalies from normal samples by supervised methods due to the given class imbalance, whereas unsupervised methods lack extensive knowledge about true anomalies(Gong et al. 2019). Means, in order to predict black swan events, data scarcity issues have to be addressed, for instance by applying GANs (Goodfellow et al. 2020) for generating time-series synthetic data (Ehrhart et al. 2022; Yoon et al. 2019). Conditional Generative Adversarial Networks (cGANs) (Mirza and Osindero 2018) have proven to be highly effective in generating labeled time-series samples that closely resemble real-world data (Esteban et al. 2017; fu et al. 2019), making them a promising solution for data scarcity in black swan events. However, the effectiveness of cGANs depends on the amount of available labeled training data, i.e., black swan events for generating realistic data (Ding et al. 2022). Furthermore, historical data on black swan events may not be representative for future occurrences due to the unique circumstances of each event (Taleb 2007a). Therefore, incorporation of contextual information related to the distinct circumstances of each event is required going beyond the scope of their mere label.

In this work, we present SwanSynthetiX - an approach for generating context-driven synthetic black swan events in open domains that closely resemble real-world data associated with extreme events. The approach combines EVT with cGANs and addresses the challenge of data scarcity through EVT-based Monte Carlo-sampling. SwanSynthetiX makes use of the fact that all crises, i.e., also black swan events send out early warning signals; but mostly weak and difficult to detect amidst the noise of everyday life (Diks et al. 2019; Fu and Zhu 2020). In order to capture these signals across the full spectrum of magnitudes, we divide time-series data into regions that exhibit different distributions. Unique attributes of each region within the time-series, e.g., GEV parameters, contextual information occurring simultaneously as well as label, are then represented as collection of feature-value pairs called scenario. The resulting set of scenarios enables SwanSynthetiX to address data scarcity by sampling points for each scenario and providing context to the GAN, making it context-aware to generate realistic synthetic data points for black swan events. Thus, in crisis management, the proposed approach is able to empower signal detection mechanisms for early identification of signals of real-world black swan events (Hensgen et al. 2003). The contributions of this paper are summarized as:

- We propose SwanSynthetiX: A generative approach that combines EVT-based Monte Carlo-sampling with context-driven conditioning of cGANs in a single model to generate synthetic black swan events.
- Our proposed method conditions cGANs on collection of feature-value pairs called scenarios, that combine GEV parameters, contextual information occurring simultaneously as well as scenario labels, allowing cGANs to generate context-driven black swan events with distinct characteristics.
- The experiments show the outstanding performance of our SwanSynthetiX approach in generating synthetic black swan events resembling real-world data in comparison with recent approaches on synthetic time-series generation.

Literature Review

Black Swan Events

There has been well-established works for predicting black swan events especially in financial markets using EVT-based methods. Such approaches use statistical methods to determine the probability of occurrence of a black swan, e.g. extreme price fluctuation in financial markets (Marohn 1998; Allen et al. 2011; Smith

2002). More recently, neural networks have been leveraged to model the existence of black swan events in the data, with ensemble methods being particularly effective in this area. Wabartha et al. introduced Diversely Extrapolated Neural Networks (DENN), which use an ensemble of neural networks with a diversity term in the loss function to generalize more effectively to novel data points and produce highly uncertain predictions for unexpected inputs (Wabartha et al. 2020). Liu et al. proposed a method that enhances outlier detection and model stability by combining multiple neural networks to improve prediction accuracy and resilience against out of distribution data (Saha et al. 2020). Another line of research explored the use of deep learning methods for estimating the magnitude of sudden fluctuations in time-series data using LSTMs (Bhanja and Das 2022), Autoencoders (Gong et al. 2019), GANs (Perera et al. 2019; Schlegl et al. 2019), and self-supervised methods (Cho et al. 2021; Golan and El-Yaniv 2018).

Synthetic Temporal Data Generation

Synthetic data generation has been an area of active research, addressing privacy concerns and data scarcity challenges in various domains (Donahue et al. 2019; Esteban et al. 2017; Mirza and Osindero 2018). Autoregressive recurrent networks, such as those trained using maximum likelihood principles, suffer from large prediction errors in multi-step sampling due to discrepancies between training and inference (Yoon et al. 2019). Several works such as Scheduled Sampling (Bengio et al. 2015) and Professor Forcing (Goyal et al. 2016) have been proposed to tackle these issues but are inherently deterministic and do not explicitly account for sampling from a learned distribution, which is crucial for synthetic data generation. GANs have been extended in multiple works to handle the temporal settings with frameworks like C-RNN-GAN (Mogren 2016) and Recurrent Conditional GAN (RCGAN) (Esteban et al. 2017). These methods have been applied in various domains, but they rely solely on binary adversarial feedback, which may not be sufficient to efficiently capture temporal dynamics in the training data (Yoon et al. 2019). Recently, representation learning for time-series data has been introduced to learn compact encodings for various downstream tasks (Dai and Le 2015; Eldele et al. 2021). Some approaches combine autoencoders with adversarial training (Larsen et al. 2016), but these methods might not generalize well to arbitrary time-series data or sufficiently capture stochasticity at each time step. TimeGAN (Yoon et al. 2019), a more advanced method, addresses these limitations by incorporating stochasticity and employing an embedding network to identify a lower-dimensional space for the generative model, allowing it to learn the stepwise distributions and latent dynamics of the data. However, a key limitation of TimeGAN and similar methods is their inability to handle anomalies in time-series data, which can significantly impact the performance of synthetic data generation and limit the applicability of generated data for tasks like anomaly detection and forecasting.

Supervised Anomaly Detection

Supervised Anomaly Detection is a recent research direction that addresses the scarcity of labeled anomaly data by utilizing a small number of available anomaly examples to learn models capable of identifying abnormal instances. These approaches include one-class metric learning with anomalies as negative samples (Huang and Li 2021; Pang et al. 2018), and one-sided anomaly-focused deviation loss (Pang et al. 2019). Despite their novelty, these models heavily depend on very small sets of observed anomalies and may overfit to the known abnormal patterns.

Preliminaries

In this section we describe our problem statement and provide definitions for different terms used in our work.

Given a time-series dataset containing both normal data points and anomalies represented by sudden fluctuations, our objective is to generate synthetic data that accurately models both the normal and anomalous regions of the dataset. The following are the definitions of the terms used in our work:

Black Swan Event: In the context of our work, we define a black swan event as an occurrence that lies in the far tail of the probability distribution of a given system's behavior represented by time-series, such as financial returns or energy prices. Formally, a black swan event occurs when a system metric X exceeds a threshold x with a tail probability $P(X \geq x) \leq \epsilon$ where ϵ is a small positive number.

Observational Set (O): The observational set is a collection of multivariate time-series data points for which our objective is to generate corresponding synthetic data. Formally, $O = \{(t, f_1, f_2, \dots, f_N) \mid t \in \{1, 2, \dots, T\}\}$ where N is the number of features in the dataset and T is the number of time steps.

Contextual Information Set (C): The contextual information set is a collection of multivariate time-series data points representing features that affect the observational set O and occur simultaneously. Formally, $C = (t, c_1, c_2, \dots, c_M) \mid t \in 1,2, \dots, T$ where M is the number of features in the dataset and T is the number of time steps.

Generalized Extreme Value Distribution (GEV): The GEV distribution is a continuous probability distribution that unifies the Gumbel, Fréchet, and Weibull families of extreme value distributions (Smith 2002). The GEV distribution is commonly used for modeling the extreme values of a dataset. It is characterized by three parameters: shape (ξ), location (μ), and scale (σ).

Scenario: A scenario is a collection of features characterizing distinct, non-overlapping regions within the Observation Set O. Each scenario is composed of the fitted GEV parameters P for the respective regions, the corresponding features from the Contextual Information Set C within the same time frame t, and a unique label l that identifies the scenario. A scenario s_i can be expressed as: $s_i = \left(P_i, C_i^{(t)}, l_i\right)$.

Method

Our approach consists of two main components: (A) Scenario Generator and (B) Context-Aware GAN, as shown in Figure 1. The Scenario Generator processes the time-series data within O by transforming it into scenarios using EVT-based Monte Carlo sampling and contextual information from C. This transformation results in a collection of sequences X_s, each corresponding to a specific scenario, along with the associated scenarios S. The sequences X_s and scenarios S are used as inputs to the Context-Aware GAN component, which consists of two interconnected modules: (1) AutoEncoder and (2) Scenario-Conditioned GAN (sc-GAN). The AutoEncoder learns to represent both the time-series sequences and the corresponding scenarios, effectively capturing the essential features of the dataset. Subsequently, the sc-GAN module utilizes these learned representations to generate synthetic data points that exhibit distinctive features tailored to the input scenario.

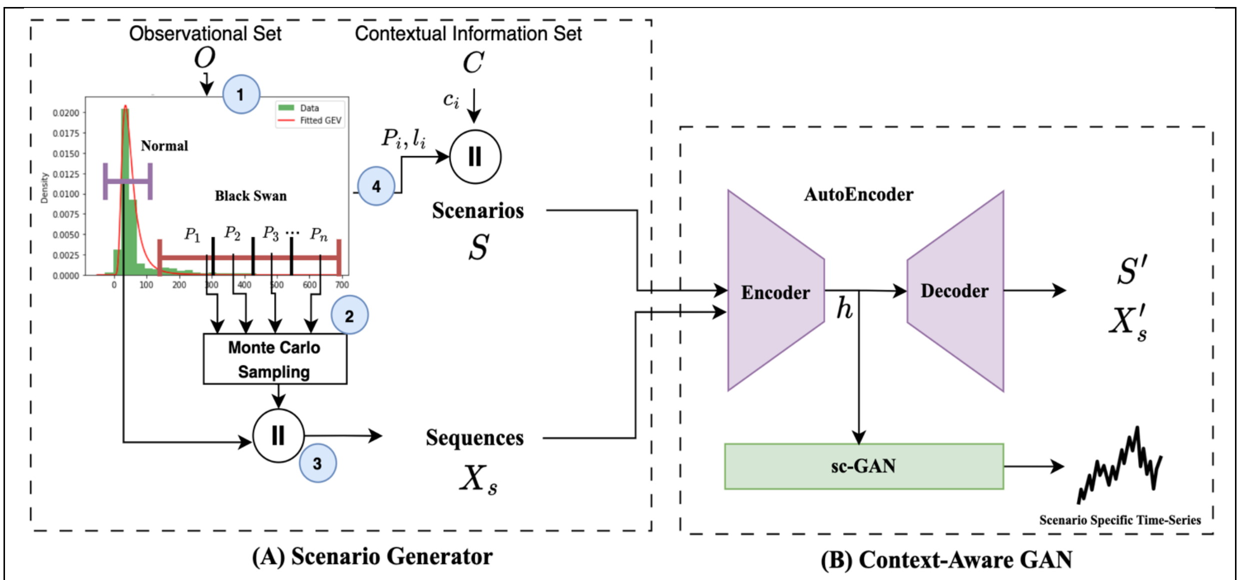


Figure 1. Overview of our approach for generating context-driven synthetic black swan events.

Scenario Generator

The Scenario Generator takes in two inputs: the observational set O and the contextual information set C. The processing steps, numbered in Figure 1 (A), are performed as follows: First, the observational set O is fitted to a Generalized Extreme Value (GEV) distribution to capture long-tail behavior. The goodness-of-fit is assessed to ensure it is appropriate for the dataset (Step 1). Next, by leveraging the fitted GEV distribution, we partition the data into n regions that exhibit different distribution by considering higher percentiles such as 90^{th}, 95^{th}, 99^{th}, and so on, which are located within the original dataset (Step 2). Mathematically, the process can be represented as: $R_i = x \in O \mid (X \leq x) \geq P_i$, where R_i represents the region associated with the i^{th} percentile, P_i, and $P(X \leq x)$ denotes the cumulative probability of observing a value less than or equal to x in the GEV distribution. Then, for each region R_i, a GEV distribution is fitted to capture the unique distribution of that particular region, characterized by GEV parameters P_i; shape (ξ_i), location (μ_i), and scale (σ_i). Finally, Monte Carlo sampling is used to draw data points for each region from the fitted GEV distributions, which are enhanced by seasonality and trend features (St, Tr) extracted from the original data. This process ensure that the sampled data accurately represent original data characteristics. Then, the original data corresponding to normal data points are used as sequences without any modifications. This is because normal data points are abundant, and therefore, readily available for processing (Step 3). Lastly, the scenarios are constructed by concatenating three inputs: the region's GEV distribution parameters P, the corresponding Contextual Information $C^{(t)}$ at the same time frame t, and the label of the region l (Step 4). In the end, the Scenario Generator produces two outputs for training the Context-Aware GAN: a set of sequences, $X_s = \{x_{s,1}, x_{s,2}, \ldots, x_{s,N}\}$, $\forall s \in S$, and the associated set of scenarios, $S = \{(P_1, C_1, l_1,), \ldots, (P_m, C_m, l_m)\}$.

Context-Aware GAN

The Context-Aware GAN component is a generative model designed to produce realistic and scenario-specific synthetic data points. As shown in Figure 2, it is composed of two main modules: an AutoEncoder and a scenario-conditioned GAN (sc-GAN).

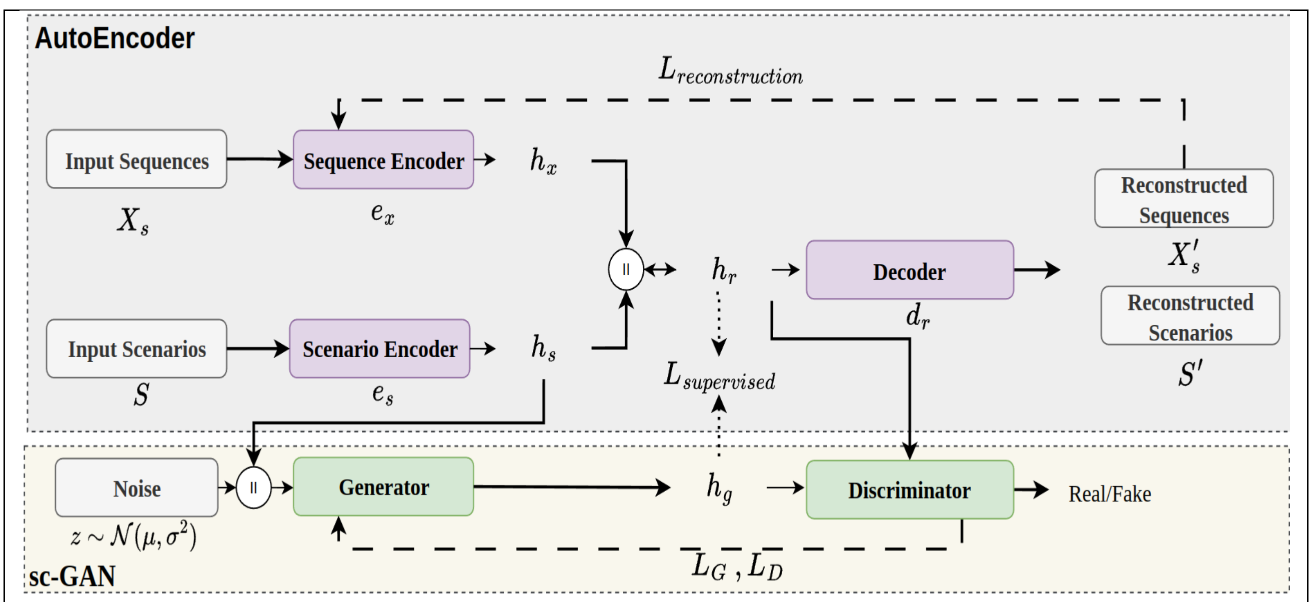


Figure 2. Context-Aware GAN

The AutoEncoder module is responsible for learning an efficient representation of the input sequences and scenarios generated from the Scenario Generator. It consists of three sub-modules: SequenceEncoder (e_x), ScenarioEncoder (e_s), and decoder (d_r). The three components are designed to enable the adversarial network to learn the underlying temporal dynamics of the data while considering the conditioning on input scenarios. e_x transforms the features of the input sequences X_s from the feature space into hidden representations h_r in the latent space using 3-layer LSTM network. The transformation can be expressed as:

$$h_x = e_x(h_{x-1}, x_t),$$

where e_x accounts for the previous real data representation h_{x-1} and the current feature vector x_t to produce the new real data representation h_x. Similarly, a separate scenario encoder e_s is used to produce scenario representations h_s using 3-layer LSTM Network. This encoder is designed to capture the characteristics of specific scenarios, which can be used to guide the generation of synthetic data that adheres to these scenarios. The representations obtained from both networks are then concatenated to form a combined hidden representation h_r. Finally, the decoder d_r takes the combined hidden representation h_r and reconstructs the original input sequences and scenarios as follows:

$$X_s', \ S' \ = r(h_r),$$

where r is a 3-layer LSTM and X_s', S' are the reconstructed sequences and scenarios, respectively. To integrate the AutoEncoder with adversarial modules, SwanSynthetiX incorporates a scenario-conditioned GAN to generate synthetic data points that exhibit the distinctive features of the input scenarios. The sc-GAN consists of two sub-modules: a conditional generator and a discriminator. Unlike traditional GANs, where the generator directly produces outputs in the feature space, our generator generates outputs in the latent space, and the discriminator distinguishes between real and fake hidden representations. Operating in the latent space, instead of output space, offers several advantages, including dimensionality reduction, noise filtering, improved generalization, and transferability (Bengio et al. 2013; Yoon et al. 2019). Let Z denote a Gaussian vector space from which a random vector $z_{1:T}$ is drawn. The generator takes the scenario embeddings, h_s, concatenated with $z_{1:T}$, and processes them through an LSTM network to obtain a hidden representation, h_g. This process conditions the generator to produce hidden representations specific to the input scenarios, which can be expressed as:

$$h_g = G(\mathrm{h_{s,}z_{1:T}})$$

where G is a 3-layer LSTM network. The discriminator receives the output from the generator and the combined hidden representation from the AutoEncoder to predict whether the received representation is real (from the AutoEncoder) or fake (from the generator). This setup enables the generator to learn more efficiently, focusing on the underlying structure and patterns in the data, and helps it to generate representations that accurately capture the context within the input scenarios.

Training and Loss Optimization

During the training phase, the Context-Aware GAN is optimized through a two-step process. First, the AutoEncoder is trained using the reconstruction loss $L_{reconstruction}$, which is defined as:

$$L_{reconstruction} = \frac{1}{n}\sum_{i=1}^{n}\left||x_i - \widehat{x_\iota}|\right|^2$$

This loss function ensures that the AutoEncoder learns to accurately reconstruct the input sequences and scenarios. Once it converges, it is integrated with the sc-GAN for joint training, leveraging a combination of loss functions including generator loss L_G, discriminator loss L_D, and supervised loss $L_{supervised}$, as shown in Figure 2. This sequential approach ensures that the AutoEncoder learns to effectively reconstruct the input data before it is integrated with the sc-GAN for generating synthetic samples.

In contrast to TimeGAN (Yoon et al. 2019) which samples directly from the generator, our approach conditions the generator on the encoded scenario representation to ensure that the synthetic data is more contextually relevant. Our L_G loss is defined as follows:

$$L_G = -\frac{1}{m}\sum_{i=1}^{m} \log D\ (G(h_s, z)|h_s)$$

where m is the number of samples, D is the discriminator, G is the generator, h_s is the scenario embeddings, and z is the noise vector. This loss function aims to optimize the generator network so that it can produce hidden representations $G(h_s, z)$ that are indistinguishable from real representations while being conditioned on the scenario representation h_s. The discriminator loss, on the other hand, is defined as:

$$L_D = -\frac{1}{m}\sum_{i=1}^{m}\left[\log D\ (h_r|h_s) + \log\left(1 - D(g(h_s, z)|h_s)\right)\right]$$

This loss function is composed of two terms: The first term $\log D\ (h_r|h_s)$ encourages the discriminator to assign high probabilities to the real hidden representations, h_r, conditioned on the scenario representation, h_s. The second term, $\log\left(1 - D(G(h_s, z)|h_s)\right)$, encourages the discriminator to assign low probabilities to the generated (fake) hidden representations produced by the generator, $G(h_s, z)$, conditioned on h_s.

Finally, inspired by (Yoon et al. 2019), to ensure that the generator captures the real distribution of the data and generates similar stepwise transitions, we adopt an additional supervised loss $L_{supervised}$ in closed-loop mode. In this mode, the generator receives the embedding sequences of actual data (h_r^{t-1}) conditioned onh_s, and generates the next latent vector h_g^t. This loss can be expressed as:

$$L_{supervised} = \sum_t \left|\left|h_r^t - g(h_r^{t-1}, z_t, h_s)\right|\right|^2$$

where $G(h_r^{t-1}, z_t, h_s)$ represents the generator's next step h_g^t with sample z_t, conditioned on h_s. In summary, the generator uses L_G and $L_{Supervised}$ losses to produce realistic sequences with similar stepwise transitions. L_G promotes indistinguishable hidden representations, while $L_{supervised}$ ensures accurate data distribution and transitions, all conditioned on the scenario representation h_s.

Once the training process is completed, SwanSynthetiX leverages the trained components to generate synthetic data points for a given input scenario. The process begins with scenario encoding, in which the input scenario is passed through the trained ScenarioEncoder e_s, generating the scenario embedding h_s (cf. Figure 2). Next, a noise vector sampled from a Gaussian distribution is concatenated with the scenario embedding and passed through the trained generator, producing a hidden representation h_g specific to the input scenario. Finally, the generated hidden representation is passed through the trained decoder d_r of the AutoEncoder, which reconstructs the synthetic data points that exhibit the characteristics of the input scenario while maintaining the structural patterns learned from the time-series dataset.

Evaluation

In this section, we conduct several experiments to evaluate whether SwanSynthetiX can generate synthetic data that accurately represents both normal and anomalous regions in time-series datasets.

Datasets

For our experiments, we collected historical time-series data on key indicators related to the German energy market and inflation from various sources, spanning from 2016 to 2023. As shown in Table 1, the collected data includes electricity wholesale prices, gas prices, crude oil prices, consumer price index, and index of import price, which we use as our Observational Set O. Furthermore, for the Contextual Information Set C, we gathered time-series datasets that affect the energy prices including electricity generation, electricity demand, gas imports, and weather information. We specifically chose these datasets as contextual features because these variables have a direct and substantial impact on energy prices (Min 2022; Sensfuß et al. 2008; Hartley et al. 2008; Bessec et al. 2008). These contextual features were selected to capture the most relevant and immediate factors contributing to price fluctuations, particularly during extreme events like the COVID-19 pandemic and the Russia-Ukraine conflict that led to drastic fluctuations in prices and overall system instability beginning mid-2021, as observed in Figure 3. The datasets were acquired from a diverse

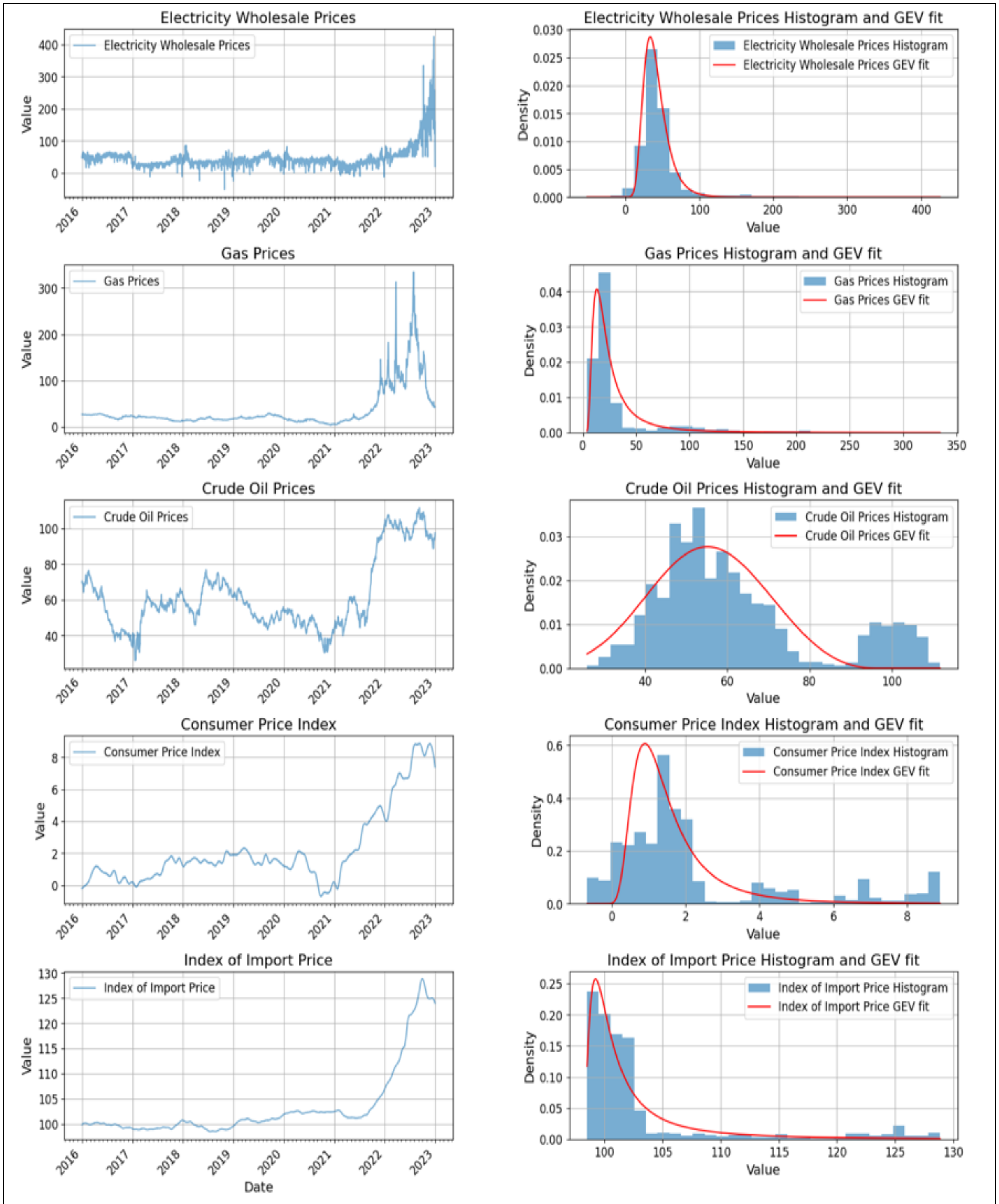


Figure 3. Observational Set (O) data (left) consisting of electricity wholesale prices, gas prices, crude oil prices, consumer price index (CPI), and index of import price along with corresponding fitted GEV distribution (right).

range of reliable providers, including ENTSO-E[6], Eurostat[7], International Energy Agency[8], German Federal Statistical Office (Destatis)[9], and NOAA[10].

Variable	Source	Set	Frequency	Number of Data Points
Electricity Wholesale Prices	ENTSO-E	O	Hourly	61,320
Gas Prices	Eurostat	O	Daily	2,556
Crude Oil Prices	IEA	O	Daily	2,556
Consumer Price Index	Destatis	O	Daily	2,556
Index of Import Price	Destatis	O	Daily	2,556
Gas Imports	ENTSO-G	C	Daily	2,556
Electricity Generation	ENTSO-E	C	Hourly	61,320
Electricity Load	ENTSO-E	C	Hourly	61,320
Weather	NOAA	C	Daily	2,556

Table 1. Summary of collected datasets used for evaluating the proposed approach, including variable types, sources, frequencies, and number of data points.

Setting

In this section, we describe the process of preparing our datasets to train and evaluate SwanSynthetiX. We then present the training details as well as the benchmarks used for evaluation. First, to ensure consistency across all features in the dataset, we perform a data preprocessing step where we take the daily average of the datasets that contain hourly values. Then, the Observational set O features are fitted to a GEV distribution, with goodness of fit assessed through qualitative analysis (cf. Figure 3). We additionally perform K-S test (Massey 1951) and show the results in Table 2, showing high p-values for all features, which confirms the distribution's suitability for the dataset.

Variable	Mean	Std	p-value
Electricity Wholesale Prices	45.67	35.92	0.64
Gas Prices	35.18	44.23	0.58
Crude Oil Prices	61.90	19.57	0.42
Consumer Price Index	2.210	2.360	0.54
Index of Import Price	103.3	7.05	0.57

Table 2. Goodness of fit results of the fitted GEV distribution

Since the Observational set O features started to fluctuate mid-2021, we consider data points from January 2016 to Jun 2021 as normal data and those from Jul 2021 to December 2022 as black swan data. Normal data consists of 1981 data points each characterized by the five features of Observational set O. We identified

[6] https://www.entsoe.eu/

[7] https://ec.europa.eu/eurostat/

[8] https://www.iea.org/

[9] https://www.destatis.de/

[10] https://www.noaa.gov/

the period after July 2021 as black swan events due to significant shifts in the dataset, leading to long-tail behavior when fitted to GEV distribution, as shown in Figure 3. However, we recognize that not all fluctuations may represent black swan events. To address this, we divided the black swan data into six distinct regions, each containing 92 data points, reflecting different characteristics of the data during this period. This approach helps capture the variations in the fluctuations and ensures a more precise classification rather than a catch-all categorization. To address the imbalance between normal and black swan data, Monte Carlo sampling uses the fitted GEV distribution for each region to generate additional 1888 samples. These samples are then enhanced with STL decomposition features (St, Tr) extracted from the original 92 points. Now, normal data, as well as each black swan region, is characterized by three features: (1) GEV parameters (P; shape (ξ), location (μ), and scale (σ)), (2) corresponding context from the Contextual information set C, and (3) region label l. To form the set of scenarios S, we concatenate these features for each data point. Scenarios that include normal data are assigned a label of 0, while scenarios from the six black swan regions are assigned labels from 1 to 6. Leveraging the sequences and scenarios generated from the Scenario Generator, we initially trained the AutoEncoder for 150,000 iterations. Subsequently, we conducted joint training of the whole Context-Aware GAN component for additional 150,000 iterations.

Our evaluation considers three key aspects: (1) Fidelity, requiring that generated samples are indistinguishable from the real data, (2) usefulness, meaning that synthetic samples should perform similarly to real data for the same predictive purposes, and (3) diversity, ensuring that generated samples cover the real data distribution. We use four evaluation metrics to assess the three aforementioned aspects:

- Discriminative Score (D-Score): We train an LSTM-based classification model to distinguish between original and generated sequences. Lower classification scores indicate higher fidelity in the synthetic data.
- MAE: To further assess fidelity, we compute the Mean Absolute Error (MAE) between the original and synthetic data, providing a measure of the discrepancies between them. Lower MAE scores represent higher fidelity.
- Predictive Score (P-Score): To evaluate usefulness, we train an LSTM-based sequence-prediction model on the synthetic data to predict the next temporal hidden representation and test its performance on the original dataset, using MAE as a quantitative measure of the model's ability to reproduce the predictive properties of the original data.
- Visualization: We use PCA to reduce data dimensionality and qualitatively assess how closely generated samples resemble the original data distribution in 2D space, which asses the diversity of the generated data.

For all metrics, we use an equal distribution of real and synthetic data to ensure a fair comparison and accurate assessment.

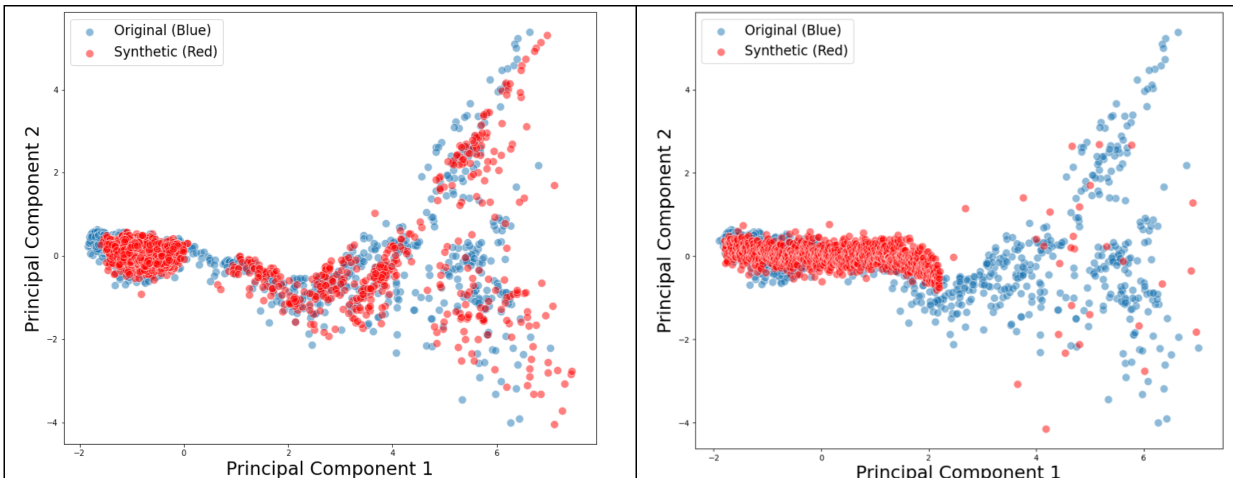
Results

We assess the performance of our proposed approach quantitatively by comparing it with various time-series generation methods, including TimeGAN (Yoon et al. 2019), RCGAN (Esteban et al. 2017), CRNNGAN (Mogren 2016), P-Forcing (Goyal et al. 2016), and WaveGAN (Donahue et al. 2019). For evaluation, we focus on two dataset regions: the normal data points spanning from Jan 2016 to Jun 2021, and the black swan data points from Jul 2021 to Dec 2022. As shown in Table 3, we observe that SwanSynthetiX outperforms the state-of-the-art methods for all metrics for generating both normal and black swan data points. For normal data, SwanSynthetiX has a Discriminative Score of 0.09 and a Predictive Score of 0.21, which is considerably better than the other methods. More importantly, for black swan region, our approach obtains a Discriminative Score of 0.08 and a Predictive Score of 0.28, outperforming TimeGAN by 29.56% in the Normal region and by 52.71% in the black swan region.

	Normal			Black Swan		
	D-Score	P-Score	MAE	D-Score	P-Score	MAE
RCGAN	0.27	0.38	0.32	0.50	0.52	0.53
CRNNGAN	0.29	0.41	0.34	0.54	0.56	0.55
WaveGAN	0.23	0.37	0.29	0.48	0.50	0.54
TimeGAN	0.15	0.32	0.28	0.42	0.44	0.49
SwanSynthetiX	0.09	0.21	0.24	0.08	0.28	0.29

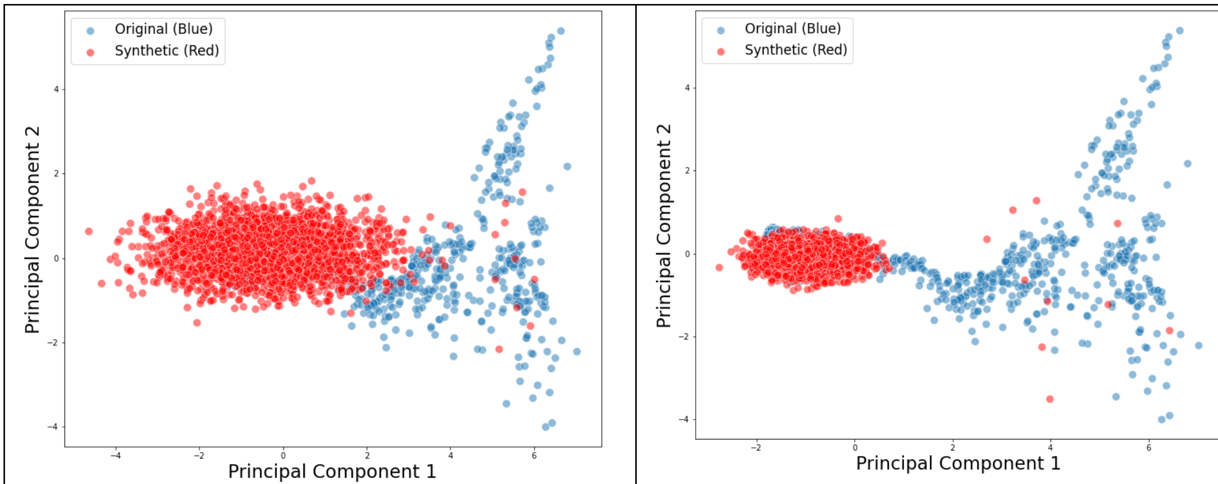
Table 3. Comparison of our approach to other state-of-the-art methods for generating synthetic data. Normal refers to data from Jan 2016 till Jun 2021, while Black Swan refers to data from Jul 2021 to Dec 2022.

To visually assess the performance of SwanSynthetiX, we conducted a qualitative analysis by plotting Principal Component Analysis (PCA) representations for the original and synthetic data generated by our method, TimeGAN, RCGAN, and CRNNGAN. Figure 4 shows that SwanSynthetiX effectively generates synthetic data that closely resembles the original data distribution, capturing both normal and black swan patterns. This sets our model apart from these methods, which fail to fully capture the distribution of the data and struggle to represent rare and extreme values.



(A) SwanSynthetiX

(C) TimeGAN



(B) RCGAN

(D) CRNNGAN

Figure 4. Comparison of PCA plots for SwanSynthetiX with RCCGAN, TimeGAN and CRNNGAN.

Finally, to evaluate the contribution of scenarios in SwanSynthetiX, we conducted an ablation study in which we measure the impact of removing features from the scenario. Table 4 shows the discriminative score, predictive score, and MAE for different combinations of the three scenario features: GEV parameters P, Contextual information set C, and label l. We observe that SwanSynthetiX by incorporating P, C, and l into scenarios achieves the best performance across all three metrics. Removing any of the features leads to performance degradation, confirming their importance in generating high-quality synthetic data. Notably, while removing P and C has a relatively small impact on MAE (only a 1% difference), it results in a sharp error increase in the discriminative score (5%) and the predictive score (3%). This suggests that although the synthetic data generated after removing these factors may be close in magnitude to those generated with all factors, they do not exhibit the same predictive capabilities. Furthermore, by removing all scenario features, i.e., removing conditioning on the GAN, we observe a substantial drop in performance. This highlights the importance of contextual scenarios in generating synthetic data that closely resemble real-world data associated with extreme events.

Label l	**Contextual Information C**	**GEV Params P**	**D-Score**	**P-Score**	**MAE**
✓	✓	✓	0.09	0.23	0.25
✓	✓	x	0.13	0.24	0.25
x	✓	✓	0.16	0.29	0.28
✓	x	x	0.14	0.26	0.26
x	✓	x	0.19	0.25	0.27
x	x	✓	0.14	0.24	0.28
x	x	x	0.18	0.26	0.27

Table 4. Ablation study of the impact of the scenario features, including GEV parameters, contextual information, and labels, on the performance of our approach. The symbol ✓ indicates this feature is used, while x indicates it is removed.

Implications

The proposed approach, SwanSynthetiX, aims to enhance the identification of crises by generating synthetic data that exhibit the characteristics of rare and unexpected high-impact events like black swan events. Our approach provides a plug-and-play method for generating data of any extreme event that can be represented by time-series data. The inclusion of contextual features as conditions for the generated data allows it to be adapted and extended to accommodate a wide range of crisis events. This flexibility is especially crucial for crises driven by complex and unique patterns, such as those influenced by political and socio-economic factors. Here, SwanSynthetiX leverages a Scenario Encoder to encode and integrate diverse types of contextual data—whether numerical or text-based—directly into the GAN's training process, making it generalizable across various domains.

Such synthetic data can be utilized in evaluation studies where state-of-the-art ML models are trained to predict these events, which is particularly beneficial for organizations seeking to proactively detect and manage potential crises. This is especially relevant in sectors like finance and healthcare, where the ability to model and predict black swan events can significantly impact decision-making and risk management strategies. As presented in the results section, the generated data from our approach not only shows high fidelity but also demonstrates high-impact usability when used to train ML models for forecasting the next temporal value in time-series data, which is explained by its low Predictive Score (P-score).

In general, synthetic data generation tries to capture the most prominent statistical properties of a dataset and samples similar data points, but the design of SwanSynthetiX allows for capturing the properties of anomalies in the data, which are considered a strong minority class, especially in the case of black swans. This indicates that the data generated by SwanSynthetiX is highly effective in enabling accurate predictions of future values in time-series data with anomalies, making it particularly valuable for classification and forecasting tasks. Subsequent studies will be conducted to evaluate the results from our approach for predicting potential black swan instances in several domains and observe the outcomes for organizations in terms of mitigating risks and minimizing negative consequences.

Limitations

Despite the promising results of SwanSynthetiX, several limitations need to be addressed in future work. Firstly, the current scenario generation process assumes that black swan events emerge slowly and then exponentially, which may not account for all possible scenarios of such events (Taleb 2007a). Moreover, the use of GANs, which are inherently biased towards the data they are trained on, can lead to issues like mode collapse, limiting the diversity of the generated data (Grover et al. 2019). Techniques such as mode regularization, multiple discriminators, or using Transformers (Vaswani et al. 2017) could help address these limitations by capturing long-range dependencies and generating more diverse samples. Additionally, our approach relies on historical data to generate black swan events, which contradicts the inherently unpredictable nature of these events. Although this reliance on historical data provides context and parameters necessary for simulation, it also assumes the existence of black swan events in historical records, which may not always be the case in different contexts. To overcome this, future work would explore integrating expert knowledge or domain-specific heuristics to build a prediction model that can more accurately forecast unpredictable black swan scenarios. Lastly, further investigation is required to validate the applicability of our approach across different domains and types of black swan events, ensuring its effectiveness in diverse contexts.

Conclusion

In this work, we introduced SwanSynthetiX, a novel approach for generating context-driven synthetic black swan events that closely resemble real-world data. Combining Extreme Value Theory, conditional GANs, and Monte Carlo sampling, our method addresses data scarcity and the uniqueness of black swan events. By using scenarios that incorporate contextual information, SwanSynthetiX generates distinct, context-aware synthetic black swan events. Our experiments demonstrate SwanSynthetiX's performance compared to recent methods in synthetic time-series generation. Our approach holds the potential to improve early signal detection mechanisms in crisis management, leading to more effective responses to real-world black swan events and paving the way for future research in generating anomalies for time-series datasets.

References

Allen, D., Singh, A., and Powell, R. 2011. "Extreme Market Risk-An Extreme Value Theory Approach," *Mathematics and Computers in Simulation* (94:August).

Aven, T. 2014. "Risk, Surprises and Black Swans: Fundamental Ideas and Concepts in Risk Assessment and Risk Management," *Risk, Surprises and Black Swans: Fundamental Ideas and Concepts in Risk Assessment and Risk Management*. (https://doi.org/10.4324/9781315755175).

Bengio, S., Vinyals, O., Jaitly, N., and Shazeer, N. 2015. "Scheduled Sampling for Sequence Prediction with Recurrent Neural Networks," in *Advances in Neural Information Processing Systems* (Vol. 2015-January).

Bengio, Y., Courville, A., and Vincent, P. 2013. "Representation Learning: A Review and New Perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence* (35:8). (https://doi.org/10.1109/TPAMI.2013.50).

Bhanja, S., and Das, A. 2022. "A Black Swan Event-Based Hybrid Model for Indian Stock Markets' Trends Prediction," *Innovations in Systems and Software Engineering*. (https://doi.org/10.1007/s11334-021-00428-0).

Cho, H., Seol, J., and Lee, S. G. 2021. "Masked Contrastive Learning for Anomaly Detection," in *IJCAI International Joint Conference on Artificial Intelligence*. (https://doi.org/10.24963/ijcai.2021/198).

Dai, A. M., and Le, Q. V. 2015. "Semi-Supervised Sequence Learning," in *Advances in Neural Information Processing Systems* (Vol. 2015-January).

Devarajan, J. P., A, M., and Sreedharan, V. R. 2021. "Healthcare Operations and Black Swan Event for COVID-19 Pandemic: A Predictive Analytics," *IEEE Transactions on Engineering Management*. (https://doi.org/10.1109/TEM.2021.3076603).

Diks, C., Hommes, C., and Wang, J. 2019. "Critical Slowing down as an Early Warning Signal for Financial Crises?," *Empirical Economics* (57:4). (https://doi.org/10.1007/s00181-018-1527-3).

Ding, C., Pang, G., and Shen, C. 2022. "Catching Both Gray and Black Swans: Open-Set Supervised Anomaly Detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (Vol. 2022-June). (https://doi.org/10.1109/CVPR52688.2022.00724).

Donahue, C., McAuley, J., and Puckette, M. 2019. "Adversarial Audio Synthesis," in *7th International Conference on Learning Representations, ICLR 2019*.

Ehrhart, M., Resch, B., Havas, C., and Niederseer, D. 2022. "A Conditional GAN for Generating Time Series Data for Stress Detection in Wearable Physiological Sensor Data," *Sensors* (22:16). (https://doi.org/10.3390/s22165969).

Eldele, E., Ragab, M., Chen, Z., Wu, M., Kwoh, C. K., Li, X., and Guan, C. 2021. "Time-Series Representation Learning via Temporal and Contextual Contrasting," in *IJCAI International Joint Conference on Artificial Intelligence*. (https://doi.org/10.24963/ijcai.2021/324).

Esteban, C., Hyland, S. L., and Rätsch, G. 2017. "Real-Valued (Medical) Time Series Generation with Recurrent Conditional GANs," *ArXiv* (abs/1706.02633). (https://api.semanticscholar.org/CorpusID:29681354).

fu, rao, Chen, J., Zeng, S., zhuang, yiping, and Sudjianto, A. 2019. "Time Series Simulation by Conditional Generative Adversarial Net," *SSRN Electronic Journal*. (https://doi.org/10.2139/ssrn.3373730).

Fu, K. wa, and Zhu, Y. 2020. "Did the World Overlook the Media's Early Warning of COVID-19?," *Journal of Risk Research*. (https://doi.org/10.1080/13669877.2020.1756380).

Golan, I., and El-Yaniv, R. 2018. "Deep Anomaly Detection Using Geometric Transformations," in *Advances in Neural Information Processing Systems* (Vol. 2018-December).

Gong, D., Liu, L., Le, V., Saha, B., Mansour, M. R., Venkatesh, S., and Van Den Hengel, A. 2019. "Memorizing Normality to Detect Anomaly: Memory-Augmented Deep Autoencoder for Unsupervised Anomaly Detection," in *Proceedings of the IEEE International Conference on Computer Vision* (Vol. 2019-October). (https://doi.org/10.1109/ICCV.2019.00179).

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. 2020. "Generative Adversarial Networks," *Communications of the ACM* (63:11). (https://doi.org/10.1145/3422622).

Goyal, A., Lamb, A., Zhang, Y., Zhang, S., Courville, A., and Bengio, Y. 2016. "Professor Forcing: A New Algorithm for Training Recurrent Networks," in *Advances in Neural Information Processing Systems*.

Grover, A., Song, J., Agarwal, A., Tran, K., Kapoor, A., Horvitz, E., and Ermon, S. 2019. "Bias Correction of Learned Generative Models Using Likelihood-Free Importance Weighting," in *Advances in Neural Information Processing Systems* (Vol. 32).

Hensgen, T., Desouza, K. C., and Kraft, G. D. 2003. "Games, Signal Detection and Processing in the Context of Crisis Management," *Journal of Contingencies and Crisis Management* (11:2). (https://doi.org/10.1111/1468-5973.1102003).

Huang, R., and Li, Y. 2021. "MOS: Towards Scaling Out-of-Distribution Detection for Large Semantic Space," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. (https://doi.org/10.1109/CVPR46437.2021.00860).

Larsen, A. B. L., Sønderby, S. K., Larochelle, H., and Winther, O. 2016. "Autoencoding beyond Pixels Using a Learned Similarity Metric," in *33rd International Conference on Machine Learning, ICML 2016* (Vol. 4).

Massey, F. J. 1951. "The Kolmogorov-Smirnov Test for Goodness of Fit," *Journal of the American Statistical Association* (46:253). (https://doi.org/10.1080/01621459.1951.10500769).

Mirza, M., and Osindero, S. 2018. "Conditional Generative Adversarial Nets Mehdi," *ArXiv:1411.1784v1 [Cs.LG] 6 Nov 2014 Conditional*.

Mogren, O. 2016. "C-RNN-GAN: Continuous Recurrent Neural Networks with Adversarial Training," *ArXiv* (abs/1611.09904). (https://api.semanticscholar.org/CorpusID:259842).

Pang, G., Chen, L., Cao, L., and Liu, H. 2018. "Learning Representations of Ultrahigh-Dimensional Data for Random Distance-Based Outlier Detection," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. (https://doi.org/10.1145/3219819.3220042).

Pang, G., Shen, C., and Van Den Hengel, A. 2019. "Deep Anomaly Detection with Deviation Networks," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. (https://doi.org/10.1145/3292500.3330871).

Paté-Cornell, E. 2012. "On 'Black Swans' and 'Perfect Storms': Risk Analysis and Management When Statistics Are Not Enough," *Risk Analysis* (32:11). (https://doi.org/10.1111/j.1539-6924.2011.01787.x).

Perera, P., Nallapati, R., and Xiang, B. 2019. "OCGAN: One-Class Novelty Detection Using Gans with Constrained Latent Representations," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (Vol. 2019-June). (https://doi.org/10.1109/CVPR.2019.00301).

Schlegl, T., Seeböck, P., Waldstein, S. M., Langs, G., and Schmidt-Erfurth, U. 2019. "F-AnoGAN: Fast Unsupervised Anomaly Detection with Generative Adversarial Networks," *Medical Image Analysis* (54). (https://doi.org/10.1016/j.media.2019.01.010).

Smith, E. P. 2002. "An Introduction to Statistical Modeling of Extreme Values," *Technometrics* (44:4). (https://doi.org/10.1198/tech.2002.s73).

Swango, D. L. 2020. "Black Swans: When the Impossible Occurs.," *Appraisal Journal* (88:2).

Taleb, N. N. 2007a. *The Black Swan: The Impact of the Highly Improbable*, Random House.

Taleb, N. N. 2007b. "Black Swans and the Domains of Statistics," *American Statistician* (61:3). (https://doi.org/10.1198/000313007X219996).

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. 2017. "Attention Is All You Need," in *Advances in Neural Information Processing Systems* (Vol. 2017-December).

Yoon, J., Jarrett, D., and van der Schaar, M. 2019. "Time-Series Generative Adversarial Networks," in *Advances in Neural Information Processing Systems* (Vol. 32), H. Wallach, H. Larochelle, A. Beygelzimer, F. d Alché-Buc, E. Fox, and R. Garnett (eds.), Curran Associates, Inc. (https://proceedings.neurips.cc/paper_files/paper/2019/file/c9efe5f26cd17ba6216bbe2a7d26d490-Paper.pdf).

Min, H. 2022. Examining the impact of energy price volatility on commodity prices from energy supply chain perspectives. *Energies*, *15*(21), 7957.

Sensfuß, F., Ragwitz, M., & Genoese, M. (2008). The merit-order effect: A detailed analysis of the price effect of renewable electricity generation on spot market prices in Germany. *Energy policy*, *36*(8), 3086-3094.

Hartley, P. R., Medlock III, K. B., & Rosthal, J. E. (2008). The relationship of natural gas to oil prices. *The Energy Journal*, *29*(3), 47-66.

Bessec, M., & Fouquau, J. (2008). The non-linear link between electricity consumption and temperature in Europe: A threshold panel approach. Energy Economics, 30(5), 2705-2721.

Marohn, F. (1998). Testing the Gumbel hypothesis via the POT-method. *Extremes*, *1*(2), 191-213.

Wabartha, M., Durand, A., Francois-Lavet, V., & Pineau, J. (2021, January). Handling black swan events in deep learning with diversely extrapolated neural networks. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence* (pp. 2140-2147).

Saha, A., Chatterjee, A., Ghosh, S., Kumar, N., & Sarkar, R. (2021). An ensemble approach to outlier detection using some conventional clustering algorithms. Multimedia Tools and Applications, 80, 35145-35169.