Ca' Foscari University of Venice

Department of Economics

Working Paper

Oleksandr Castello
Marco Corazza

Machine Learning techniques for synthetic data generation in Energy and Financial Markets

ISSN: 1827-3580
No. 11/WP/2026

Working Papers
Department of Economics
Ca' Foscari University of Venice
No. 11/WP/2026
ISSN 1827-3580

Machine Learning techniques for synthetic data generation in Energy and Financial Markets

Oleksandr Castello
Ca' Foscari University of Venice

Marco Corazza
Ca' Foscari University of Venice

Abstract
The availability of sufficiently large, reliable, and high-quality datasets represents a fundamental prerequisite for quantitative analysis and data-driven decision-making in economics and finance. In practice, however, financial data are often limited, noisy, or subject to restricted access, creating significant empirical constraints for both researchers and practitioners. Recent advances in Generative Machine Learning (GenML) provide promising tools to overcome these limitations by enabling the generation of synthetic data capable of preserving the main statistical features of original data. Despite the rapid diffusion of these techniques, most existing studies focus on replicating stylized facts of financial time series or producing forward-looking simulations, while less attention has been devoted to a systematic assessment of the generative fidelity and generalization capacity of alternative models across different distributional environments. Motivated by this gap, this study provides a comparative evaluation of several Deep Generative Machine Learning (Deep-GenML) families by assessing their ability to reproduce both theoretical statistical distributions and empirical financial and commodity market data. The analysis spans multiple Deep-GenML architectures, distributional settings and market regimes, while also examining model performance under alternative training configurations that reflect varying degrees of data availability. The empirical evidence indicates that deep generative models are capable of accurately reproducing complex distributional features—including heavy tails, asymmetry, and multimodality—across a wide range of scenarios. Overall, the results highlight the potential of deep generative approaches as flexible tools for synthetic data generation and distributional modeling in financial and energy market applications.

Keywords
Deep Generative Machine Learning, Synthetic data generation, GAN, VAE, EBM, Financial and Energy market data

JEL Codes
C45, C46, C58, C63

Address for correspondence:
Oleksandr Castello
Department of Economics
Ca' Foscari University of Venice
Cannaregio 873, Fondamenta S.Giobbe
30121 Venezia - Italy
e-mail: oleksandr.castello@unive.it

The Working Paper Series
is available only on line
(https://www.unive.it/dec-workingpapers)
For editorial correspondence, please contact:
wp.dse@unive.it

Department of Economics
Ca' Foscari University of Venice
Cannaregio 873, Fondamenta San Giobbe
30121 Venice Italy
Fax: ++39 041 2349210

Machine Learning techniques for synthetic data generation in Energy and Financial Markets

Oleksandr Castello*, Marco Corazza

Department of Economics, Cà Foscari University of Venice, Sestiere Cannaregio 873, Venice, 30121, Italy

Abstract

The availability of sufficiently large, reliable, and high-quality datasets represents a fundamental prerequisite for quantitative analysis and data-driven decision-making in economics and finance. In practice, however, financial data are often limited, noisy, or subject to restricted access, creating significant empirical constraints for both researchers and practitioners. Recent advances in Generative Machine Learning (GenML) provide promising tools to overcome these limitations by enabling the generation of synthetic data capable of preserving the main statistical features of original data. Despite the rapid diffusion of these techniques, most existing studies focus on replicating stylized facts of financial time series or producing forward-looking simulations, while less attention has been devoted to a systematic assessment of the generative fidelity and generalization capacity of alternative models across different distributional environments. Motivated by this gap, this study provides a comparative evaluation of several Deep Generative Machine Learning (Deep-GenML) families by assessing their ability to reproduce both theoretical statistical distributions and empirical financial and commodity market data. The analysis spans multiple Deep-GenML architectures, distributional settings and market regimes, while also examining model performance under alternative training configurations that reflect varying degrees of data availability. The empirical evidence indicates that deep generative models are capable of accurately reproducing complex distributional features—including heavy tails, asymmetry, and multimodality—across a wide range of scenar-

*Corresponding Author.

Email addresses: oleksandr.castello@unive.it (Oleksandr Castello), corazza@unive.it (Marco Corazza)

ios. Overall, the results highlight the potential of deep generative approaches as flexible tools for synthetic data generation and distributional modeling in financial and energy market applications.

Keywords:
Deep Generative Machine Learning, Synthetic data generation, GAN, VAE, EBM, Financial and Energy market data

JEL: C45, C46, C58, C63

---

1. Introduction

In the economic and financial domain, the availability of sufficiently large and high-quality datasets represents a cornerstone for quantitative analysis, data-driven decision-making, and empirical research. Tasks such as market modeling, scenario generation, stress testing, risk management, trading strategy design, and asset pricing rely fundamentally on the statistical richness and representativeness of the underlying data. However, real-world financial data frequently suffer from limitations related to scarcity, imperfect quality, or restricted accessibility due to confidentiality constraints, regulatory barriers, acquisition costs, missing observations, measurement errors, or technological frictions in the data collection process [3, 20]. These structural limitations impose significant empirical constraints on practitioners and researchers, particularly when modeling rare events, conducting regime-dependent analyses, or testing methodological robustness across heterogeneous distributional environments.

In recent years, Generative Machine Learning (GenML) has emerged as a powerful and increasingly adopted framework for data augmentation and synthetic data generation, offering the possibility of producing realistic datasets that preserve the salient distributional, structural, and statistical properties of original data [15]. These techniques have achieved remarkable results across a wide range of fields, from healthcare [2, 13] and climate modeling [16, 22] to engineering [17, 6] and data science [8, 19]. The economic and financial sectors are no exception [12, 11], as this methodological advancement constitutes a viable solution to data scarcity, while simultaneously enhancing empirical analysis and expanding the set of quantitative tools available to both practitioners and researchers.

Despite this growing interest, the overwhelming majority of financial applications of GenML have focused primarily on two directions. First, the

replication of dynamic stylized facts, such as volatility clustering and autocorrelation patterns. Second, the implementation of forward-looking simulations aimed at scenario generation [24] or synthetic asset-path construction [7], predominantly within equity markets. This narrow focus leaves a critical gap in the literature: a systematic and rigorous evaluation of the generative power of GenML families with respect to their ability to replicate the distributional properties of both theoretical probability densities — commonly assumed in quantitative modeling — and real-world datasets drawn from multiple markets operating under diverse regimes, including periods characterized by structural breaks, crises, and exogenous shocks.

Furthermore, existing literature lacks a global and exhaustive comparative framework, failing to integrate diverse model families, various model configurations, and a broad spectrum of data inputs and distributions.

Motivated by these considerations, the present study develops a comprehensive, multi-layered assessment of the generative capabilities of several families of Deep Generative Machine Learning (Deep-GenML) models. The objective is twofold: (i) to rigorously evaluate their effectiveness in learning and reproducing a wide array of statistical distributions, both theoretical and market-based; and (ii) to construct an integrated and comparative benchmark that spans multiple model classes, data-generating processes, and performance metrics.

Specifically, we benchmark the performance of representative models from three core GenML paradigms, namely Adversarial [4], Latent-Encoding [9], and Energy-Based [1] frameworks, including: the Multilayer Perceptron Generative Adversarial Network (MLP-GAN), Deep Convolutional GAN (DC-GAN), β-Variational Autoencoder (β-VAE), Adversarial-VAE (A-VAE), Stochastic Gradient Hamiltonian Monte Carlo Energy-Based Model (SGHMC-EBM), and Score-Based EBM (S-EBM). The models were tasked with learning and replicating the statistical properties of several theoretical distributions frequently employed in quantitative finance — Standard Normal, Student's t, Weibull, Johnson's, and Gaussian Mixture. In addition, we evaluated models' performance on a diverse set of real-world daily financial and commodity market data, including the STOXX Europe 600 Climate Transition Benchmark (CTB) index returns, Russian and Chinese government bond spot rates, and Natural Gas and Coal futures prices. These datasets exhibit heterogeneous lengths (depending on the availability of historical information) and incorporate multiple market regimes, including periods of structural turbulence, geopolitical shocks, supply-chain disruptions, and crisis episodes, thereby

enabling a robust stress-testing environment for model evaluation.

To further examine generalization, the networks were trained on input sequences of varying length, mimicking typical historical trading windows used by practitioners.

By jointly exposing the models to analytically tractable distributions and complex, noisy, and potentially non-stationary empirical data, the proposed framework enables a unified and robust evaluation of generative performance. Moreover, this setting allows us to distinguish models that excel at replicating idealized statistical structures from those capable of generalizing to realistic financial environments, thereby providing a meaningful assessment of their suitability for practical applications in quantitative finance. To the best of the authors' knowledge, this study is the first to provide a comprehensive, multi-family benchmark for generative models applied to various theoretical and real financial densities.

Overall, our study contributes to the existing literature along several dimensions, summarized as follows: (i) we offer an integrated, cross-family and cross-distribution benchmark of Deep-GenML models; (ii) we address the issue of model generalization across a heterogeneous set of theoretical and real-world distributional environments; (iii) we investigate models' generative power under alternative training regimes; (iv) finally, we provide a comprehensive stress test of the framework using datasets characterized by periods of extreme market turmoil, including global shocks and recent financial, geopolitical, and energy crises.

The remainder of the paper is organized as follows. Section 2 discusses the Deep-GenML models in use; Section 3 presents the empirical framework and discusses the results; Section 4 concludes.

2. Deep Generative Machine Learning Models

Generative Machine Learning (GenML) refers to a broad class of models designed to capture data-generating mechanisms and salient statistical structure of complex datasets, with the objective of producing novel, high-fidelity synthetic samples. Over recent years, the field has evolved into a diverse set of methodological frameworks, which can be broadly grouped into four major categories based on their fundamental approach to density estimation and sample generation. These include: Adversarial models [5], Inference-based or Latent-Encoding models [9], Energy-based models [1], Autoregressive [10, 23] and Diffusion models [21].

In this section, we present six representative GenML architectures covering three prominent paradigms: Adversarial, Inference-based, and Energy-based. These models have been selected for their relevance and diverse approaches to data generation and include the Multilayer Perceptron Generative Adversarial Network (MLP-GAN), Deep Convolutional-GAN (DC-GAN), β-Variational Autoencoder (β-VAE), Adversarial-VAE (A-VAE), Stochastic Gradient Hamiltonian Monte Carlo Energy-based model (SGHMC-EBM), as well as the Score-based EBM (S-EBM). The key features of each model are briefly discussed below. For comprehensive methodological details and formal derivations, readers can refer to the original contributions.

2.1. Generative Adversarial Networks

Adversarial models are grounded in a competitive learning paradigm involving two neural networks - a Generator and a Discriminator - trained simultaneously in a zero-sum game. The Generator processes input noise through a parametric transformation to synthesize samples that are statistically indistinguishable from the target data distribution, while the Discriminator operates as a binary classifier tasked with separating real from generated observations. Through this minimax interaction, the Generator learns an implicit representation of the underlying data-generating process relying on observational signals rather than functional estimation.

Within this class, we considered two representative architectures: the Multilayer Perceptron GAN (MLP-GAN) and the Deep Convolutional GAN (DC-GAN). The MLP-GAN employs fully connected layers and provides a flexible baseline architecture capable of capturing complex global non-linear relationships in low-dimensional settings. By contrast, the DC-GAN leverages convolutional layers to exploit local structures and hierarchical feature representations, making it particularly suitable for data exhibiting grid-like or ordered dependencies.

GANs bypass the need for explicit likelihood specifications or parametric density approximations. This likelihood-free formulation facilitates a flexible representation of complex, multimodal, and non-smooth data-generating distributions, which typically pose significant challenges for explicit density-based models.

2.2. Variational Autoencoders

The VAE framework combines neural networks with variational Bayesian inference, enabling explicit probabilistic modelling of the data-generating

process through latent variable inference. It is built upon an Encoder–Decoder architecture, where the Encoder maps the input data into a compressed latent space defined by a prior distribution (typically Gaussian), and the Decoder reconstructs the target distribution by sampling from these latent representations. Model optimization is formulated as the maximization of the Evidence Lower Bound (ELBO) on the data likelihood. This framework ensures a structured and continuous latent manifold, facilitating stable and regularized generative processes.

Within this paradigm, we focused on the β-Variational Autoencoder (β-VAE) and the Adversarial Variational Autoencoder (A-VAE) architectures. The β-VAE introduces an explicit regularization of the latent space by adjusting the magnitude of the Kullback–Leibler penalty within the objective function, thereby establishing a formal trade-off between latent disentanglement and reconstruction fidelity. By contrast, the A-VAE augments the variational framework with an adversarial training component, relaxing the parametric assumptions on the latent distribution while preserving the probabilistic rigor of the likelihood-based objective.

By explicitly modelling the data-generating process through latent variables and likelihood-based optimization, VAE-type models offer stable training dynamics and a well-defined probabilistic interpretation, ensuring consistent distributional replication.

2.3. Energy-based Networks

Energy-based models characterize the data distribution through a scalar energy function, where lower energy values correspond to higher-probability regions. Unlike the adversarial competition of GANs or the latent-to-data mapping of VAEs, EBMs parameterize the density landscape directly. A single network learns to assign an energy value to each point in the input space, shaping an energy landscape such that low-energy regions correspond to high-density regions of the target data distribution.

In our research, we considered two representative variants: the Stochastic Gradient Hamiltonian Monte Carlo EBM (SGHMC-EBM) and the Score-based EBM (S-EBM). The SGHMC-EBM employs Hamiltonian dynamics combined with stochastic gradients to explore the energy landscape and generate samples from low-energy regions. The S-EBM, by contrast, learns the score function - the gradient of the log-density with respect to the input - enabling sample generation through iterative denoising and Langevin-type dynamics.

By explicitly shaping the geometry of the data distribution, EBMs provide a theoretically grounded alternative to adversarial and variational approaches, capable of capturing complex, multimodal, and non-smooth distributional features that are challenging for traditional parametric or likelihood-based generative models.

3. Empirical Study

This section examines the empirical performance of the considered DeepGenML models in replicating salient statistical and distributional features of the reference data. The evaluation framework combines controlled distributional benchmarks with heterogeneous real-world financial series, enabling a systematic assessment of model fidelity, stability, and generalization across markedly different data-generating processes.

3.1. Data

The empirical investigation relies on a heterogeneous collection of input datasets chosen to span a wide range of distributional structures and market conditions. This design supports a rigorous assessment of model performance under both controlled statistical environments and realistic, market-driven settings.

To this end, we considered a set of theoretical probability distributions commonly used in quantitative finance and risk modeling, namely the Standard Normal, Student's t, Weibull, Johnson's, and Gaussian Mixture distributions, illustrated in Figure 1. Each distribution is represented by 10,000 simulated values, ensuring a sufficiently large sample size to accurately characterize its underlying statistical properties. The specific parameters are carefully selected such that the resulting distributions provide a well-defined statistical ground truth against which model performance can be benchmarked.

As shown in Figure 1, these PDFs exhibit markedly different characteristics in terms of symmetry, tail behavior, skewness, kurtosis, and modality. While the Standard Normal distribution represents a canonical baseline with finite moments and light tails, the Student's t and Gaussian Mixture introduce heavy tails and multimodality, whereas the Weibull and Johnson families allow for flexible asymmetry and tail decay. By exposing the generative models to this controlled environment, we are able to directly assess their

capacity to replicate higher-order moments and complex marginal structures beyond simple mean–variance dynamics.

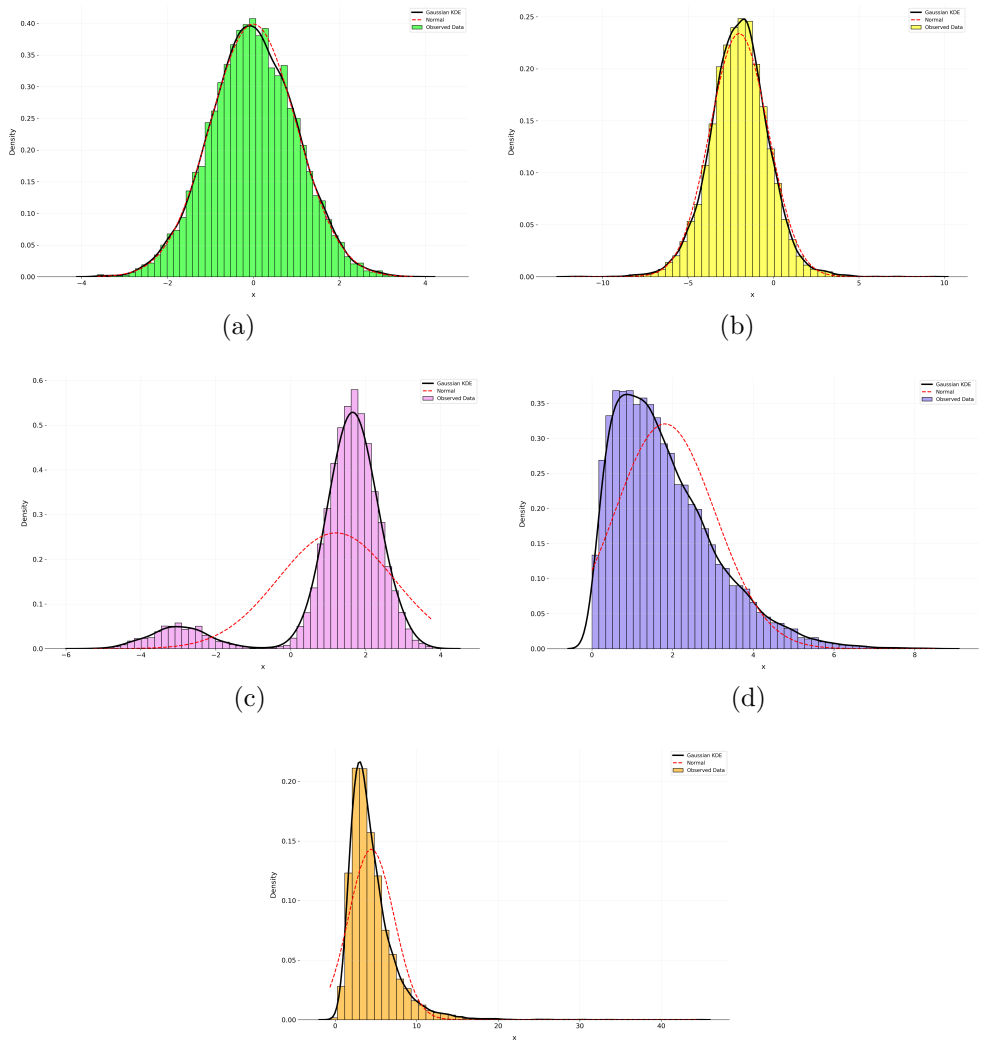

Figure 1: Plot of the Standard Normal (a), Student's t (b), Gaussian Mixture (c), Weibull (d) and Johnson's (e) theoretical densities, along with the Gaussian Kernel Density Estimate (black) and Gaussian (red) reference densities.

Additionally, the analysis incorporates a diverse set of real-world daily series, including STOXX Europe 600 Climate Transition Benchmark (EU 600 CTB) index returns, Russian and Chinese 10-year government bond spot rates, as well as Natural Gas and Coal prices on 1-year futures contracts. The observation periods differ across the examined markets. Specifically, the start/end dates are 04-2019/11-2024 for the EU 600 CTB, 01-2003/10-2022 for Russian spot rates, 01-2005/10-2022 for Chinese spot rates, and 03-2010/01-2024 and 01-2016/01-2024 for Natural Gas and Coal futures prices, respectively. Overall, the datasets span between a minimum of 3 and a maximum of 19 years of daily data and include 1449 observations for the EU 600 CTB, 5009 for Russian spot rates, 4245 for Chinese spot rates, and 3545 and 2089 for Natural Gas and Coal futures prices, respectively.

Figure 2 illustrates the empirical distributions of the analyzed assets. These datasets reflect a broad spectrum of market regimes, encompassing calm periods as well as episodes of heightened uncertainty associated with structural turbulence, geopolitical shocks, supply-chain disruptions, and global crises. As a result, their PDFs exhibit pronounced departures from normality, heavy tails, and irregular features that pose substantial challenges for generative modeling. From a modeling perspective, these data introduce multiple sources of complexity that are absent in the theoretical setting, such as noise, non-stationarity, and latent structural breaks. Consequently, they provide a stringent testing ground for evaluating the robustness and generalization capacity of the generative architectures.

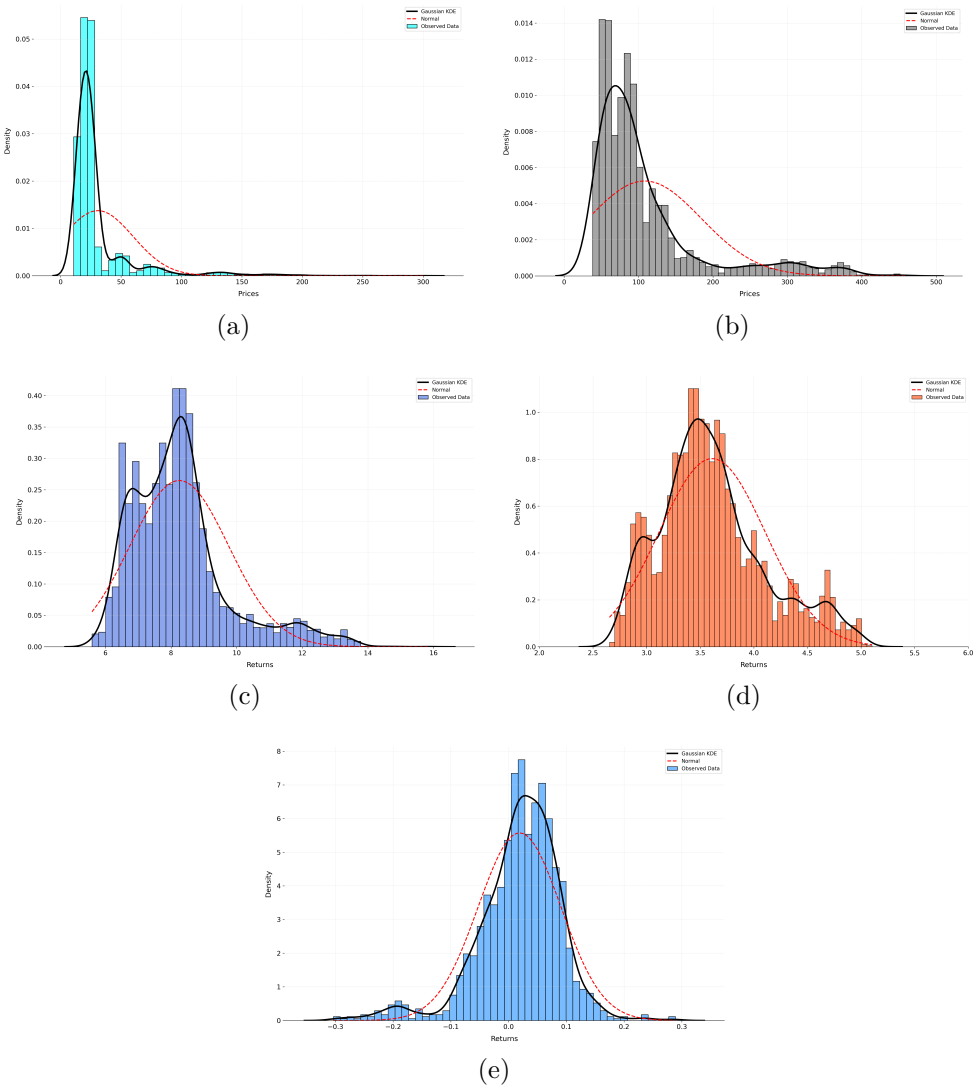

Figure 2: Plot of the Natural Gas futures prices (a), Coal futures prices (b), Russian spot rates (c), Chinese spot rates (d) and STOXX EU 600 CTB index returns (e) observed densities, along with the Gaussian Kernel Density Estimate (black) and Gaussian (red) reference densities.

3.2. Experiment setup

We evaluate the ability of the Deep-GenML models introduced in Section 2 to generate synthetic samples that provide a high-fidelity representation of the target data-generating processes.

As part of the network input preparation process, we first preprocessed real-world empirical time series to ensure numerical stability and comparability across architectures. Observations were rescaled using robust location and scale measures, producing centered and normalised inputs while preserving their distributional shape. Theoretical distributions, by contrast, were directly considered as model inputs, being analytically defined and free from scale-related issues.

Following the preprocessing stage, both theoretical and empirical datasets constituted the pool of admissible inputs to the generative models. Within this framework, three distinct training regimes were specified in order to assess model performance under limited information settings. In particular, for each model–dataset pair, the generative architecture was trained separately using input samples of 21, 63, and 126 observations randomly drawn from the reference series. This setup mimics typical historical look-back windows used by practitioners (i.e. month, quarter, and half a year) and allows for a systematic evaluation of generative performance under varying levels of data availability. Finally, the trained model was employed to generate 10,000 synthetic observations, yielding an associated synthetic distribution.

Additionally, given the inherently stochastic nature of GenML models, we executed 50 repeated simulation cycles for each model-dataset-training configuration, and for each case an averaged synthetic distribution was obtained by computing the Wasserstein Barycenter (WB) of the corresponding PDFs. This procedure mitigates the impact of stochastic variability and ensures greater robustness and stability of the empirical results.

Finally, the framework was implemented in Python (3.11.14) using Pandas (2.3.3) [14] for data preprocessing and PyTorch (2.9.1) [18] for model development.

3.3. Evaluation metrics

Evaluation proceeds along multiple complementary dimensions. To quantify the fidelity with which the generative models reproduce distributional structures, we compute a set of statistical performance metrics widely adopted in the literature and sensitive to different aspects of distributional divergence. These include the Wasserstein Distance, Energy Distance, Maximum

Mean Discrepancy (MMD), Jensen-Shannon Divergence (JS), as well as the Overlap Coefficient (OC). Together, these metrics offer a rigorous, multi-perspective assessment of generative accuracy, highlighting strengths and weaknesses across different statistical environments and model families.

3.4. Results discussion

Models' generative performance is reported in Tables A.1–A.10 (see Appendix A) for each distribution considered. Within each table, for every model–input configuration pair, we report the average results for all accuracy metrics introduced in Subsection 3.3 computed over 50 simulation runs, with the overall best-performing specification highlighted in bold.

The results can be interpreted from at least three complementary perspectives. On the one hand, they allow a comparison across models to determine the most effective generative approach within each distribution class. At the same time, they provide evidence on how the interaction between model architecture and input configuration impacts the replication of the target distributions, while also identifying the best-performing specification within each setting.

Overall, results indicate that all the implemented techniques exhibit consistently strong generative performance. Exemplifying this with the Overlap Coefficient, the models attain average values exceeding 94% in the theoretical density setting and 87% for market-based distributions, confirming the flexibility, reliability as well as generalization capacity of the analyzed framework.

Turning to the comparative analysis across model families, VAEs and EBMs emerge as the best-performing families, consistently outperforming GANs, as reflected by higher OC values. On average, these architectures attain OC levels exceeding 96% and 94% on theoretical distributions, and 90% and 89% on observed real-world distributions, respectively. EBM-based methods demonstrate consistently strong and stable performance across all distributional settings, particularly in the presence of multimodality or pronounced asymmetry, likely due to their ability to learn the target distribution without relying on an explicit parametric specification. VAE-based architectures, in turn, exhibit improved and more consistent performance across both theoretical and empirical distributions, likely due to their structured latent representation, which stabilizes learning and supports robust, consistent generative modeling. Within these classes, A-VAE and SGHMC-EBM represent the top-performing specifications in the theoretical density setting, attaining

average OC values of 96.4% and 95.9%, respectively, followed by β-VAE and S-EBM with corresponding values of 95.8% and 92.9%. When focusing on empirical distributions, the strongest results are instead obtained by A-VAE and S-EBM, reaching average OC values of 90.6% and 89.3%, respectively, followed by β-VAE and SGHMC-EBM with OC values of 89.1% and 88.8%. With regard to GAN-based models, performance remains competitive, albeit slightly lower and more variable compared with VAE- and EBM-based architectures, with average OC values marginally above 93% under idealized conditions and around 84% when applied to empirical series. GANs tend to deliver stronger performance when dealing with relatively symmetric distributions. Within this family, the MLP-GAN emerges as the leading specification, attaining approximately 94.1% in OC on theoretical datasets and 84.8% on real-world series, while the DC-GAN ranks lowest, with OC values of 92.9% and 82.5% for the corresponding settings.

With regard to model generalization, a modest decline in accuracy metrics is observed when transitioning from idealized distributions to real-world series, with the decrease in OC of 5.6% for EBM-based models, 6.5% for VAEs and of 10.5% for GAN-based architectures. It is important to note, however, that the models are stress-tested under substantially more complex observed distributions of financial and energy assets, characterized by multiple market regimes and shocks. These results allow us to affirm that the generative capabilities of the models are robustly preserved, highlighting their flexibility and the presence of a coherent, stable generative mechanism that remains effective across structurally heterogeneous distributions.

It is worth noting that the models achieve strong generative performance even under the most limited input configuration. Notably, VAE-based architectures remain the most accurate in this setting, attaining average OC levels of approximately 95% for theoretical densities and 89% for empirical ones. EBM-based models follow closely, with corresponding values of 92% and 87%, while GAN-based approaches display comparatively lower, though still robust, performance, reaching approximately 92% and 81% under the same conditions. As the amount of available input data increases, performance further improves, reflecting the models' enhanced ability to learn richer distributional structures and more stable statistical dependencies. In particular, larger input sets allow the networks to better capture higher-order features and tail behaviors, thereby reducing variability across simulation runs and improving overall generative consistency.

From a practical perspective, these findings are especially relevant, as

they indicate that the proposed framework remains effective even in data-constrained environments commonly encountered in financial applications, while naturally benefiting from additional information when available.

Overall, the evidence points to a methodologically consistent framework operating effectively across heterogeneous distributional environments typical of quantitative modeling, alongside a learning process that progressively refines distributional representations as data availability increases.

4. Conclusion

In this paper, we provided a Deep Generative Machine Learning (Deep-GenML) framework for synthetic data generation within Financial and Energy markets. We tested and compared the abilities of the Multilayer Perceptron Generative Adversarial Network (MLP-GAN), Deep Convolutional GAN (DC-GAN), β-Variational Autoencoder (β-VAE), Adversarial-VAE (A-VAE), Stochastic Gradient Hamiltonian Monte Carlo Energy-based Model (SGHMC-EBM), and Score-based EBM (S-EBM) to replicate the distributional features of theoretical statistical distributions commonly used by practitioners. The analysis was then extended to real-world financial and commodity market data spanning multiple market regimes, thereby assessing the extent to which Deep-GenML models can generalize beyond idealized conditions. Finally, by introducing three training regimes reflecting progressively limited data availability, the study investigated model robustness under realistic information constraints faced by practitioners.

Empirical evidence confirms that the suggested framework accurately reproduces the full set of distributional properties of the reference data. The considered models demonstrated a high degree of flexibility, enabling them to capture complex density shapes, tail behavior, and higher-order statistical features with remarkable precision, as testified by the consistently strong values of the adopted accuracy metrics. Furthermore, the results indicate that the generative quality of the models is preserved when transitioning from theoretical distributions to real-world financial series, indicating that the models rely on a coherent and stable generative mechanism that remains effective across structurally different distributions.

The comparative analysis indicates that VAEs and EBMs emerge as the best-performing model families, consistently outperforming alternative architectures, as reflected by higher Overlap Coefficient (OC) values. Within these classes, the A-VAE, the SGHMC-EBM and the S-EBM rank among the

top-performing specifications. Although differences among the leading models are not always statistically significant, the results highlight that model selection remains inherently context-dependent, with practitioners and researchers facing trade-offs between distributional fidelity, architectural complexity, and computational efficiency.

Nevertheless, despite the encouraging results obtained thus far, several avenues for future research remain open. Further developments may explore alternative model architectures, specifications, and learning algorithms to better balance generative accuracy and computational efficiency. In addition, extending the generative framework to incorporate conditional information (e.g. macroeconomic variables, market sentiment indicators, policy-related signals) may substantially enhance its flexibility and practical applicability. Finally, the proposed approach can be extended to other complex financial systems and asset classes. Actually, all these topics represent a part of our ongoing research.

Funding information

This study was funded by the European Union in the framework of: Next Generation EU – *"Just Energy Transition – JET: Stochastic and Machine Learning Methods for the Evaluation, Mitigation and Geographical Hedging of Involved Natural Risks (with climate in view)"*, National Recovery and Resilience Plan (NRRP), Mission 4, C2, Intervention 1.1 - Notice: Progetti di Rilevante Interesse Nazionale (PRIN) 2022, DD No. 104 dated 2/2/2022, P2022XTLM2, CUP J53D23015530001.

Data availability

The authors do not have permission to share data.

References

[1] Ackley, D.H., Hinton, G.E., Sejnowski, T.J., 1985. A learning algorithm for boltzmann machines. Cognitive Science 9, 147–169. doi:`https://doi.org/10.1016/S0364-0213(85)80012-4`.

[2] Chen, J., Chun, D., Patel, M., Chiang, E., James, J., 2019. The validity of synthetic clinical data: a validation study of a leading synthetic data

generator (Synthea) using clinical quality measures. BMC Medical Informatics and Decision Making 19. doi:10.1186/s12911-019-0793-0.

[3] Figueira, A., Vaz, B., 2022. Survey on Synthetic Data Generation, Evaluation Methods and GANs. Mathematics 10. doi:10.3390/math10152733.

[4] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2020. Generative adversarial networks. Commun. ACM 63, 139–144. doi:10.1145/3422622.

[5] Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial networks.

[6] Hong, Y., Park, S., Kim, H., Kim, H., 2021. Synthetic data generation using building information models. Automation in Construction 130, 103871. doi:https://doi.org/10.1016/j.autcon.2021.103871.

[7] Huang, A., Khushi, M., Suleiman, B., 2023. Regime-Specific Quant Generative Adversarial Network: A Conditional Generative Adversarial Network for Regime-Specific Deepfakes of Financial Time Series. Applied Sciences 13. doi:10.3390/app131910639.

[8] Jain, S., Seth, G., Paruthi, A., Soni, U., Kumar, G., 2022. Synthetic data augmentation for surface defect detection and classification using deep learning. Journal of Intelligent Manufacturing 33, 1007 – 1020.

[9] Kingma, D.P., Welling, M., 2013. Auto-encoding variational bayes. URL: https://arxiv.org/abs/1312.6114, arXiv:1312.6114.

[10] Larochelle, H., Murray, I., 2011. The Neural Autoregressive Distribution Estimator, in: Gordon, G., Dunson, D., Dudík, M. (Eds.), Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, PMLR, Fort Lauderdale, FL, USA. pp. 29–37.

[11] Lee, D., Guan, C., Yu, Y., Ding, Q., 2024. A Comprehensive Review of Generative AI in Finance. FinTech 3, 460–478. doi:10.3390/fintech3030025.

[12] Liu, D., Chen, K., Cai, Y., Tang, Z., 2024. Interpretable EU ETS Phase 4 prices forecasting based on deep generative data augmentation approach. Finance Research Letters 61, 105038. doi:https://doi.org/10.1016/j.frl.2024.105038.

[13] Liu, Y., Acharya, U.R., Tan, J.H., 2025. Preserving privacy in healthcare: A systematic review of deep learning approaches for synthetic data generation. Computer Methods and Programs in Biomedicine 260, 108571. doi:https://doi.org/10.1016/j.cmpb.2024.108571.

[14] Wes McKinney, 2010. Data Structures for Statistical Computing in Python, in: Stéfan van der Walt, Jarrod Millman (Eds.), Proceedings of the 9th Python in Science Conference, pp. 56 – 61. doi:10.25080/Majora-92bf1922-00a.

[15] Meldrum, J., Suleiman, B., Rabhi, F., Alibasa, M.J., 2025. New money: A systematic review of synthetic data generation for finance. doi:https://doi.org/10.48550/arXiv.2510.26076, arXiv:2510.26076.

[16] Meyer, D., Nagler, T., Hogan, R.J., 2021. Copula-based synthetic data augmentation for machine-learning emulators. Geoscientific Model Development 14, 5205–5215. URL: https://gmd.copernicus.org/articles/14/5205/2021/, doi:10.5194/gmd-14-5205-2021.

[17] Pacheco, F., Hermosilla, G., Piña, O., Villavicencio, G., Allende-Cid, H., Palma, J., Valenzuela, P., García, J., Carpanetti, A., Minatogawa, V., Suazo, G., León, A., López, R., Novoa, G., 2022. Generation of Synthetic Data for the Analysis of the Physical Stability of Tailing Dams through Artificial Intelligence. Mathematics 10. doi:10.3390/math10234396.

[18] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S., 2019. Pytorch: An imperative style, high-performance deep learning library. URL: https://arxiv.org/abs/1912.01703, arXiv:1912.01703.

[19] Pathare, A., Mangrulkar, R., Suvarna, K., Parekh, A., Thakur, G., Gawade, A., 2023. Comparison of tabular synthetic data generation techniques using propensity and cluster log metric. International Journal

of Information Management Data Insights 3, 100177. doi:https://doi.org/10.1016/j.jjimei.2023.100177.

[20] Ramzan, F., Sartori, C., Consoli, S., Recupero, D.R., 2024. Generative Adversarial Networks for Synthetic Data Generation in Finance: Evaluating Statistical Similarities and Quality Assessment. AI 5, 667–685. doi:10.3390/ai5020035.

[21] Sohl-Dickstein, J., Weiss, E.A., Maheswaranathan, N., Ganguli, S., 2015. Deep unsupervised learning using nonequilibrium thermodynamics, in: Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37, JMLR.org. p. 2256–2265.

[22] Tiwari, V., Wang, J., 2025. GANterpolate: Improving Climate Modeling through Synthetic Data Generation, in: 2025 International Conference on Artificial Intelligence, Computer, Data Sciences and Applications (ACDSA), pp. 1–7. doi:10.1109/ACDSA65407.2025.11165906.

[23] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need, in: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), Advances in Neural Information Processing Systems, Curran Associates, Inc.

[24] Wang, Z., Wang, B., Li, Y., Liu, S., Li, H., Watada, J., 2024. Cross-modal scenario generation for stock price forecasting using Wasserstein GAN and GCN. Applied Soft Computing 167, 112342. doi:https://doi.org/10.1016/j.asoc.2024.112342.

Appendix A.

Table A.1: Standard Normal distribution ($N = 10000$)

Performance Indicator	n	**MLP-GAN**	**DC-GAN**	β**-VAE**	**A-VAE**	**SGHMC-EBM**	**S-EBM**
ED	21	0.0568	0.0740	0.0135	0.0205	0.0627	0.0407
	63	0.0341	0.0439	0.0278	0.0137	0.0697	0.0361
	126	0.0380	0.0258	0.0169	0.0110	0.0259	0.0213
WD	21	0.0777	0.0937	0.0193	0.0284	0.0840	0.0567
	63	0.0476	0.0638	0.0419	0.0195	0.0917	0.0486
	126	0.0482	0.0332	0.0237	0.0144	0.0378	0.0269
MMD	21	0.0020	0.0022	0.0001	0.0002	0.0012	0.0008
	63	0.0009	0.0005	0.0005	0.0001	0.0019	0.0004
	126	0.0007	0.0003	0.0001	0.00004	0.0005	0.0002
JSD	21	0.0608	0.0535	0.0393	0.0427	0.0493	0.0816
	63	0.0507	0.0351	0.0408	0.0340	0.0547	0.0482
	126	0.0415	0.0405	0.0359	0.0404	0.0391	0.0375
OC	21	94.70%	95.09%	96.76%	96.38%	94.63%	93.68%
	63	95.47%	96.82%	97.18%	96.39%	95.08%	95.34%
	126	95.82%	97.16%	97.28%	97.27%	96.79%	96.83%

Table A.2: Student's-t distribution ($N = 10000$)

Performance Indicator	n	**MLP-GAN**	**DC-GAN**	β**-VAE**	**A-VAE**	**SGHMC-EBM**	**S-EBM**
ED	21	0.0525	0.1188	0.0546	0.0401	0.0825	0.0810
	63	0.0490	0.1185	0.0461	0.0231	0.0605	0.0627
	126	0.0252	0.0531	0.0222	0.0254	0.0331	0.0362
WD	21	0.1170	0.2505	0.1355	0.0830	0.1805	0.1737
	63	0.1010	0.3170	0.1047	0.0549	0.1148	0.1429
	126	0.0597	0.1063	0.0562	0.0581	0.0807	0.0804
MMD	21	0.0008	0.0049	0.0002	0.0003	0.0015	0.0008
	63	0.0007	0.0020	0.0003	0.0002	0.0005	0.0005
	126	0.0004	0.0003	0.0001	0.0003	0.0003	0.0005
JSD	21	0.0609	0.0912	0.0472	0.0435	0.0560	0.0816
	63	0.0421	0.0799	0.0497	0.0412	0.0358	0.0710
	126	0.0464	0.0386	0.0364	0.0382	0.0405	0.0544
OC	21	95.49%	90.89%	97.51%	96.63%	94.95%	95.75%
	63	96.53%	93.66%	97.33%	97.05%	96.84%	96.26%
	126	96.88%	97.21%	97.82%	97.25%	97.35%	96.58%

Table A.3: Weibull distribution ($N = 10000$)

Performance Indicator	n	**MLP-GAN**	**DC-GAN**	β-**VAE**	**A-VAE**	**SGHMC-EBM**	**S-EBM**
ED	21	0.0297	0.1083	0.0602	0.0312	0.0854	0.0711
	63	0.0112	0.0534	0.0272	0.0201	0.0309	0.0523
	126	0.0138	0.0302	0.0254	0.0175	0.0273	0.0200
WD	21	0.0436	0.1447	0.0812	0.0326	0.1012	0.1027
	63	0.0156	0.0725	0.0401	0.0268	0.0489	0.0711
	126	0.0185	0.0395	0.0357	0.0210	0.0393	0.0317
MMD	21	0.0004	0.0038	0.0016	0.0006	0.0033	0.0021
	63	0.0001	0.0012	0.0004	0.0002	0.0002	0.0016
	126	0.00005	0.0005	0.0002	0.0002	0.0003	0.0001
JSD	21	0.0569	0.0625	0.0627	0.0709	0.0644	0.0695
	63	0.0413	0.0581	0.0534	0.0609	0.0399	0.0662
	126	0.0467	0.0408	0.0507	0.0514	0.0457	0.0443
OC	21	95.87%	93.91%	95.43%	95.30%	93.98%	93.56%
	63	96.26%	94.45%	96.28%	95.84%	96.36%	94.28%
	126	96.25%	96.26%	96.38%	95.93%	96.69%	96.11%

Table A.4: Johnson's distribution ($N = 10000$)

Performance Indicator	n	**MLP-GAN**	**DC-GAN**	β**-VAE**	**A-VAE**	**SGHMC-EBM**	**S-EBM**
ED	21	0.0747	0.1255	0.1235	0.0480	0.0955	0.4375
	63	0.0625	0.0713	0.1231	0.0317	0.0266	0.3465
	126	0.0470	0.1162	0.0622	0.0232	0.0217	0.1587
WD	21	0.2064	0.2957	0.2689	0.1174	0.2170	0.9265
	63	0.1771	0.1667	0.2518	0.0859	0.0734	0.7691
	126	0.1406	0.2092	0.1616	0.0676	0.0631	0.3684
MMD	21	0.0018	0.0051	0.0027	0.0003	0.0009	0.0268
	63	0.0017	0.0019	0.0021	0.0002	0.0002	0.0145
	126	0.0006	0.0023	0.0005	0.0001	0.0001	0.0024
JSD	21	0.0875	0.0817	0.0678	0.0423	0.0445	0.1561
	63	0.0757	0.0769	0.0594	0.0370	0.0345	0.1239
	126	0.0678	0.0687	0.0443	0.0359	0.0282	0.0732
OC	21	93.11%	91.74%	93.80%	96.91%	96.51%	82.41%
	63	94.55%	94.53%	94.37%	97.77%	97.29%	86.71%
	126	95.81%	94.54%	97.44%	98.02%	98.72%	93.82%

Table A.5: Gaussian Mixture distribution ($N = 10000$)

Performance Indicator	n	**MLP-GAN**	**DC-GAN**	β**-VAE**	**A-VAE**	**SGHMC-EBM**	**S-EBM**
ED	21	0.3028	0.1333	0.0701	0.0670	0.1010	0.1862
	63	0.3034	0.1217	0.0515	0.0302	0.0777	0.0876
	126	0.3224	0.1011	0.0391	0.0222	0.0326	0.0335
WD	21	0.5344	0.1899	0.1218	0.0873	0.1867	0.2868
	63	0.5161	0.1566	0.0863	0.0504	0.1252	0.1576
	126	0.5621	0.1424	0.0672	0.0345	0.0594	0.0458
MMD	21	0.0120	0.0141	0.0023	0.0023	0.0038	0.0118
	63	0.0162	0.0091	0.0021	0.0005	0.0020	0.0052
	126	0.0155	0.0044	0.0008	0.0004	0.0005	0.0008
JSD	21	0.1979	0.1914	0.1222	0.0756	0.0984	0.1275
	63	0.1973	0.1531	0.1009	0.0584	0.0578	0.1070
	126	0.1985	0.1375	0.0805	0.0528	0.0498	0.0514
OC	21	88.91%	80.89%	92.51%	93.18%	92.14%	86.72%
	63	87.88%	86.21%	92.97%	96.08%	94.68%	90.93%
	126	87.43%	90.23%	94.51%	96.25%	96.56%	95.48%

Table A.6: STOXX Europe 600 Climate Transition Benchmark distribution ($N = 1449$)

Performance Indicator	n	**MLP-GAN**	**DC-GAN**	β**-VAE**	**A-VAE**	**SGHMC-EBM**	**S-EBM**
ED	21	0.0193	0.0780	0.0674	0.0138	0.0518	0.0084
	63	0.0134	0.0615	0.0374	0.0086	0.0299	0.0107
	126	0.0107	0.0373	0.0329	0.0105	0.0611	0.0060
WD	21	0.0071	0.0241	0.0925	0.0059	0.0696	0.0032
	63	0.0052	0.0198	0.0503	0.0033	0.0443	0.0038
	126	0.0040	0.0134	0.0426	0.0043	0.0751	0.0028
MMD	21	0.00002	0.0007	0.0019	0.40×10^{-5}	0.0017	0.38×10^{-6}
	63	0.00003	0.0004	0.0009	0.18×10^{-4}	0.0004	0.84×10^{-5}
	126	0.00001	0.0001	0.0006	0.24×10^{-4}	0.0017	0.95×10^{-6}
JSD	21	0.1285	0.2130	0.1116	0.1191	0.1068	0.1057
	63	0.1263	0.1727	0.0994	0.1123	0.1026	0.0857
	126	0.1120	0.1503	0.0974	0.1202	0.1001	0.0818
OC	21	89.18%	76.45%	91.51%	91.25%	89.84%	90.88%
	63	89.66%	81.77%	91.65%	91.37%	92.52%	91.32%
	126	89.92%	86.38%	92.41%	93.05%	90.34%	92.64%

Table A.7: Russian 10-year government bond spot rates distribution ($N = 5009$)

Performance Indicator	n	**MLP-GAN**	**DC-GAN**	β**-VAE**	**A-VAE**	**SGHMC-EBM**	**S-EBM**
ED	21	0.1034	0.1238	0.0584	0.0346	0.0664	0.1119
	63	0.0977	0.0506	0.0407	0.0456	0.0563	0.0395
	126	0.0892	0.0419	0.0519	0.0482	0.0562	0.0241
WD	21	0.1599	0.2269	0.0613	0.0470	0.0832	0.1731
	63	0.1549	0.0802	0.0474	0.0772	0.0747	0.0677
	126	0.1409	0.0718	0.0578	0.0710	0.0627	0.0368
MMD	21	0.0100	0.0033	0.0021	0.0008	0.0027	0.0031
	63	0.0087	0.0021	0.0008	0.0010	0.0010	0.28×10^{-3}
	126	0.0075	0.0013	0.0017	0.0013	0.0015	0.23×10^{-3}
JSD	21	0.1489	0.1105	0.1093	0.0880	0.1094	0.1151
	63	0.1410	0.0973	0.1028	0.0880	0.0876	0.0795
	126	0.1351	0.1046	0.1001	0.0880	0.0994	0.0721
OC	21	82.46%	89.87%	89.02%	90.83%	88.85%	87.03%
	63	83.50%	90.10%	90.24%	91.04%	90.81%	91.71%
	126	84.44%	90.12%	91.00%	91.94%	90.94%	92.38%

Table A.8: Chinese 10-year government bond spot rates distribution ($N = 4245$)

Performance Indicator	n	**MLP-GAN**	**DC-GAN**	β**-VAE**	**A-VAE**	**SGHMC-EBM**	**S-EBM**
ED	21	0.0351	0.0579	0.0651	0.0275	0.1376	0.0597
	63	0.0332	0.0405	0.0545	0.0245	0.0776	0.0529
	126	0.0326	0.0394	0.0277	0.0292	0.0392	0.0368
WD	21	0.0330	0.0495	0.0798	0.0243	0.1598	0.0514
	63	0.0298	0.0352	0.0653	0.0231	0.0795	0.0438
	126	0.0311	0.0319	0.0342	0.0239	0.0422	0.0338
MMD	21	0.0001	0.0015	0.0031	0.0002	0.0083	0.0016
	63	0.0004	0.0001	0.0022	0.0003	0.0034	0.0013
	126	0.0004	0.0005	0.0004	0.0005	0.0007	0.0006
JSD	21	0.1211	0.1398	0.1392	0.0973	0.1424	0.1167
	63	0.1175	0.1322	0.1206	0.0885	0.1280	0.1219
	126	0.1210	0.1191	0.0991	0.1125	0.1101	0.1054
OC	21	87.33%	84.94%	87.46%	89.84%	87.56%	87.11%
	63	88.28%	86.40%	88.34%	92.10%	87.73%	87.97%
	126	88.25%	88.24%	90.20%	89.58%	88.62%	89.79%

Table A.9: Natural Gas prices distribution on 1 year futures contracts ($N = 3545$)

Performance Indicator	n	**MLP-GAN**	**DC-GAN**	β-**VAE**	**A-VAE**	**SGHMC-EBM**	**S-EBM**
ED	21	1.2827	0.6643	0.3114	0.3146	0.3218	0.5632
	63	1.4427	0.7430	0.1813	0.2683	0.1635	0.5823
	126	1.4806	0.5190	0.1111	0.2111	0.1370	0.2333
WD	21	9.3386	2.9573	0.6552	2.4135	0.8619	4.4558
	63	10.2672	2.6784	0.4040	2.1432	0.3537	4.5031
	126	10.4617	2.1996	0.2780	1.7650	0.3309	1.7586
MMD	21	0.0143	0.0420	0.0192	0.0035	0.0209	0.0049
	63	0.0161	0.0733	0.0083	0.0018	0.0055	0.0043
	126	0.0167	0.0356	0.0027	0.0032	0.0040	0.0033
JSD	21	0.2595	0.2457	0.1764	0.1232	0.1547	0.1890
	63	0.2531	0.2393	0.1439	0.0977	0.1208	0.1512
	126	0.2566	0.2314	0.1252	0.1176	0.0957	0.0959
OC	21	78.42%	79.16%	82.64%	89.70%	83.05%	85.59%
	63	80.33%	80.15%	88.03%	93.03%	89.30%	87.27%
	126	78.71%	80.68%	90.81%	91.24%	92.80%	94.51%

Table A.10: Coal prices distribution on 1 year futures contracts ($N = 2089$)

Performance Indicator	n	**MLP-GAN**	**DC-GAN**	β**-VAE**	**A-VAE**	**SGHMC-EBM**	**S-EBM**
ED	21	0.7708	1.0504	0.0794	0.4186	0.1058	0.4397
	63	0.4619	1.3641	0.0662	0.2883	0.1021	0.5071
	126	0.3543	1.2720	0.0559	0.4751	0.0537	0.4652
WD	21	8.4480	9.1114	0.1192	4.1335	0.1337	5.0287
	63	5.9950	13.4714	0.0882	3.4110	0.1607	5.4951
	126	4.7831	12.5047	0.0816	5.3646	0.0703	4.4231
MMD	21	0.0048	0.0110	0.0033	0.0022	0.0058	0.0025
	63	0.0020	0.0052	0.0020	0.0016	0.0024	0.0024
	126	0.0021	0.0048	0.0012	0.0025	0.0010	0.0026
JSD	21	0.1917	0.2647	0.1526	0.1159	0.1312	0.1196
	63	0.1317	0.2159	0.1232	0.1129	0.1229	0.1142
	126	0.1328	0.1987	0.1227	0.1333	0.1175	0.1229
OC	21	77.69%	68.79%	87.13%	88.92%	86.38%	87.34%
	63	86.33%	76.82%	87.63%	90.07%	86.98%	87.47%
	126	87.60%	78.07%	88.31%	85.02%	87.40%	86.68%