technische universität
dortmund

Essays in Finance: Generative Probabilistic Models, Firm Efficiency, and Investor Relations

Dissertation zur Erlangung des akademischen Grades
Doctor rerum politicarum
der Fakultät für Wirtschaftswissenschaften
der Technischen Universität Dortmund

Miguel Krause

September 19, 2022

Contents

List of Figures

List of Tables

Acknowledgments

This dissertation is the result of my work at the Chair of Finance of the TU Dortmund University. I would like to thank a number of people who helped me to accomplish this dissertation.

I want to start by expressing my gratitude to my doctoral supervisor Peter N. Posch for giving me the opportunity to write this dissertation at the Chair of Finance. Thank you for your guidance and support on all the projects I worked on and for giving me the freedom to conduct my own research.

I am deeply grateful to my co-authors Daniel and Nils for the time we spent on our COVID-19 projects right at the beginning of the pandemic, which led to far more successful projects than I could have ever imagined. I hope our friendship continues.

Additionally, I would also like to thank my co-author Nico for introducing me to the machine learning research community and for the in-depth talks we had concerning our paper.

I also want to thank my former Master's thesis supervisor, Timo, for encouraging me to pursue a Ph.D. in the first place and for all the discussions we had in our joint seminars. Furthermore, I would like to thank my colleagues Gerrit, Aydin, Lars, Philipp, Nicole and all the assistants for the great time at the chair.

Finally, I would like to thank my family. Many thanks to Francisca, Jörg, Michael, Andrea, my brother David and my love, Annkathrin. Without all of you, I would never have managed to finish this dissertation.

Thank you.

1 Introduction

Empirical research in finance and economics usually focuses on estimating (causal) relationships between inputs and an output, also known as statistical inference. As researchers in these fields can rarely conduct randomized controlled experiments because these are often impractical or unethical, they must rely on econometric methods applied to observational data where, according to the principle of ceteris paribus, all other factors remain fixed (Wooldridge, 2010; Tiffin, 2019). This observational approach makes it possible to interpret how a causal effect is identified and how likely it is that the observed effect is due to chance, as the effect could have arisen from sampling error alone (Einav and Levin, 2014).

Machine learning research, on the other hand, focuses mainly on finding generalizable predictive patterns to predict an output given certain inputs (Bzdok et al., 2018). Here, researchers are less interested in estimating (causal) effects from in-sample data than in making predictions for out-of-sample data. However, machine learning research is strongly linked to the technology sector and therefore mainly interested in image, speech or text data (Krizhevsky et al., 2012; Hinton et al., 2012; Devlin et al., 2018). Due to the increasing amount of data and available processing capacity, these methods are also heavily used by the financial industry, e.g. for fraud detection, robo-advisory, chatbots for banks or algorithmic trading (Buchanan, 2019).

This dissertation consists of four independently written essays dealing with both, inference and prediction of financial data sets. Following this brief introduction, the remainder will provide detailed summaries of the individual essays and publications details.

The first part of this dissertation covers two chapters that explore the question of how financial markets priced companies' stocks during the market collapse caused by the COVID-19 pandemic in the beginning of 2020. As the COVID-19 pandemic and the subsequent economic lockdown represented one of the most impacting exogenous shocks to financial markets in recent history, it led to a huge increase in uncertainty about a firm's future cash flows (Fahlenbrach et al., 2020). This environment thus allowed us to examine the drivers and characteristics that may make firms more resilient to crises and help to reduce investor uncertainty.

Against this background, **Chapter 2** examines how firm efficiency influenced stock returns during the COVID-19 crisis. There are two opposing theories on how firm efficiency may influence stock returns given that a firm's operational efficiency should determine its future cash flows and associated returns. On the one hand, inefficient firms likely have a higher risk associated with their future cash flows, and hence these firms must offer higher rates of return to attract risk-averse investors (Nguyen and Swanson, 2009). One the other hand, investors' demand for efficiently operating firms may be high and so their stock returns, since these firms have more secured future cash flows and are consequently exposed to a lower risk of default (Frijns et al., 2012).

We investigate the link between firm efficiency and stock returns during the COVID-19 crisis using a sample of 884 US firms. We follow Frijns et al. (2012) and argue that firms using their resources more efficiently should be more resilient during the crisis since they have a significant lower risk of corporate default. This should lead to a higher valuation by investors. Our empirical analysis supports this hypothesis. Based on efficiency scores obtained from a Stochastic Frontier Analysis (SFA) and a Data Envelopment Analysis (DEA), we find that highly efficient firms experienced at least 9.44 percentage points higher cumulative returns than inefficient firms during the COVID-19 crisis. Additionally, we find that a long-short portfolio consisting of efficient and inefficient firms would have realized a significant positive weekly return of 3.5% on average in a backtesting setting.

Chapter 3 investigates the value of investor relations (IR) during the COVID-19 crisis. As mentioned above, the COVID-19 pandemic and the respective lockdown measures have caused uncertainty on capital markets. Particularly, the rumours and news in the media and on the internet, which mainly revolved around whether companies were able to overcome the crisis might have overwhelmed investors and might have led to information fictions. This is because investors have been shown to possess only limited information processing capacities (Hirshleifer and Teoh, 2003). In this paper, we therefore hypothesize that an effective communication by a company's IR department with its investors might have helped to reduce uncertainty and to alleviate information frictions; and thus, this should have paid off during the crisis.

In our main tests as well as in our additional tests, we find consistent evidence to support our hypothesis. Using a sample of almost 1,000 companies from 16 different European countries and IR rankings from Institutional Investor, we find that companies with better-quality IR experienced significantly higher returns during the crisis period. Particularly, we document that firms with strong IR experienced between five and eight percentage points higher stock returns than firms with weak IR during the COVID-19 crisis. This result is robust to controlling for a several firm and governance characteristics, industry and country-fixed effects, and to using an entropy balanced sample, which helps us to address possible endogeneity concerns.

To tighten the link between a firm's IR quality and its stock performance during the crisis, we also run regressions with weekly stock returns over the entire year 2020 as the dependent variable and interactions of all independent variables with the weekly growth of COVID-19 cases, similar to Ding et al. (2021). Further, we run daily cross-sectional regressions during the first quarter of 2020 and difference-in-differences regressions with daily abnormal returns following Albuquerque et al. (2020) and Lins et al. (2017). But despite these different settings, the results confirm our earlier findings.

Apart from our main tests, we also focus in Chapter 3 on how a firm's IR functions might have increased its firm value. We first examine two different IR functions, namely public and private IR. We find that mainly private IR functions, such as one-to-one meetings of senior management with investors, are the main drivers of our results, and that the public IR function does not seem to have increased the performance of firms with strong IR. Further, our results provide evidence that a firm's (private) IR functions have boosted its stock performance by increasing credibility with incumbent shareholders and by diversifying the shareholder base. Strong IR firms have not only managed to retain incumbent institutional investors, but also seem to have managed to attract new institutional investors during the crisis.

In further tests, we also find differences in the value of (private) IR depending on the countries the firms are headquartered in. In line with Karolyi et al. (2020) we find that strong (private) IR was even more valuable for firms headquartered in countries with lower-quality legal institutions. Furthermore, we find that firms with better-quality IR benefited significantly more in countries with a low level of societal trust and in uncertainty-avoidance countries.

The last two essays of this dissertation move away from the inference element and deal with the prediction of financial time-series data using unsupervised machine learning methods. In the finance literature so far, machine learning models are mainly used for discriminative tasks, such as point forecasts or classifications. However, in this dissertation, we show how the finance literature can be extended by using generative probabilistic models, which aim to learn the underlying distribution of the data and are able to generate realistic artificial samples. Since time-series in the real world are highly stochastic, probabilistic sampling has the advantage of providing a complete distribution of possible scenarios instead of a single prediction.

Chapter 4 studies measures for the evaluation of generative probabilistic models for time-series data. Prior literature has shown that generative machine learning models are capable of generating artificial data of exceptional quality in areas such as natural language processing or computer vision (Karras et al., 2018), but there is still a gap in the literature regarding time-series data. This may be due to the fact that humans can easily judge whether a generated image or a text is realistic, deciding whether generated time-series are realistic is more challenging because of a lack of reliable quality measures. Against this background, we aim to quantify the similarity between real time-series samples and samples from generative models by using the popular maximum mean discrepancy (MMD) (Gretton et al., 2006) as well as our new proposal, the Hausdorff discrepancy (HD). The latter takes the shape of the generated samples into account by using the well-known Hausdorff distance. Since both discrepancies require an inner core distance, we consider the Euclidean, the dynamic time warping and the Fréchet distance to finally obtain six discrepancy measures.

We compare implicit and explicit probabilistic models, to find the best generator for time-series data. For implicit models, which are likelihood-free, we employ generative adversarial networks (Goodfellow et al., 2014) and variational autoencoders (Kingma and Welling, 2013; Rezende et al., 2014). For explicit models, we use Markow random fields (Piatkowski et al., 2013), which allow us to estimate an explicit likelihood function. However, both model classes allow us to sample from the underlying data distribution.

In our empirical analysis, we train several hundred generative models and evaluate them using the discrepancy measures. To do so, we employ two real-world time series data sets: (I) hourly day-ahead electricity prices from the European Power Exchange and (II) hourly humidity measurements from the Intel Berkeley Research Lab.

Using both data sets, our results suggest implicit models outperform explicit models based on our discrepancy measures. For the MMD, we find that generative adversarial networks outperform variational autoencoders and Markow random fields in 4 out of 6 distance measures. Considering our new proposal, the Hausdorff discrepancy, we find the variational autoencoders to win in 4 out of 6 cases. A visual examination of the generated sample paths shows that GANs and VAEs tend to generate samples close to the mean of the data set, while Markow random fields generate more realistic paths with extreme values close to those the original data set. This visual result is strengthened in our additional tests, where we compare the first four statistical moments of the original data distributions with the generated samples of all models. In these tests, Markow random fields outperform the implicit models and are able to generate samples with statistical properties similar to the moments of the original data.

Overall, our study suggests that multiple measures should be considered when assessing the quality of generative models for time-series data, and that even the most advanced measures can be misleading.

Finally, **Chapter 5** shifts the focus from the simulation to the prediction of time-series data using generative models. In order to make reliable decisions in high uncertainty environments, probabilistic forecasting allowing to quantify the underlying uncertainty and to obtain a sample of forecasts instead of a deterministic point forecast is vital. For instance, especially decision-makers in the energy sector rely on accurate probabilistic forecasting given the high levels of uncertainty due to climate change and the threat of the COVID-19 pandemic. In this chapter, we therefore investigate the predictive power of generative models on energy time-series. We employ two popular frameworks of generative models, namely generative adversarial networks (GAN) (Goodfellow et al., 2014) and variational autoencoders (VAE) (Kingma and Welling, 2013; Rezende et al., 2014), which can be used as drop-in Monte Carlo samplers (Piatkowski et al., 2021a). To incorporate conditional information in the generative sampling process, we apply the conditional forms of the generative frameworks, the conditional GAN (CGAN) (Mirza and Osindero, 2014) and conditional VAE (CVAE) (Sohn et al., 2015).

In our empirical analysis, we conduct tests on two real-world energy time-series. First, a data set from the European Energy Exchange (EEX) that includes hourly day-ahead power prices, available at SMARD.de. Second, a data set from an IEEE-dataport competition (Farrokhabadi et al., 2022) with hourly electricity loads from an unnamed but real city with a median consumption of 1.1 GW. We train over 800 generative models on these data sets and automate the hyper-parameter search using the state-of-the-art optimization tool Optuna (Akiba et al., 2019). Here we apply the maximum mean discrepancy and the Hausdorff discrepancy as optimization objective and also propose the use of both discrepancies as a final goodness-of-fit for time-series forecasting tasks. Finally, we evaluate the predictive power of generative models on out-of-sample test sets using the discrepancy measures as well as point-wise error metrics.

Our results show that generative models, which were optimized by the maximum mean discrepancy, outperform the models optimized by the Hausdorff discrepancy on both data sets. In a visual inspection and an additional analysis of the prediction interval, we find those models to generate larger prediction intervals as well as more accurate sample median predictions. Comparing several generative model frameworks, our results are not in favour of one framework. Instead, multiple generative models should be evaluated to select the most suitable for the respective task.

1.1 Publication Details

Paper I (Chapter 2):
Firm Efficiency and Stock Returns during the COVID-19 Crisis

Authors:
Daniel Neukirchen, Nils Engelhardt, Miguel Krause, and Peter N. Posch

Abstract:
We investigate the relationship between firm efficiency and stock returns during the COVID-19 pandemic. We find that highly efficient firms experienced at least 9.44 percentage points higher cumulative returns during the market collapse. A long-short portfolio consisting of efficient and inefficient firms would have also yielded a significantly positive weekly return of 3.53% on average. Overall, our results show that firm efficiency has significant explanatory power for stock returns during the crisis period.

Publication Details:
Finance Research Letters (2021), 102037.
`https://doi.org/10.1016/j.frl.2021.102037`

Paper II (Chapter 3):

The Value of (Private) Investor Relations during the COVID-19 Crisis

Authors:

Daniel Neukirchen, Nils Engelhardt, Miguel Krause, and Peter N. Posch

Abstract:

We investigate the value of investor relations (IR) and find firms with strong IR to experience between five and eight percentage points higher stock returns than those with weak IR during the COVID-19 crisis. Firms with better-quality IR are also associated with higher investor loyalty and appear to have attracted significantly more institutional investors over the crisis period. This suggests that a firm's IR contributes to value generation by enhancing credibility with shareholders and by diversifying its shareholder base. After decomposing IR into public and private transmission channels, we find the private IR function to be the main driver of our results.

Publication Details:

The Journal of Banking & Finance (2022), 106450.
https://doi.org/10.1016/j.jbankfin.2022.106450

Paper III (Chapter 4):
How to Trust Generative Probabilistic Models for Time-Series Data?

Authors:
Nico Piatkowski, Peter N. Posch and Miguel Krause

Abstract:
Generative machine learning methods deliver unprecedented quality in the fields of computer vision and natural language processing. When comparing models for these task, the user can fast and reliably judge generated data with her bare eye—for humans, it is easy to decide whether an image or a paragraph of text is realistic. However, generative models for time-series data from natural or social processes are largely unexplored, partially due to a lack of reliable and practical quality measures. In this work, measures for the evaluation of generative models for time-series data are studied—in total, over 1000 models are trained and analyzed. The well-established maximum mean discrepancy (MMD) and our novel proposal: the Hausdorff discrepancy (HD) are considered for quantifying the disagreement between the sample distribution of each generated data set and the ground truth data. While MMD relies on the distance between mean-vectors in an implicit high-dimensional feature space, the proposed HD relies on intuitive and explainable geometric properties of a "typical" sample. Both discrepancies are instantiated for three underlying distance measures, namely Euclidean, dynamic time warping, and Frechét distance. The discrepancies are applied to evaluate samples from generative adversarial networks, variational autoencoders, and Markov random fields. Experiments on real-world energy prices and humidity measurements suggest, that considering a single score is insufficient for judging the quality of a generative model.

Publication Details:
In Proceedings of the International Conference on Learning and Intelligent Optimization (2021), (pp. 283-298).
`https://doi.org/10.1007/978-3-030-92121-7_23`

Paper IV (Chapter 5):
Deep Generative Models for Probabilistic Time Series Forecasting

Authors:
Miguel Krause and Peter N. Posch

Abstract:
We investigate the predictive power of deep generative models on energy times series. As electricity markets face high levels of uncertainty due to the climate change and the ongoing COVID-19 pandemic, decision-makers can particularly benefit from probabilistic forecasts from generative models to obtain robust estimates for underlying risks and predictions. Particularly, we use the conditional forms of generative adversarial networks (CGAN) and variational autoencoders (CVAE) to generate probabilistic forecasts on two real-world energy data sets, and propose a goodness-of-fit for generative models in a forecasting setting using the maximum mean and Hausdorff discrepancies. In total, 800 models are trained and evaluated using automated Optuna optimization. Our experiments show that the generative models optimized with the maximum mean discrepancy generate realistic and accurate time-series forecasts.

Publication Details:
Working Paper. Submitted to the AAAI Conference 2023 on Artificial Intelligence.

2 Firm Efficiency and Stock Returns during the COVID-19 Crisis

The following is based on Neukirchen et al. (2021a).

3 The Value of (Private) Investor Relations during the COVID-19 Crisis

The following is based on Neukirchen et al. (2021b).

3.3.2 Key Variables

4 How to Trust Generative Probabilistic Models for Time-Series Data?

The following is based on Piatkowski et al. (2021b).

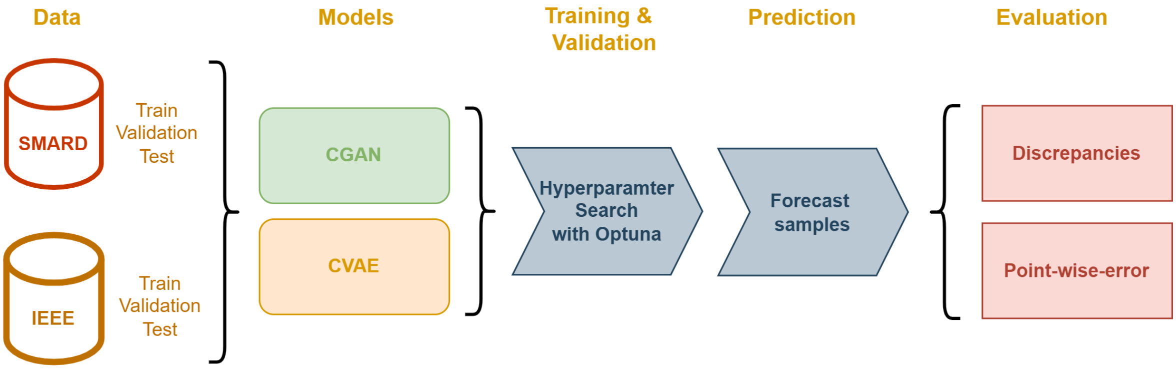
5 Deep Generative Models for Probabilistic Time Series Forecasting

The following is based on Krause and Posch (2022).

5.1 Introduction

The global electricity markets are confronted with two major challenges: one, the rising impact of renewable energy production as a consequence of the transition to a carbon-free society in order to limit climate change, and two, more recently, the ongoing COVID-19 pandemic, which has resulted in massive changes in energy consumption behaviour in industry and households. However, both difficulties increase the uncertainty in the context of decision-making for both energy producers and distributors. As a result, anticipating power prices and demand is crucial for grid operators, as the power generation has to balance the consumption due to the lack of storage capacity compared to the net load (de Vilmarest and Goude, 2021). Point forecasting and heuristic Monte Carlo approaches are commonly used in the energy industry to handle this problem, although both have significant limitations. While Monte Carlo simulations are often based on theoretical assumptions that do not always hold in reality, point forecasts are deterministic and can lead to low-quality predictions in a highly stochastic environment. To alleviate these limitations we apply generative probabilistic models, which are widely used in the machine learning community for generating artificial images (Karras et al., 2018, 2019; Kang and Park, 2020) or in natural language processing (de Masson d'Autume et al., 2019; Li et al., 2017). Generative models aim to learn the distribution of the underlying data generating process and can capture inherent uncertainty through latent variables which account for components of the target that are not explainable by the observed input (Henaff et al., 2017). We can use generative models as drop-in samplers for Monte Carlo techniques and create forecast samples and

Figure 5.1: Experiment framework



prediction intervals without the need for theoretical assumptions (Piatkowski et al., 2021a; Dumas et al., 2022).

In this work, we focus of on two of the most popular frameworks for unsupervised generative modeling, namely generative adversarial networks (GAN) and variational autoencoders (VAE). In order to enforce valid forecast sample scenarios, we employ the conditional forms of these generative models. Conditional models allow us to include additional information and delimit the sample space to a smaller subspace of realistic forecasting samples. To improve forecast sampling quality we combine different conditional feature types of previous studies (Koochali et al., 2019; Fu et al., 2019).

We conduct experiments on two energy data sets to assess the predictive power of our generative probabilistic models. We use the SMARD day-ahead electricity prices data set from Piatkowski et al. (2021a) and an electricity loads data set from a recent IEEE-dataport competition ("day-ahead-electricity-demand-forecasting-post-covid-paradigm") (Farrokhabadi et al., 2022). Figure 5.1 provides the general framework of the experiments. For reproducibility, all data sets and code are available online after acceptance.[1] In summary, this paper makes the following contributions:

1. We investigate the predictive power of conditional generative models on two types of energy time-series, day-ahead prices and electricity loads, to demonstrate the benefits of deep generative models for different aspects of the energy market.

2. Up to our knowledge, we are the first study that applies the Hausdorff and the maximum mean discrepancy in a forecasting setting as goodness-of-fit of

[1] https://github.com/firrm/ConditionalGenerativeModels

generative models. Furthermore, we automate the hyper-parameter search using the state-of-the-art optimization tool Optuna (Akiba et al., 2019) based on both discrepancies.

3. We combine multiple conditional features types in our generative models and generate fully data driven multi-step ahead forecasts.

5.2 Related Work

In this section, we provide relevant studies concerning generative models for time-series. An overview is provided in Table 5.1.

Most of the existing literature on generative models for time-series focuses on learning the underlying data distribution from the historical data without any additional conditional features. For example, Ge et al. (2020) examine GANs and VAEs on electricity load data and show that the generative models are able to capture the spatial-temporal correlation of daily load profiles. Piatkowski et al. (2021a) use GANs, VAEs and Markow random fields to generate realistic samples of electricity prices and investigate the use of discrepancies as evaluation measures. Cramer et al. (2022) apply GANs, WGANs and VAEs to energy time-series and discuss validation methods for generated scenario data. Wu et al. (2021) propose a deep generative model based on dynamic Gaussian mixture noise (DGM^2) to predict sparse multivariate time-series, and show its robustness and effectiveness on real-life data sets.

While these approaches are adequate for general sample generation, recent papers include conditional information such as numerical or categorical features to generate samples for forecasting tasks. For example, Dumas et al. (2022) show the forecasting competitiveness of normalizing flows, CGANs and CVAEs on electricity load, wind and solar time-series using weather forecasts as conditions. Ravuri et al. (2021) apply a CGAN on a radar field data and show that the probabilistic forecasts improve the forecast value and provide fast and accurate short-term weather predictions. Koochali et al. (2019) introduce the ForGAN, a conditional GAN that uses the lagged target values as conditional input and show that the ForGAN outperforms a comparable discriminative model. Salazar et al. (2022) employ a CVAE and generate reliable wind time-series forecasts using weather forecasts and spatio-temporal encodings as

Table 5.1: Overview of related work about generative models for time-series forecasting. The column OPT indicated which kind of hyper-parameter optimization was conducted.

Paper	Models	Data sets	OPT
This work	CGAN, CVAE	electricity prices electricity loads	Optuna
Dumas et al. (2022)	NF, CGAN, CVAE	solar, wind, electricity loads	hand-tuning
Ravuri et al. (2021)	CGAN	radar field	hand-tuning
Koochali et al. (2019)	CGAN	Lorenz, Mackey-Glass, Internet Traffic	genetic algo.
Salazar et al. (2022)	CVAE	wind speed, wind power	hand-tuning
Jeha et al. (2021)	PSA-GAN	electricity loads, solar, traffic, M4	grid search
Ge et al. (2020)	NICE, GAN, VAE	London smart meter	hand-tuning
Piatkowski et al. (2021a)	GAN, VAE, MRF	electricity prices intel humidity readings	grid search
Cramer et al. (2022)	GAN, WGAN, VAE	wind, solar electricity prices	hand-tuning
Wu et al. (2021)	DGM^2	weather, air quality hospital admission records	hand-tuning

conditional features and and an increase in performance compared to a conventional numerical weather prediction model. Jeha et al. (2021) introduce the PSA-GAN which uses the popular self-attention architecture (Vaswani et al., 2017) and progressive growing (Karras et al., 2018), and show that the PSA-GAN can improve downstream forecasting tasks on various data sets. Furthermore, Fu et al. (2019) propose the use of CGANs with categorical and continuous variables as conditions and show that CGANs can be used for a variety of financial time-series applications.

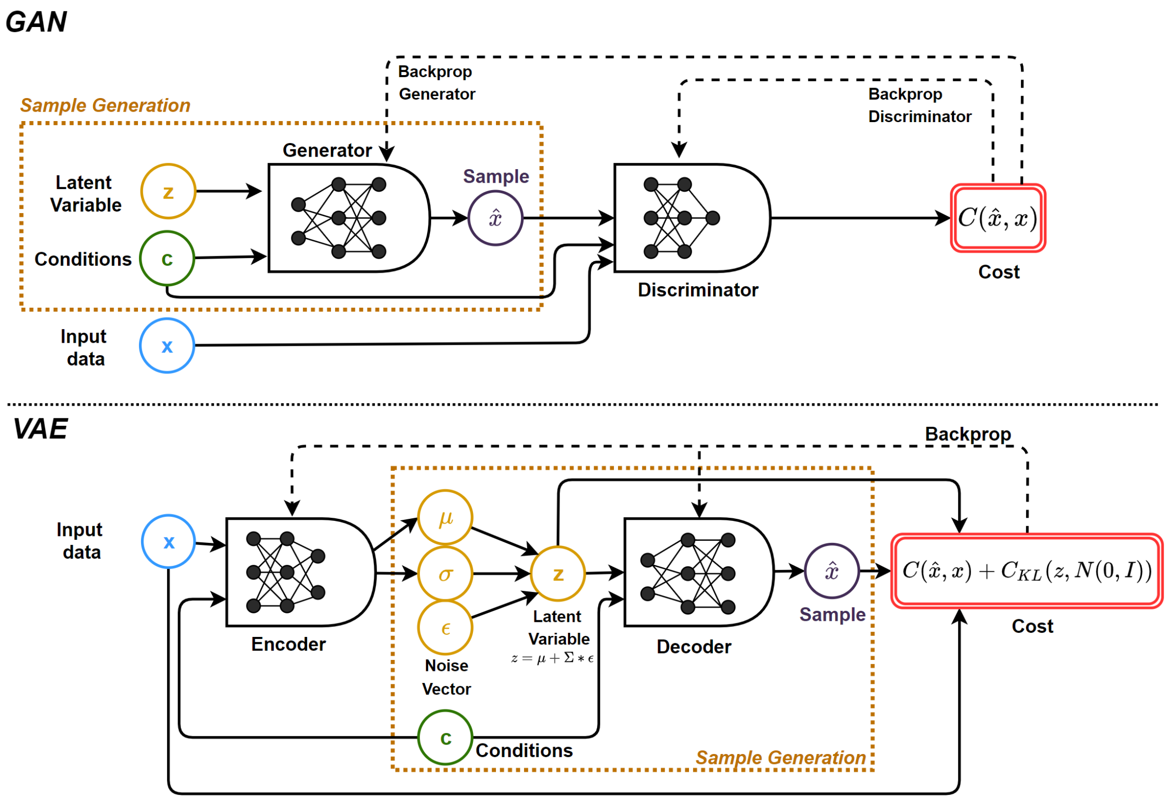
5.3 Problem Formulation

In this paper we focus on the task of multi-step ahead forecasting for energy time-series. Formally, we are given a data set $\mathcal{D} = \{x^{(t)}, c^{(t)}\}_{1\leq t\leq T}$ of targets $x^{(t)} = (x_1, \dots, x_N)^{(t)}$, consisting of N values corresponding to measurements over time, such as hourly electricity prices over a day and additional conditions $c^{(t)} = (c_1, \dots, c_M)^{(t)}$ consisting of M feature values, such as the hourly weather forecast or categorical data such as the weekday or the season. Our goal is now to generate multiple forecast samples $\hat{x}^S_{pred} = (\hat{x}^{(t+1)}, \dots, \hat{x}^{(t+k)})^{(S)}$ with forecast horizon k and number of samples S, given known conditions $c_{pred} = (c^{(t+1)}, \ldots, c^{(t+k)})$. If we include the lagged target realizations $x^{(t-l)}$ as conditions, the forecast horizon reduces to $k = l$.

5.4 Methods

Most of recent generative probabilistic models are based on the use of a differentiable generator network that maps an input from a latent space to the data space (Goodfellow et al., 2016). In this section, we describe two types of unsupervised deep generative models, namely generative adversarial networks and variational autoencoders, which differ in their learning process and how they include the generator in the network structure. However, both frameworks can be modified to include conditional information in their training and generation process.

Figure 5.2: GAN and VAE training and generation



5.4.1 Conditional Generative Adversarial Network (CGAN)

Generative adversarial networks (GAN) (Goodfellow et al., 2014) are one of the most prominent generative frameworks of the recent decade. Their popularity stems mainly from their adversarial training technique, which involves two models competing against each other in a game theoretic scenario. One of the models is a generator network that learns to generate realistic data and tries to fool the other network, called the discriminator. The discriminator simultaneously learns to improve its ability to distinguish between real and artificial data from the generator. This process enables the GAN to learn the dynamics of a multivariate stochastic process without having to make explicit assumptions about its form (Madeka et al., 2018). Since we want to use GANs to predict time-series data, we include additional information to delimit the generation process and to generate conditional samples. We therefore use a conditional generative adversarial network (CGAN) (Mirza and Osindero, 2014). Here, two neural networks are trained simultaneously in a mini-max game: The generator $G : \mathcal{Z} \times \mathcal{C} \to \mathcal{X}$ takes a random vector Z from a measure $\mathbb{Q}$ over $\mathcal{Z}$, such as a multivariate Gaussian and conditions C from the conditional space $\mathcal{C}$ as inputs and generates a new point in the data domain $\mathcal{X}$. The discriminator $D : \mathcal{X} \times \mathcal{C} \to [0; 1]$ estimates the probability of drawing a data point from the original data distribution $\mathbb{P}$. In other words, G and D play a two-player minimax game to determine their parameters with value function $V(D; G)$:

$$\min_G \max_D V(G; D) = \mathbb{E}_{\mathbb{P}}[\log D(X|C)] + \mathbb{E}_{\mathbb{Q}}[\log(1 - D(G(Z|C)))] \tag{5.4.1}$$

With sufficient capacity and training iterations, the minimax game reaches a Nash equilibrium where the discriminator can no longer distinguish between generated and real data, which implies that the generator can produce realistic samples as if they came from the original data distribution (Chen et al., 2018).

To predict future signals $\tilde{\mathcal{X}}$ we pass random input vectors from the prior $\mathbb{Q}$ and conditional features $\mathcal{C}$ to the trained generator. The generator maps these inputs to the data space $\mathcal{X}$. This allows us to predict a wide range of possible future samples. Additionally, we can easily compute a point forecast by the sample median and are able to estimate prediction intervals using the sample quantiles to calculate a range of the most likely predictions. The training and sample generation process is depicted in figure 5.2.

5.4.2 Conditional Variational Auto-Encoder (CVAE)

A variational autoencoder (VAE) (Kingma and Welling, 2013; Rezende et al., 2014) is a generative latent variable model based on a variational Bayesian approach. A VAE learns an approximation of the true underlying data distribution $\mathbb{P}_\theta(x)$ by encoding latent attributes of the data X as probability distributions in the latent space (Ivanovic et al., 2020). Similar to an autoencoder the VAE consist in two models, an encoder and an decoder, which in practice are parameterized by deep neural networks and can be trained jointly with gradient-based methods. However, the decoder works as a generator network which maps samples from the latent space distribution to the data space (Dumas et al., 2022; Goodfellow et al., 2016).

Similar to the GAN we focus on the conditional form, the conditional variational autoencoder (CVAE) (Sohn et al., 2015). Given conditional variables C and latent variables Z from an associated prior $\mathbb{P}_\theta(z|c)$ the encoder part of the VAE approximates it's true intractable posterior distribution $\mathbb{P}_\theta(z|x,c) \propto \mathbb{P}(z|c)\mathbb{P}_\theta(x|z,c)$ by encoding the data into parameters of the approximate posterior distribution $\mathbb{Q}_\phi(z|x,c)$, parameterized by ϕ. The decoder part $\mathbb{P}_\theta(x|z,c)$, parameterized by θ, then learns to reconstruct a sample from the posterior. We follow Kingma and Welling (2013) and choose as a prior of the latent variables a multivariate Gaussian, where the mean and variance are predicted by the encoder network. To perform inference on the marginal likelihood $\mathbb{P}_\theta(x|c)$ and train the VAE we maximize the evidence lower bound (ELBO):

$$\begin{aligned}\log \mathbb{P}_\theta(x|c) &\geq L(\phi,\theta;x,c)\\ &= \mathbb{E}_{\mathbb{Q}(z|x,c)}[\log \mathbb{P}_\theta(z,x|c) - \log \mathbb{Q}_\phi(z|x,c)]\\ &= -D_{\mathrm{KL}}(\mathbb{Q}_\phi(z|x,c)||\mathbb{P}_\theta(z|c)) + \mathbb{E}_{\mathbb{Q}_\phi(z|x,c)}[\log \mathbb{P}_\theta(x|z,c)]\end{aligned} \tag{5.4.2}$$

where D_{KL} is the Kullback-Liebler divergence that regularises the posterior $\mathbb{Q}_\phi(z|x,c)$ towards the prior $\mathbb{P}_\theta(z)$. Optimization and sampling is typically carried out using stochastic gradient descent and the reparameterization trick (Kingma and Welling, 2013). For example, let $z \sim \mathbb{Q}_\phi(z|x,c) = \mathbb{N}(\mu,\Sigma)$, where μ,Σ are predicted by the encoder. We then sample from $\mathbb{Q}$ by reparameterizing $z = \mu + \Sigma * \epsilon$, with $\epsilon \sim \mathcal{N}(0,I)$

(Camuto et al., 2021; Piatkowski et al., 2021a) .

To generate new forecast samples, random samples from the learned latent space distribution and the conditional feature data are passed to the decoder network. In this way, we can generate multiple prediction scenarios and estimate a point forecast as well as prediction intervals by calculating the sample median and respective quantiles. The training and sample generation process is shown in figure 5.2.

5.5 Empirical Evaluation

In this section, we provide all the details of the experimental setup and the final results. This includes the data sets, hyper-parameter search, goodness-of-fit and evaluation measures.

5.5.1 Data

SMARD. The first data set contains the day-ahead prices of the German power market starting from 01 January 2014 to 30 June 2020, available at SMARD.de. The European Energy Exchange (EXX) runs a daily blind auction, which means all 24 prices for each hour of the day are set at the same time through a clearing process that matches supply and demand. The traded electricity is delivered the next day. Due to the fact that supply and demand of electricity needs to be balanced, electricity prices have characteristics that differ from those of financial assets or other commodities. For example, negative prices can occur on days with low demand such as weekends or public holidays and a strong infeed of renewable energy, as storage capacities are still limited (Paraschiv et al., 2014). The data set contains $T = 2373$ data points of length $N = 24$. As conditional features, we add the one day lagged day-ahead prices, and categorical variables as one-hot vectors for months, seasons and weekends. Thus, we aim to generate one-day-ahead forecasts. We split our data set into a training part from 01 January 2014 to the 31 December 2019 and a test part from 01 January to 30 June 2020. Than we split the training part accordingly to the common $80 : 20$ ratio into a training and a validation set to obtain a data set that we can use to find the best hyper-parameters. With this setup, we can shed light on the forecast ability of our models during the beginning of the COVID-19 pandemic using pre-pandemic training and validation data. Figure 5.3 shows the day ahead prices for the train($T_{Train} = 1766$), validation ($T_{Validation} = 442$) and test sets ($T_{Test} = 165$).

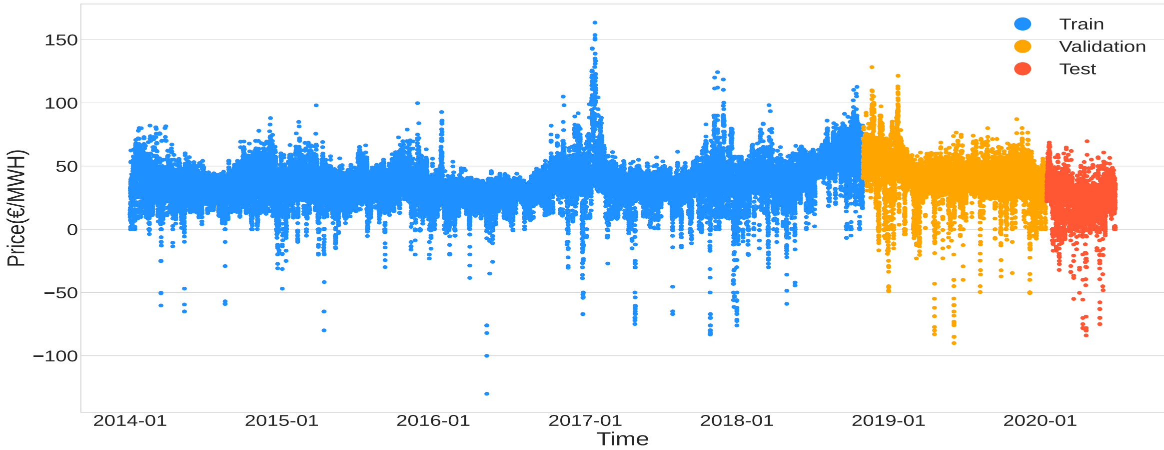


Figure 5.3: Day-ahead electricity prices from the *SMARD* data set. The figure is our own contribution.

IEEE-dataport. The second data set contains electricity demand data from the recent open-access IEEE-dataport competition ("Day-Ahead Electricity Demand Forecasting: Post-COVID Paradigm") (Farrokhabadi et al., 2022). This challenge concerns the POST-COVID aspects of electricity load forecasting and is based on real data of an unknown city over a over 30-day test period from 16 January 2021 up to 16 February 2021 (Ziel, 2021). The training data covers the period from 18 March 2017 to 15 January 2020. We split the training data into training and validation sets using the common 80:20 ratio and use the validation set to find the best hyper-parameters for the generative models. All data is provided in a hourly shape, therefore each conditional feature variable and the target have dimension $N = 24$. The data includes the historical electricity demand (MW), historical weather observations (air pressure (kpa), cloud cover (%), humidity(%), temperature (C), wind direction (deg), wind speed (kmh)) and weather forecasts (air pressure (kpa), cloud cover (%), temperature (C), wind direction (deg), wind speed (kmh)) and is obtained from an unknown real-world utility and weather service provider. A detailed data overview is described in Ziel (2021) and de Vilmarest and Goude (2021). We add additional categorical features (seasons, weeks, weekend) as one-hot vectors to pass some general information to the models and the one week lagged historical weather observations as well as the one week lagged electricity loads. We use the historical data delayed by one week because the original IEEE-dataport challenge only allows the use of historical data from at least 48 hours before the forecast time. We therefore aim to generate one-week-ahead forecasts. We adjust missing data points in the data set by linear interpolation. Figure 5.4 shows the electricity loads for the train ($T_{Train} = 1120$), validation ($T_{Validation} = 280$) and test sets ($T_{Test} = 32$).

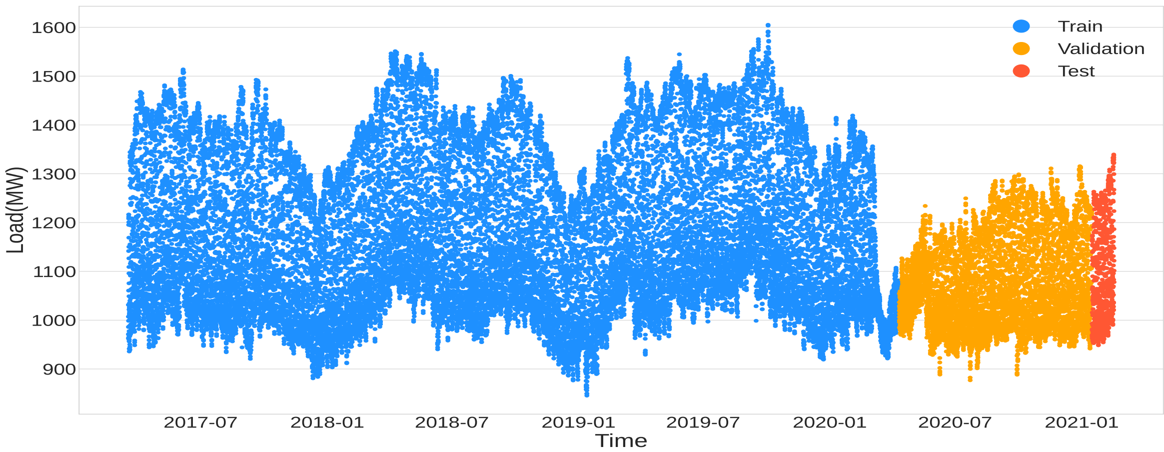


Figure 5.4: Electricity loads from *IEEE-dataport* challenge 2021. The figure is our own contribution.

5.5.2 Hyper-parameter and Network Architecture Search

We adopt the network architectures and hyper-parameters from Piatkowski et al. (2021a) and add conditional input layers to the vanilla models. The limits for the hyper-parameter search of the CGAN and CVAE models are listed in table 5.2. All experiments are implemented in TensorFlow (Abadi et al., 2016). The training procedure is accelerated by two NVIDIA GeForce RTX 2080 Ti GPUs. To stabilize and accelerate the training process, we normalize the data based on the mean and standard derivation of the training data set.

Table 5.2: Hyper-parameter search space for generative probabilistic models.

(a) CGAN hyper-parameter limits.

Parameter	**Values**
Iterations	100, 2000
Batch size	32, 512
Hidden units	32, 512
Hidden layer	1, 8
Droprate	0.0, 0.5
Iterations D	1, 5
Learning rate D	0.001, 0.009
Learning rate G	0.0001, 0.0009

(b) CVAE hyper-parameter limits.

Parameter	**Values**
Epochs	10, 300
Batch size	32, 512
Latent dim	128, 512
Hidden layer	1, 8
Hidden units	32, 512
Learning rate	0.0001, 0.001

We train the models using the Optuna optimization tool (Akiba et al., 2019), which optimizes the hyper-parameters according to an objective value using a Tree-structured Parzen Estimator algorithm (Bergstra et al., 2011). Since generative models aim to estimate the underlying data probability distribution, we follow Piatkowski et al. (2021a) and apply the maximum mean discrepancy and Hausdorff discrepancy as objective values. In general, these discrepancies serve as a goodness-of-fit for generative models and determine how well the models have learned the underlying data distribution. The discrepancies are defined as follows:

Maximum Mean Discrepancy. The maximum mean discrepancy (MMD) (Gretton et al., 2006, 2012) compares samples from two probability distributions by mapping the distance between the embedding of the two data samples into a reproducing kernel Hilbert space (RKHS).

Definition 5.5.1 (Maximum Mean Discrepancy)**.** Let k be the kernel of a RKHS $\mathcal{F}_k$ of functions on a topological space $\mathcal{X}$. We define the mean embedding of a probability measure $\mathbb{P}$ in $\mathcal{F}_k$ as $\mu_k(\mathbb{P}) \in \mathcal{F}_k$ such that $\mathbb{E}_{\mathbb{P}} f(X) = \langle f, \mu_k(\mathbb{P})\rangle_{\mathcal{F}_k}$ for all $f \in \mathcal{F}_k$. Then the MMD in $\mathcal{F}_k$ between two probability measures $\mathbb{P}$ and $\mathbb{Q}$ is defined as

$$\mathrm{MMD}_{\mathcal{F}_k}(\mathbb{P}, \mathbb{Q}) = \sup_{f \in \mathcal{F}_k} ||\mathbb{E}_{\mathbb{P}} f(X) - \mathbb{E}_{\mathbb{Q}} f(X)||^2_{F_k} \tag{5.5.1}$$

Given two sets $\mathcal{D}_{\mathbb{P}} = \{x_1, \ldots, x_m\} \sim \mathbb{P}$ and $\mathcal{D}_{\mathbb{Q}} = \{y_1, \ldots, y_n\} \sim \mathbb{Q}$ an unbiased estimate of the MMD is defined as

$$\begin{aligned}\tilde{\mathrm{MMD}}_{\mathcal{F}_k}(\mathbb{P},\mathbb{Q}) = \frac{1}{|\mathcal{D}_\mathbb{P}|^2}\sum_{x\in\mathcal{D}_\mathbb{P}}\sum_{y\in\mathcal{D}_\mathbb{P}} k(x,y) - \frac{2}{|\mathcal{D}_\mathbb{P}||\mathcal{D}_\mathbb{Q}|}\sum_{x\in\mathcal{D}_\mathbb{P}}\sum_{y\in\mathcal{D}_\mathbb{Q}} k(x,y) \\ + \frac{1}{|\mathcal{D}_\mathbb{Q}|^2}\sum_{x\in\mathcal{D}_\mathbb{Q}}\sum_{y\in\mathcal{D}_\mathbb{Q}} k(x,y)\ .\end{aligned} \tag{5.5.2}$$

As in Piatkowski et al. (2021a) we use the radial basis function (RBF) kernel, which is a characteristic kernel. This choice implies that MMD is a metric and that $\mathrm{MMD}(\mathbb{P},\mathbb{Q}) = 0$ if and only if $\mathbb{P} = \mathbb{Q}$ (Sriperumbudur et al., 2010; Sutherland et al., 2016).

The Hausdorff Discrepancy. The Hausdorff discrepancy (HD) (Piatkowski et al., 2021a) is a geometrical discrepancy between two probability distributions and is based on the well-known Hausdorff distance. The Hausdorff distance, which is often used in computer vision tasks (Huttenlocher et al., 1993; Karimi and Salcudean, 2019), measures how far two subsets in metric space are from being isometric. Using the Hausdorff discrepancy, distributions are similar if the generated data points from these distributions are at similar positions.

Definition 5.5.2 (Hausdorff Discrepancy)**.** Let X, Y be random variables both having state space $\mathcal{X}$ and following distributions $\mathbb{P},\mathbb{Q}$, respectively. Given two (random) data sets $\mathcal{D} = \{x_1,\ldots,x_N\}, \mathcal{E} = \{y_1,\ldots,y_N\}$ of size N, containing independent samples from $\mathbb{P}$ and $\mathbb{Q}$, the Hausdorff discrepancy between $\mathbb{P},\mathbb{Q}$ is defined as

$$\mathrm{HD}(\mathbb{P},\mathbb{Q}) = \max\{\mathbb{E}_\mathcal{E}\left[\max_{x\in\mathcal{D}}\min_{y\in\mathcal{E}} d(x,y)\mid\mathcal{D}\right], \mathbb{E}_\mathcal{E}\left[\max_{x\in\mathcal{E}}\min_{y\in\mathcal{D}} d(x,y)\mid\mathcal{D}\right]\}. \tag{5.5.3}$$

with $d(.,.)$ the euclidean distance.

Algorithm 1 shows the pseudo code of the training and optimization process.

Algorithm 1 Optuna optimization

Input: Training set (X_{Train}, C_{Train}),
validation set $(X_{Validation}, C_{Validation})$,
optimization objective $s(.,.)$

initialize hyper-parameters h_1
for trial $i = 1, 2, \ldots, 100$ **do**
  Train the model on (X_{Train}, C_{Train}) with h_i
  Generate 1000 forecast samples $\hat{X}$ with $C_{Validation}$
  Compute objective scores $\hat{s} = s(\hat{X}, X_{Validation})$
  Aggregate scores $\bar{s} = mean(\hat{s})$
  $h_{i+1} \leftarrow OptunaSampler(\bar{s}, h_i)$
end for
return best model based on the minimal score $\bar{s}$

5.5.3 Evaluation

We investigate the performance of the best generative models using the test data set. For this purpose, we generate 1000 forecast samples per model and calculate the HD and MMD between each of these samples and the original test data set. We aggregate the discrepancies using the mean and calculate the standard derivation. In addition, we compute point-wise error metrics which are commonly used in prediction tasks. As point forecast we use the median sample of the generated forecast samples and report the root mean squared error (RMSE) and mean absolute error (MAE)

$$RMSE = \sqrt{\frac{1}{S}\sum_{i}^{N}(x^{(i)} - \hat{x}^{(i)})^2}, \quad MAE = \frac{1}{S}\sum_{i}^{N}|x^{(i)} - \hat{x}^{(i)}| \tag{5.5.4}$$

with S number of target samples $x^{(i)}$, and forecasts $\hat{x}^{(i)}$. However, the discrepancies show how accurately the generative models have learned the underlying data distribution, while the point-wise error metrics examine how well the models have accounted for the conditional features (Koochali et al., 2019).

In an additional test, we want to shed light on the advantages of probabilistic forecasts for decision-makers in energy trading and power generation scheduling. For this purpose, we evaluate the prediction interval coverage percentage (PICP), which indicates the probability that the target values are covered by an $(1-\alpha)\%$ prediction interval. A higher PICP is associated with more target values falling within the constructed prediction interval and idyllically it should be close to its nominal value $(1-\alpha)\%$ (Khosravi et al., 2010).

For example, to construct a 90% prediction interval, we estimate the the 95% and 5% quantiles of the forecast samples as the upper (U_i) and lower (L_i) bounds. The PICP is defined as

$$PICP = \frac{1}{S}\sum_{i=1}^{S}\epsilon_i, \quad \epsilon_i = \begin{cases} 1, & if\ x^{(i)} \in [L_i, U_i] \\ 0, & if\ x^{(i)} \notin [L_i, U_i] \end{cases} \tag{5.5.5}$$

with S number of target samples $x^{(i)}$.

5.5.4 Results

In this section we present the results of our experiments on the out-of-sample test sets. Table 5.3 presents the discrepancies and the point-wise error metrics of the *SMARD* data set. Table 5.4 shows the discrepancies and the point-wise error of the *IEEE-dataport* data set.

Table 5.3: This table shows the forecasting results for the *SMARD* data set with the Hausdorff Discrepancy (HD), maximum mean discrepancy (MMD), mean absolute error (MAE) and root mean squared error (RMSE). For HD and MMD we report the mean and standard derivation over the forecast samples. The objective of the previous hyper-parameter search is noted column OPT. All hyper-parameters for the GAN and VAE models are described Table C1 in the appendix.

		Discrepancies		**Point-wise**	
OPT	**Model**	**HD**	**MMD**	**MAE**	**RMSE**
HD	CGAN	114.695 ±14.146	0.054 ±0.007	7.168	10.731
	CVAE	**100.189±3.766**	0.046 ±0.003	7.020	10.603
MMD	CGAN	174.567 ±9.260	**0.020±0.003**	**6.530**	**10.533**
	CVAE	102.332 ±7.776	0.031 ±0.002	6.817	10.938

For the *SMARD* data set, the GAN (optimized by the MMD objective) outperforms the VAE models in all metrics except the HD measure. In a visual inspection, we can see that the VAE models only provide a small variation in the forecast samples around the median sample forecast, while the GAN models exhibit larger variations and also cover stronger variations of the original day-ahead prices. This is especially true for the GAN, which was optimized by the MMD objective. Figure 5.5 depicts the forecasts and 90% prediction intervals for all models for the first 20 days of the COVID-19 Lockdown in Germany starting from the 22 of March 2020. The PICP evaluation in table 5.5 support these visual results. The GAN prediction interval covers 70% of all target realizations, while the best VAE prediction interval covers only 39% of all realizations.

Table 5.4: This table shows the forecasting results for the *IEEE-dataport* data set with the Hausdorff Discrepancy (HD), maximum mean discrepancy (MMD), mean absolute error (MAE) and root mean squared error (RMSE). The objective of the previous hyper-parameter search is noted column OPT. All hyper-parameters for the GAN and VAE models are described Table C2 in the appendix.

		Discrepancies		**Point-wise**	
OPT	**Model**	**HD**	**MMD**	**MAE**	**RMSE**
HD	CGAN	209.727 ± 35.644	0.116 ± 0.012	21.851	31.414
	CVAE	168.052 ± 15.535	0.092 ± 0.009	19.459	**26.271**
MMD	CGAN	258.952 ± 19.938	0.109 ± 0.005	28.037	42.671
	CVAE	**156.475 ±15.996**	**0.077 ±0.008**	**19.139**	27.666

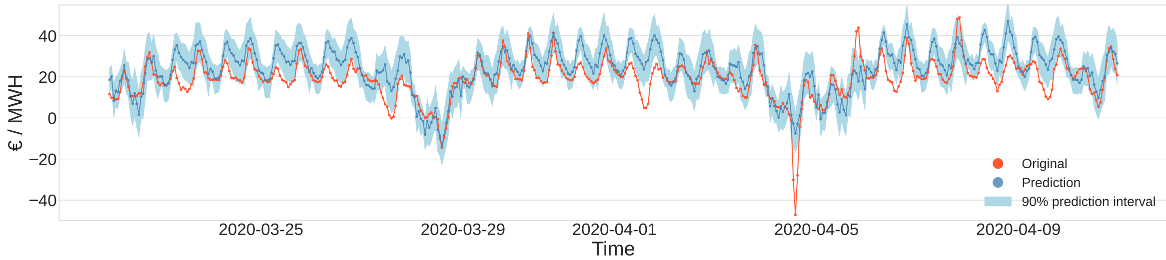
For the *IEEE-dataport* test data, the VAE (optimized by the MMD objective) outperforms the other models in all metrics except the RMSE measure. Again as depicted in Figure 5.6 the MMD optimization is in favour of a model that exhibits a large variation in forecasts and therefore a wide prediction interval while the median forecast is close to the target. The visual results on the prediction intervals are strengthened by the PICP evaluation in table 5.5, where the VAE prediction interval covers 81% of all target realizations. The CGAN (optimized by the HD objective) exhibits the highest PICP of 93% but also high discrepancy values, indicating that the CGAN is able to generate a large variation in forecast samples but these samples are less realistic as they deviate from the geometric shape and statistical properties of the real data. Overall, it can be seen for both data sets that the generative models optimized by the MMD objective outperform the HD optimized models.

Table 5.5: This table shows the prediction interval coverage percentage (PICP) results for the *IEEE-dataport* data set and the *SMARD* data set. The objective of the previous hyper-parameter search is noted column OPT. All hyper-parameters for the GAN and VAE models are described Table C1 and Table C2 in the appendix.

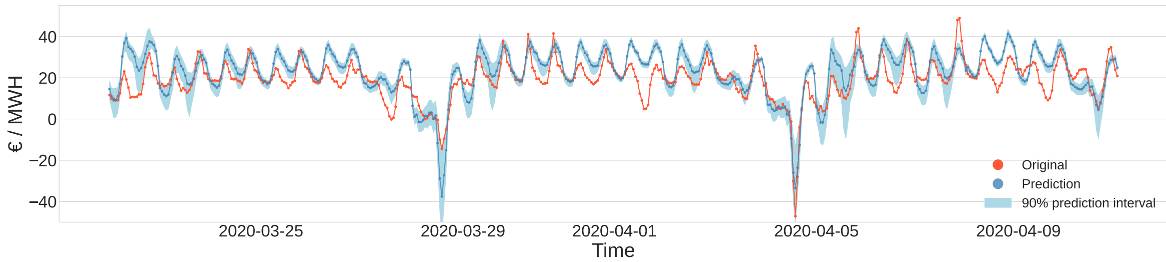
		PICP(%)	
OPT	**Model**	*SMARD*	*IEEE-dataport*
HD	CGAN	0.677	**0.930**
	CVAE	0.389	0.863
MMD	CGAN	**0.696**	0.253
	CVAE	0.256	0.810

Figure 5.5: (a) CGAN (HD), (b) CGAN (MMD), (c) CVAE (HD) and (d) CVAE (MMD) predictions over first 20 days of the lockdown time of the test data set. The optimization objective is denoted in parenthesis. The realized electricity price shapes over time are shown in orange, the median sample forecast in blue and the 90% prediction interval in lightblue. The figure is our own contribution based on data from *SMARD*.

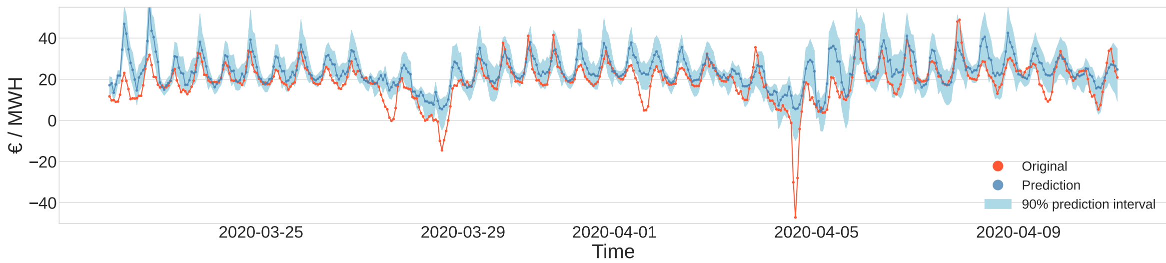
(a) CGAN (HD)



(b) CVAE (HD)



(c) CGAN (MMD)



(d) CVAE (MMD)

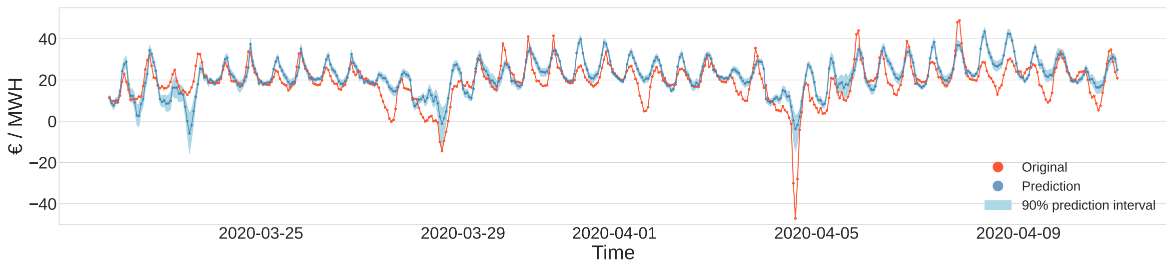
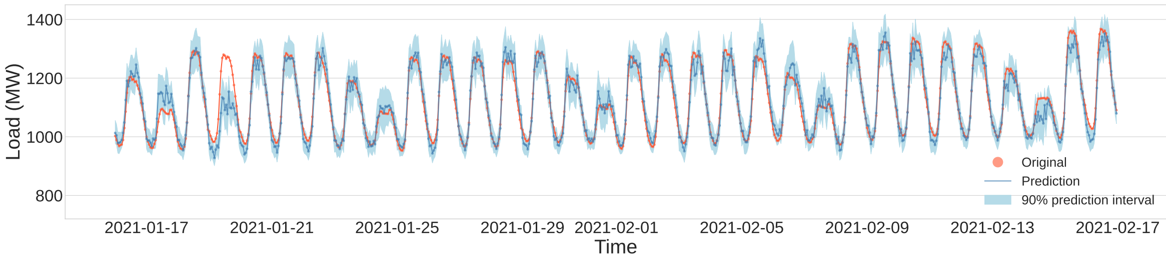
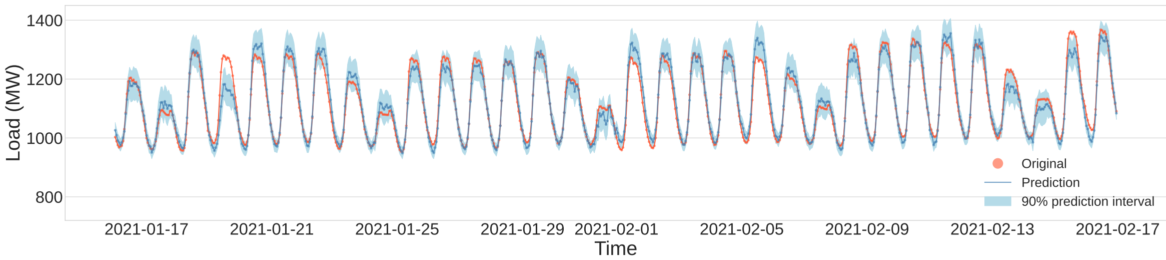


Figure 5.6: (a) CGAN (HD), (b) CGAN (MMD), (c) CVAE (HD), (d) CVAE (MMD) predictions of the *IEEE-dataport* test sample. The optimization objective is denoted in parenthesis. The realized electricity loads over the time are shown in orange, the median sample forecast in blue and the 90% prediction interval in lightblue. The figure is our own contribution based on data from *IEEE-dataport*.

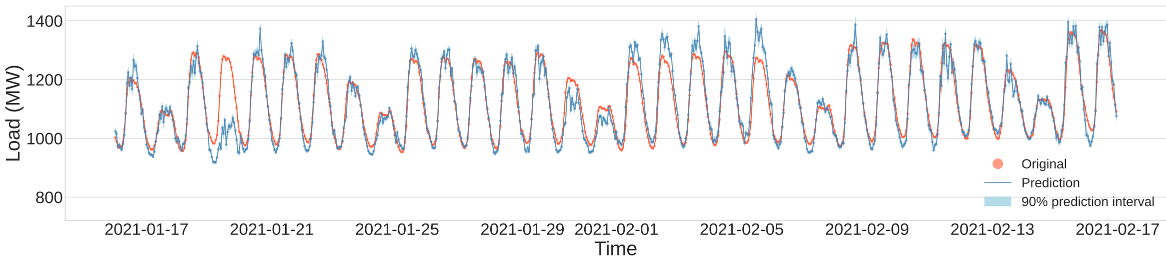
(a) CGAN (HD)



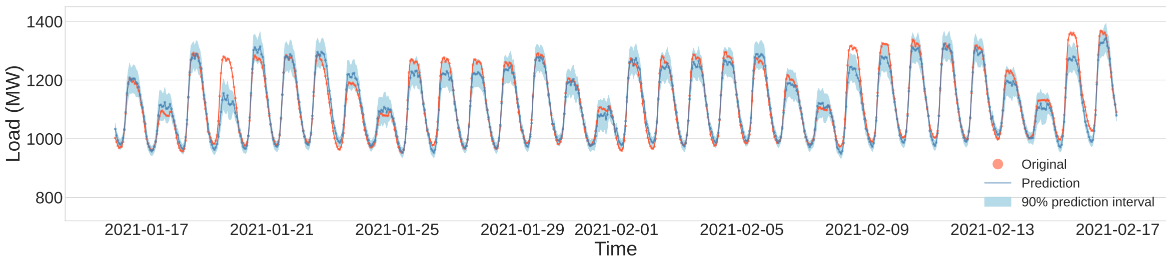
(b) CVAE (HD)



(c) CGAN (MMD)



(d) CVAE (MMD)



5.6 Conclusion

In this work, we investigated the predictive power of deep generative models for real world energy time-series. For this purpose, we trained and optimized several hundred models to find the optimal hyper-parameters, using the maximum mean discrepancy and the Hausdorff discrepancy as optimization objectives. In the final out-of-sample evaluation, we find that generative models optimized with the maximum mean discrepancy perform better than those optimized with the Hausdorff discrepancy. Interestingly, the maximum mean discrepancy favours models that generate larger variations in forecast samples and a more accurate forecast sample median. When comparing the model classes, the GAN model performs better in the electricity price data set, while the VAE model performs better in the electricity load data set. Overall, our results argue for the use of the maximum mean discrepancy when optimizing the hyper-parameters of a generative probabilistic model for time-series forecasting and do not argue for a superior deep generative model class for energy time-series forecasting tasks.

A Appendix for Chapter 2

B Appendix for Chapter 3

C Appendix for Chapter 5

Table C1: Final hyper-parameter for the SMARD dataset optimized with Optuna

(a) GAN hyperparameter

Parameter	(HD)	(MMD)
Iterations	1200	1000
Batch size	64	288
Hidden units	320	352
Hidden layer	4	2
Droprate	0.0	0.0
Learning rate G	0.00061	0.00069
Learning rate D	0.00480	0.00100
Iterations D	3	3

(b) VAE hyperparameter

Parameter	(HD)	(MMD)
Epochs	260	250
Batch size	512	128
Latent dimension	512	128
Hidden layer	7	2
Hidden units	64	352
Learning rate	0.0010	0.0008

Table C2: Final hyper-parameter for the IEEE dataset optimized with Optuna

(a) GAN hyperparameter

Parameter	(HD)	(MMD)
Iterations	1650	1850
Batch size	32	224
Hidden units	224	96
Hidden layer	2	5
Droprate	0.1	0.1
Learning rate G	0.00023	0.00086
Learning rate D	0.00440	0.00210
Iterations D	5	2

(b) VAE hyperparameter.

Parameter	(HD)	(MMD)
Epochs	280	230
Batch size	384	256
Latent dimension	496	512
Hidden layer	1	2
Hidden units	288	448
Learning rate	0.0003	0.0001

Bibliography

Abadi, M., A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, et al. (2016). Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint*. arXiv:1603.04467.

Acharya, V. V. and S. Steffen (2020). The risk of being a fallen angel and the corporate dash for cash in the midst of COVID. *Review of Corporate Finance Studies 9*(3), 430–471. doi:10.1093/rcfs/cfaa013.

Agarwal, V., A. Liao, R. Taffler, and E. Nash (2008). The impact of effective investor relations on market value. *SSRN Electronic Journal*. doi:10.2139/ssrn.1102644.

Aigner, D., C. K. Lovell, and P. Schmidt (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics 6*(1), 21–37. doi:10.1016/0304-4076(77)90052-5.

Akiba, T., S. Sano, T. Yanase, T. Ohta, and M. Koyama (2019). Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 2623–2631. doi:10.1145/3292500.3330701.

Albuquerque, R., Y. Koskinen, S. Yang, and C. Zhang (2020). Resiliency of environmental and social stocks: An analysis of the exogenous COVID-19 market crash. *Review of Corporate Finance Studies 9*(3), 593–621. doi:10.1093/rcfs/cfaa011.

Alfaro, L., A. Chari, A. N. Greenland, and P. K. Schott (2020). Aggregate and firm-level stock returns during pandemics, in real time. *NBER Working Paper Series*. doi:10.3386/w26950.

Alt, H. and M. Godau (1995). Computing the fréchet distance between two polygonal curves. *International Journal of Computational Geometry & Applications 5*(01n02), 75–91. doi:10.1142/S0218195995000064.

Altig, D., S. Baker, J. M. Barrero, N. Bloom, P. Bunn, S. Chen, S. J. Davis, J. Leather, B. Meyer, E. Mihaylov, et al. (2020). Economic uncertainty before and during the COVID-19 pandemic. *Journal of Public Economics 191*, 104274. doi:10.1016/j.jpubeco.2020.104274.

Arlitt, M., M. Marwah, G. Bellala, A. Shah, J. Healey, and B. Vandiver (2015). Iotabench: an internet of things analytics benchmark. In *Proceedings of the 6th ACM/SPEC International Conference on Performance Engineering*, pp. 133–144. doi:10.1145/2668930.2688055.

Baker, S. R., N. Bloom, S. J. Davis, and S. J. Terry (2020). COVID-induced economic uncertainty. *NBER Working Paper Series*. doi:10.3386/w26983.

Bartov, E., L. Faurel, and P. S. Mohanram (2018). Can Twitter help predict firm-level earnings and stock returns? *Accounting Review 93*(3), 25–57. doi:10.2308/accr-51865.

Bergstra, J., R. Bardenet, Y. Bengio, and B. Kegl (2011). Algorithms for hyper-parameter optimization. *Advances in neural information processing systems 24*. Available at: https://proceedings.neurips.cc/paper/2011/file/86e8f7ab32cfd12577bc2619bc635690-Paper.pdf. Last checked on Aug 16, 2022.

Brennan, M. J. and C. Tamarowski (2000). Investor relations, liquidity, and stock prices. *Journal of Applied Corporate Finance 12*(4), 26–37. doi:10.1111/j.1745-6622.2000.tb00017.x.

Brochet, F., P. Limbach, D. Bazhutov, A. Betzer, and M. Doumet (2021). The Supply and Effectiveness of Investor Relations in Insider- vs. Outsider-Oriented Markets. *SSRN Electronic Journal*. doi:10.2139/ssrn.3331936.

Brophy, E., Z. Wang, and T. E. Ward (2019). Quick and easy time series generation with established image-based gans. *arXiv preprint*. arXiv:1902.05624.

Buchanan, B. (2019). *Artificial intelligence in finance*. The Alan Turing Institute.

Bushee, B. J. and G. S. Miller (2012). Investor relations, firm visibility, and investor following. *Accounting Review 87*(3), 867–897. doi:10.2308/accr-10211.

Bzdok, D., M. Krzywinski, and N. Altman (2018). Machine learning: supervised methods. *Nature methods 15*(1), 5.

Camuto, A., M. Willetts, S. Roberts, C. Holmes, and T. Rainforth (2021). Towards a theoretical understanding of the robustness of variational autoencoders. In *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, Volume 130, pp. 3565–3573. PMLR. Available at: http://proceedings.mlr.press/v130/camuto21a/camuto21a.pdf. Last checked on Aug 16, 2022.

Carhart, M. M. (1997). On persistence in mutual fund performance. *Journal of Finance 52*(1), 57–82. doi:10.1111/j.1540-6261.1997.tb03808.x.

Cepoi, C.-O. (2020). Asymmetric dependence between stock market returns and news during COVID-19 financial turmoil. *Finance Research Letters 36*, 101658. doi]:10.1016/j.frl.2020.101658.

Chahine, S., G. Colak, I. Hasan, and M. Mazboudi (2020). Investor relations and IPO performance. *Review of Accounting Studies 25*(2), 474–512. doi:10.1007/s11142-019-09526-8.

Chambers, R. G., Y. Chung, and R. Färe (1996). Benefit and distance functions. *Journal of Economic Theory 70*(2), 407–419. doi:/10.1006/jeth.1996.0096.

Chapman, K., G. S. Miller, and H. D. White (2019). Investor relations and information assimilation. *Accounting Review 94*(2), 105–131. doi:10.2308/accr-52200.

Chapman, K. L., G. S. Miller, J. J. Neilson, and H. D. White (2021, 03). Investor relations, engagement, and shareholder activism. *Accounting Review*. Advance online publication. doi:10.2308/TAR-2018-0361.

Charnes, A., W. W. Cooper, and E. Rhodes (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research 2*(6), 429–444. doi:10.1016/0377-2217(78)90138-8.

Cheema-Fox, A., B. R. LaPerla, G. Serafeim, and H. S. Wang (2020). Corporate resilience and response during COVID-19. *SSRN Electronic Journal*. doi:10.2139/ssrn.3578167.

Chen, Y., X. Wang, and B. Zhang (2018). An unsupervised deep learning approach for scenario forecasts. In *2018 Power Systems Computation Conference (PSCC)*, pp. 1–7. IEEE. doi:10.23919/PSCC.2018.8442500.

Chen, Y., Y. Wang, D. Kirschen, and B. Zhang (2018). Model-free renewable scenario generation using generative adversarial networks. *IEEE Transactions on Power Systems 33*(3), 3265–3275. doi:10.1109/TPWRS.2018.2794541.

Chollet, F. (2018). *Deep learning with Python*. Shelter Island, NY: Manning Publications Company.

Contessi, S. and P. De Pace (2021). The international spread of COVID-19 stock market collapses. *Finance Research Letters*. Advance online publication. doi:10.1016/j.frl.2020.101894.

Cramer, E., L. R. Gorjão, A. Mitsos, B. Schäfer, D. Witthaut, and M. Dahmen (2022). Validation methods for energy time series scenarios from deep generative models. *IEEE Access 10*, 8194–8207. doi:10.1109/ACCESS.2022.3141875.

Crifo, P., E. Escrig-Olmedo, and N. Mottis (2019). Corporate governance as a key driver of corporate sustainability in France: The role of board members and investor relations. *Journal of Business Ethics 159*(4), 1127–1146. doi:10.1007/s10551-018-3866-6.

de Masson d'Autume, C., S. Mohamed, M. Rosca, and J. Rae (2019). Training language gans from scratch. *Advances in Neural Information Processing Systems 32*. Available at: https://proceedings.neurips.cc/paper/2019/file/a6ea8471c120fe8cc35a2954c9b9c595-Paper.pdf. Last checked on Aug 16, 2022.

de Vilmarest, J. and Y. Goude (2021). State-space models win the IEEE DataPort competition on post-covid day-ahead electricity load forecasting. *arXiv preprint*. arXiv:1411.1784.

Demsetz, H. (1973). Industry structure, market rivalry, and public policy. *Journal of Law and Economics 16*(1), 1–9. doi:10.1086/466752.

DesJardine, M., P. Bansal, and Y. Yang (2019). Bouncing back: Building resilience through social and environmental practices in the context of the 2008 global financial crisis. *Journal of Management 45*(4), 1434–1460. doi:10.1177/0149206317708854.

Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint*. arXiv:1810.04805.

Ding, W., R. Levine, C. Lin, and W. Xie (2021). Corporate immunity to the COVID-19 pandemic. *Journal of Financial Economics 141*(2), 802–830. doi:10.1016/j.jfineco.2021.03.005.

Djankov, S., R. La Porta, F. Lopez-de Silanes, and A. Shleifer (2008). The law and economics of self-dealing. *Journal of Financial Economics 88*(3), 430–465. doi:10.1016/j.jfineco.2007.02.007.

Dubuisson, M.-P. and A. K. Jain (1994). A modified hausdorff distance for object matching. In *Proceedings of 12th international conference on pattern recognition*, Volume 1, pp. 566–568. IEEE. doi:10.1109/ICPR.1994.576361.

Dumas, J., A. Wehenkel, D. Lanaspeze, B. Cornélusse, and A. Sutera (2022). A deep generative model for probabilistic energy forecasting in power systems: normalizing flows. *Applied Energy 305*, 117871. doi:10.1016/j.apenergy.2021.117871.

Einav, L. and J. Levin (2014). Economics in the age of big data. *Science 346*(6210), 1243089. doi:10.1126/science.124308.

Engelhardt, N., M. Krause, D. Neukirchen, and P. Posch (2020a). What drives stocks during the corona-crash? News attention vs. rational expectation. *Sustainability 12*(12), 5014. doi:10.3390/su12125014.

Engelhardt, N., M. Krause, D. Neukirchen, and P. N. Posch (2020b). Trust and stock market volatility during the COVID-19 crisis. *Finance Research Letters 38*, 101873. doi:10.1016/j.frl.2020.101873.

Espinosa-Méndez, C. and J. Arias (2020). COVID-19 effect on herding behaviour in European capital markets. *Finance Research Letters 38*, 101787. doi:10.1016/j.frl.2020.101787.

Esteban, C., S. L. Hyland, and G. Rätsch (2017). Real-valued (medical) time series generation with recurrent conditional gans. *arXiv preprint*. arXiv:1706.02633.

Fahlenbrach, R., K. Rageth, and R. M. Stulz (2020). How valuable is financial flexibility when revenue stops? Evidence from the COVID-19 crisis. *NBER Working Paper Series*. doi:10.3386/w27106.

Fama, E. F. (1990). Stock returns, expected returns, and real activity. *Journal of Finance 45*(4), 1089–1108. doi:10.1111/j.1540-6261.1990.tb02428.x.

Fama, E. F. and J. D. MacBeth (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy 81*(3), 607–636. doi:10.1086/260061.

Farrokhabadi, M., J. Browell, Y. Wang, S. Makonin, W. Su, and H. Zareipour (2022). Day-ahead electricity demand forecasting competition: Post-covid paradigm. *IEEE Open Access Journal of Power and Energy 9*, 185–191. doi:10.1109/OAJPE.2022.3161101.

Fischer, R., N. Piatkowski, C. Pelletier, G. I. Webb, F. Petitjean, and K. Morik (2020). No cloud on the horizon: probabilistic gap filling in satellite image series. In *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*, pp. 546–555. IEEE. doi:10.1109/DSAA49011.2020.00069.

Fortuin, V., D. Baranchuk, G. Rätsch, and S. Mandt (2020). Gp-vae: Deep probabilistic time series imputation. In *International conference on artificial intelligence and statistics*, Volume 108, pp. 1651–1661. PMLR. Available at: https://proceedings.mlr.press/v108/fortuin20a.html. Last checked on Aug 16, 2022.

Frijns, B., D. Margaritis, and M. Psillaki (2012). Firm efficiency and stock returns. *Journal of Productivity Analysis 37*(3), 295–306. doi:10.1007/s11123-011-0246-y.

Fu, R., J. Chen, S. Zeng, Y. Zhuang, and A. Sudjianto (2019). Time series simulation by conditional generative adversarial net. *arXiv preprint arXiv:1904.11419*. arXiv:1904.11419.

Ge, L., W. Liao, S. Wang, B. Bak-Jensen, and J. R. Pillai (2020). Modeling daily load profiles of distribution network for scenario generation using flow-based generative network. *IEEE Access 8*, 77587–77597. doi:10.1109/ACCESS.2020.2989350.

Glossner, S., P. Matos, S. Ramelli, and A. F. Wagner (2020). Where do institutional investors seek shelter when disaster strikes? Evidence from COVID-19. *SSRN Electronic Journal*. doi:10.2139/ssrn.3655271.

Goldman, J. (2020). Government as customer of last resort: The stabilizing effects of government purchases on firms. *Review of Financial Studies 33*(2), 610–643. doi:10.1093/rfs/hhz059.

Goodfellow, I., Y. Bengio, and A. Courville (2016). *Deep learning*. MIT press.

Goodfellow, I., J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio (2014). Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems*, Volume 27. Curran Associates, Inc. Available at: https://proceedings.neurips.cc/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf. Last checked on Aug 16, 2022.

Goodfellow, I. J., J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. C. Courville, and Y. Bengio (2014). Generative adversarial nets. In *Advances in Neural Information Processing Systems 27*, pp. 2672–2680. Available at: https://proceedings.neurips.cc/paper/2014/hash/5ca3e9b122f61f8f06494c97b1afccf3-Abstract.html. Last checked on Aug 16, 2022.

Gretton, A., K. Borgwardt, M. Rasch, B. Schölkopf, and A. Smola (2006). A kernel method for the two-sample-problem. *Advances in neural information processing systems 19*, 513–520. Available at: https://proceedings.neurips.cc/paper/2006/file/e9fb2eda3d9c55a0d89c98d6c54b5b3e-Paper.pdf. Last checked on Aug 16, 2022.

Gretton, A., K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola (2012). A kernel two-sample test. *The Journal of Machine Learning Research 13*(1), 723–773. Available at: https://www.jmlr.org/papers/volume13/gretton12a/gretton12a.pdf?ref=https://githubhelp.com. Last checked on Aug 16, 2022.

Habib, M. A. and A. Ljungqvist (2005). Firm value and managerial incentives: A stochastic frontier approach. *Journal of Business 78*(6), 2053–2094. doi:10.1086/497040.

Hainmueller, J. (2012). Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political Analysis 20*(1), 25–46. doi:10.1093/pan/mpr025.

Hale, T., N. Angrist, R. Goldszmidt, B. Kira, A. Petherick, T. Phillips, S. Webster, E. Cameron-Blake, L. Hallas, S. Majumdar, et al. (2021). A global panel database of pandemic policies (oxford covid-19 government response tracker). *Nature Human Behaviour 5*(4), 529–538. doi:10.1038/s41562-021-01079-8.

Henaff, M., J. Zhao, and Y. LeCun (2017). Prediction under uncertainty with error-encoding networks. *arXiv preprint*. arXiv:1711.04994.

Heusel, M., H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter (2017). Gans trained by a two time-scale update rule converge to a local nash equilibrium. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.), *Advances in Neural Information Processing Systems*, Volume 30. Curran Associates, Inc. Available at: https://proceedings.neurips.cc/paper/2017/file/8a1d694707eb0fefe65871369074926d-Paper.pdf. Last checked on Aug 16, 2022.

Hinton, G., L. Deng, D. Yu, G. E. Dahl, A.-r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, et al. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal processing magazine 29*(6), 82–97. doi:10.1109/MSP.2012.2205597.

Hirshleifer, D. and S. H. Teoh (2003). Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics 36*(1-3), 337–386. doi:10.1016/j.jacceco.2003.10.002.

Hockerts, K. and L. Moir (2004). Communicating corporate responsibility to investors: The changing role of the investor relations function. *Journal of Business Ethics 52*(1), 85–98. doi:10.1023/B:BUSI.0000033109.35980.16.

Holmström, B. and J. Tirole (1993). Market liquidity and performance monitoring. *Journal of Political Economy 101*(4), 678–709. doi:10.1086/261893.

Hope, O.-K., Z. Huang, and R. Moldovan (2021). Wall street analysts as investor relations officers. *Journal of Corporate Finance 67*, 101893. doi:10.1016/j.jcorpfin.2021.101893.

Huttenlocher, D. P., G. A. Klanderman, and W. J. Rucklidge (1993). Comparing images using the hausdorff distance. *IEEE Transactions on pattern analysis and machine intelligence 15*(9), 850–863. doi:10.1109/34.232073.

IR Magazine (2020). Making the most of investor meetings and events in the covid-19 era. Available at: https://www.irmagazine.com/corporate-access/making-most-investor-meetings-and-events-covid-19-era. Last checked on Aug 16, 2022.

Ivanovic, B., K. Leung, E. Schmerling, and M. Pavone (2020). Multimodal deep generative models for trajectory prediction: A conditional variational autoencoder approach. *IEEE Robotics and Automation Letters 6*(2), 295–302. doi:10.1109/LRA.2020.3043163.

Jankensgård, H. and A. Vilhelmsson (2018). The shareholder base hypothesis of stock return volatility: Empirical evidence. *Financial Management 47*(1), 55–79. doi:10.1111/fima.12184.

Jeha, P., M. Bohlke-Schneider, P. Mercado, S. Kapoor, R. S. Nirwan, V. Flunkert, J. Gasthaus, and T. Januschowski (2021). Psa-gan: Progressive self attention gans for synthetic time series. In *International Conference on Learning Representations*. Available at: https://openreview.net/pdf?id=Ix_mh42xq5w. Last checked on Aug 16, 2022.

Kaiser, Ł. and S. Bengio (2018). Discrete autoencoders for sequence models. *arXiv preprint*. arXiv:1801.09797.

Kang, M. and J. Park (2020). Contragan: Contrastive learning for conditional image generation. *Advances in Neural Information Processing Systems 33*, 21357–21369. Available at: https://proceedings.neurips.cc/paper/2020/file/f490c742cd8318b8ee6dca10af2a163f-Paper.pdf. Last checked on Aug 16, 2022.

Karimi, D. and S. E. Salcudean (2019). Reducing the hausdorff distance in medical image segmentation with convolutional neural networks. *IEEE Transactions on medical imaging 39*(2), 499–513. doi:10.1109/TMI.2019.2930068.

Karolyi, G. A., D. Kim, and R. Liao (2020). The theory and practice of investor relations: A global perspective. *Management Science 66*(10), 1–26. doi:10.1287/mnsc.2019.3405.

Karras, T., T. Aila, S. Laine, and J. Lehtinen (2018). Progressive growing of GANs for improved quality, stability, and variation. In *International Conference on Learning Representations*. Available at: https://openreview.net/forum?id=Hk99zCeAb. Last checked on Aug 16, 2022.

Karras, T., S. Laine, and T. Aila (2019). A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4401–4410. doi:10.1109/CVPR.2019.00453.

Keogh, E. J. and M. J. Pazzani (2001). Derivative dynamic time warping. In *Proceedings of the 2001 SIAM international conference on data mining*, pp. 1–11. SIAM. doi:10.1137/1.9781611972719.1.

Khosravi, A., S. Nahavandi, and D. Creighton (2010). Construction of optimal prediction intervals for load forecasting problems. *IEEE Transactions on Power Systems 25*(3), 1496–1503. doi:10.1109/TPWRS.2010.2042309.

Kingma, D. P. and J. Ba (2014). Adam: A method for stochastic optimization. *arXiv preprint*. arXiv:1412.6980.

Kingma, D. P. and M. Welling (2013). Auto-encoding variational bayes. *arXiv preprint*. arXiv:1312.6114.

Kirk, M. P. and J. D. Vincent (2014). Professional investor relations within the firm. *Accounting Review 89*(4), 1421–1452. doi:10.2308/accr-50724.

Koetter, M., A. Karmann, and E. Fiorentino (2006). The cost efficiency of German banks: A comparison of SFA and DEA. *SSRN Electronic Journal*. doi:10.2139/ssrn.947340.

Koochali, A., P. Schichtel, A. Dengel, and S. Ahmed (2019). Probabilistic forecasting of sensory data with generative adversarial networks–forgan. *IEEE Access 7*, 63868–63880. doi:10.1109/ACCESS.2019.2915544.

Krause, M. and P. N. Posch (2022). Deep Generative Models for Probabilistic Time Series Forecasting. *Working Paper*.

Krizhevsky, A., I. Sutskever, and G. E. Hinton (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems 25*. Available at: https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf. Last checked on Aug 16, 2022.

Kusner, M. J., B. Paige, and J. M. Hernández-Lobato (2017, 06–11 Aug). Grammar variational autoencoder. In *Proceedings of the 34th International Conference on Machine Learning*, Volume 70 of *Proceedings of Machine Learning Research*, pp. 1945–1954. PMLR. Available at: http://proceedings.mlr.press/v70/kusner17a/kusner17a.pdf. Last checked on Aug 16, 2022.

La Porta, R., F. Lopez-de Silanes, and A. Shleifer (2006). What works in securities laws? *Journal of Finance 61*(1), 1–32. doi:10.1111/j.1540-6261.2006.00828.x.

Landier, A. and D. Thesmar (2020). Earnings expectations during the COVID-19 crisis. *Review of Asset Pricing Studies 10*(4), 598–617. doi:10.1093/rapstu/raaa016.

Lee, L. F., A. P. Hutton, and S. Shu (2015). The role of social media in the capital market: Evidence from consumer product recalls. *Journal of Accounting Research 53*(2), 367–404. doi:10.1111/1475-679X.12074.

Lemmon, M. L. and K. V. Lins (2003). Ownership structure, corporate governance, and firm value: Evidence from the East Asian financial crisis. *Journal of Finance 58*(4), 1445–1468. doi:10.1111/1540-6261.00573.

Li, D., Q. N. Nguyen, P. K. Pham, and S. X. Wei (2011). Large foreign ownership and firm-level stock return volatility in emerging markets. *Journal of Financial and Quantitative Analysis 46*(4), 1127–1155. doi:i:10.1017/S0022109011000202.

Li, J., W. Monroe, T. Shi, S. Jean, A. Ritter, and D. Jurafsky (2017). Adversarial learning for neural dialogue generation. *arXiv preprint*. arXiv:1701.06547.

Lins, K. V., H. Servaes, and A. Tamayo (2017). Social capital, trust, and firm performance: The value of corporate social responsibility during the financial crisis. *Journal of Finance 72*(4), 1785–1824. doi:10.1111/jofi.12505.

Lins, K. V., P. Volpin, and H. F. Wagner (2013). Does family control matter? International evidence from the 2008–2009 financial crisis. *Review of Financial Studies 26*(10), 2583–2619. doi:10.1093/rfs/hht044.

Madeka, D., L. Swiniarski, D. Foster, L. Razoumov, K. Torkkola, and R. Wen (2018). Sample path generation for probabilistic demand forecasting. In *ICML workshop on Theoretical Foundations and Applications of Deep Generative Models*. Available at: https://assets.amazon.science/fa/d7/d1049a9144219b2686e51394c0f4/sample-path-generation-for-probabilistic-demand-forecasting.pdf. Last checked on Aug 16, 2022.

Meeusen, W. and J. van Den Broeck (1977). Efficiency estimation from Cobb-Douglas production functions with composed error. *International Economic Review 18*(2), 435–444. doi:10.2307/2525757.

Merton, R. C. (1987). A simple model of capital market equilibrium with incomplete information. *Journal of Finance 42*(3), 483–510. doi:10.1111/j.1540-6261.1987.tb04565.x.

Metz, L., J. Ibarz, N. Jaitly, and J. Davidson (2017). Discrete sequential prediction of continuous actions for deep rl. *arXiv preprint*. arXiv:1705.05035.

Mirza, M. and S. Osindero (2014). Conditional generative adversarial nets. *arXiv preprint*. arXiv:1411.1784.

Mitton, T. (2002). A cross-firm analysis of the impact of corporate governance on the East Asian financial crisis. *Journal of Financial Economics 64*(2), 215–241. doi:10.1016/S0304-405X(02)00076-4.

Neukirchen, D., N. Engelhardt, M. Krause, and P. N. Posch (2021a). Firm efficiency and stock returns during the COVID-19 crisis. *Finance Research Letters*. Advance online publication. doi:10.1016/j.frl.2021.102037.

Neukirchen, D., N. Engelhardt, M. Krause, and P. N. Posch (2021b). The value of (private) investor relations during the COVID-19 crisis. *Working Paper*.

Nguyen, G. X. and P. E. Swanson (2009). Firm characteristics, relative efficiency, and equity returns. *Journal of Financial and Quantitative Analysis 44*(1), 213–236. doi:10.1017/S0022109009090012.

Nguyen, T., H. Nguyen, and X. Yin (2015). Corporate governance and corporate financing and investment during the 2007-2008 financial crisis. *Financial Management 44*(1), 115–146. doi:10.1111/fima.12071.

NIRI (2020). Definition of investor relations. Available at: https://www.niri.org/about-niri. Last checked on Aug 16, 2022.

Pagano, M., C. Wagner, and J. Zechner (2020). Disaster resilience and asset prices. *SSRN Electronic Journal*. doi:10.2139/ssrn.3603666.

Papke, L. E. and J. M. Wooldridge (1996). Econometric methods for fractional response variables with an application to 401 (k) plan participation rates. *Journal of Applied Econometrics 11*(6), 619–632. doi:10.1002/(SICI)1099-1255(199611)11:6<619::AID-JAE418>3.0.CO;2-1.

Paraschiv, F., D. Erni, and R. Pietsch (2014). The impact of renewable energies on eex day-ahead electricity prices. *Energy Policy 73*, 196–210. doi:10.1016/j.enpol.2014.05.004.

Peltzman, S. (1977). The gains and losses from industrial concentration. *Journal of Law and Economics 20*(2), 229–263. doi:10.1086/466902.

Peng, L. and W. Xiong (2006). Investor attention, overconfidence and category learning. *Journal of Financial Economics 80*(3), 563–602. doi:10.1016/j.jfineco.2005.05.003.

Petersen, M. A. (2009). Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies 22*(1), 435–480. doi:10.1093/rfs/hhn053.

Piatkowski, N. (2019). Hyper-parameter-free generative modelling with deep boltzmann trees. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 415–431. Springer. doi:10.1007/978-3-030-46147-8_25.

Piatkowski, N., S. Lee, and K. Morik (2013). Spatio-temporal random fields: compressible representation and distributed estimation. *Machine learning 93*(1), 115–139. doi:10.1007/s10994-013-5399-7.

Piatkowski, N., P. N. Posch, and M. Krause (2021a). How to trust generative probabilistic models for time-series data? In *International Conference on Learning and Intelligent Optimization*, pp. 283–298. Springer. doi:10.1007/978-3-030-92121-7_234.

Piatkowski, N., P. N. Posch, and M. Krause (2021b). How to trust generative probabilistic models for time-series data? In *International Conference on Learning and Intelligent Optimization*, pp. 283–298. Springer. doi:10.1007/978-3-030-92121-7_23.

Ramelli, S. and A. F. Wagner (2020, 07). Feverish stock price reactions to COVID-19. *Review of Corporate Finance Studies 9*(3), 622–655. doi:10.1093/rcfs/cfaa012.

Ravuri, S., K. Lenc, M. Willson, D. Kangin, R. Lam, P. Mirowski, M. Fitzsimons, M. Athanassiadou, S. Kashem, S. Madge, et al. (2021). Skilful precipitation nowcasting using deep generative models of radar. *Nature 597*(7878), 672–677. doi:10.1038/s41586-021-03854-z.

Rezende, D. J., S. Mohamed, and D. Wierstra (2014). Stochastic backpropagation and approximate inference in deep generative models. In *Proceedings of the 31st International Conference on Machine Learning*, Volume 32, pp. 1278–1286. PMLR. Available at: http://proceedings.mlr.press/v32/rezende14.pdf. Last checked on Aug 16, 2022.

Salazar, A. A., J. Zheng, Y. Che, and F. Xiao (2022). Deep generative model for probabilistic wind speed and wind power estimation at a wind farm. *Energy Science & Engineering*. doi:10.1002/ese3.1086.

Sankoff, D. and J. B. Kruskal (1983). Time warps, string edits, and macromolecules: The theory and practice of sequence comparison. *Reading: Addison-Wesley Publication*.

Available at: https://ci.nii.ac.jp/naid/10007989594/en/. Last checked on Aug 16, 2022.

Schmidt, D. (2020). Stock market rumors and credibility. *Review of Financial Studies 33*(8), 3804–3853. doi:10.1093/rfs/hhz120.

Sohn, K., H. Lee, and X. Yan (2015). Learning structured output representation using deep conditional generative models. *Advances in neural information processing systems 28*. Available at: https://proceedings.neurips.cc/paper/2015/file/8d55a249e6baa5c06772297520da2051-Paper.pdf. Last checked on Aug 16, 2022.

Sriperumbudur, B. K., A. Gretton, K. Fukumizu, B. Schölkopf, and G. R. Lanckriet (2010). Hilbert space embeddings and metrics on probability measures. *The Journal of Machine Learning Research 11*, 1517–1561. Available at: https://www.jmlr.org/papers/volume11/sriperumbudur10a/sriperumbudur10a.pdf. Last checked on Aug 16, 2022.

Subrahmanyam, A. and S. Titman (2001). Feedback from stock prices to cash flows. *Journal of Finance 56*(6), 2389–2413. doi:10.1111/0022-1082.00409.

Sutherland, D. J., H.-Y. Tung, H. Strathmann, S. De, A. Ramdas, A. Smola, and A. Gretton (2016). Generative models and model criticism via optimized maximum mean discrepancy. *arXiv preprint*. arXiv:1611.04488.

Takahashi, S., Y. Chen, and K. Tanaka-Ishii (2019). Modeling financial time-series with generative adversarial networks. *Physica A: Statistical Mechanics and its Applications 527*, 121261. doi:T10.1016/j.physa.2019.121261.

Tiffin, M. A. J. (2019). *Machine learning and causality: The impact of financial crises on growth*. International Monetary Fund.

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin (2017). Attention is all you need. *Advances in neural information processing systems 30*. Available at: https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf. Last checked on Aug 16, 2022.

Vuolteenaho, T. (2002). What drives firm-level stock returns? *Journal of Finance 57*(1), 233–264. doi:10.1111/1540-6261.00421.

Wan, Z., Y. Zhang, and H. He (2017). Variational autoencoder based synthetic data generation for imbalanced learning. In *2017 IEEE symposium series on computational intelligence (SSCI)*, pp. 1–7. IEEE. doi:10.1109/SSCI.2017.8285168.

Wiese, M., R. Knobloch, R. Korn, and P. Kretschmer (2020). Quant gans: Deep generation of financial time series. *Quantitative Finance 20*(9), 1419–1440. doi:10.1080/14697688.2020.1730426.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data.* MIT press.

Wu, Y., J. Ni, W. Cheng, B. Zong, D. Song, Z. Chen, Y. Liu, X. Zhang, H. Chen, and S. B. Davidson (2021). Dynamic gaussian mixture based deep generative model for robust forecasting on sparse multivariate time series. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Volume 35, pp. 651–659. Available at: https://ojs.aaai.org/index.php/AAAI/article/view/16145. Last checked on Aug 16, 2022.

Xu, H., W. Chen, N. Zhao, Z. Li, J. Bu, Z. Li, Y. Liu, Y. Zhao, D. Pei, Y. Feng, J. Chen, Z. Wang, and H. Qiao (2018). Unsupervised anomaly detection via variational auto-encoder for seasonal kpis in web applications. In *Proceedings of the 2018 World Wide Web Conference*, WWW '18, Republic and Canton of Geneva, CHE, pp. 187–196. International World Wide Web Conferences Steering Committee. doi:10.1145/3178876.3185996.

Yoon, J., D. Jarrett, and M. van der Schaar (2019). Time-series generative adversarial networks. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (Eds.), *Advances in Neural Information Processing Systems*, Volume 32. Curran Associates, Inc. Available at: https://proceedings.neurips.cc/paper/2019/file/c9efe5f26cd17ba6216bbe2a7d26d490-Paper.pdf. Last checked on Aug 16, 2022.

Zaremba, A., R. Kizys, D. Y. Aharon, and E. Demir (2020). Infected markets: Novel coronavirus, government interventions, and stock return volatility around the globe. *Finance Research Letters 35*, 101597. doi:10.1016/j.frl.2020.101597.

Zhang, C., S. R. Kuppannagari, R. Kannan, and V. K. Prasanna (2018). Generative adversarial network for synthetic time series data generation in smart grids. In *2018 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, pp. 1–6. IEEE. doi:10.1109/SmartGridComm.2018.8587464.

Zhang, D., M. Hu, and Q. Ji (2020). Financial markets under the global pandemic of COVID-19. *Finance Research Letters 36*, 101528. doi:10.1016/j.frl.2020.101528.

Ziel, F. (2021). Smoothed bernstein online aggregation for day-ahead electricity demand forecasting. *arXiv preprint*. arXiv:2107.06268.