CMS-VAE: A Strategy-aware Variational AutoEncoder for High-Fidelity Crypto Market Simulation

Yihao Ang[1] **Yifan Bao**[1] **Qiang Huang**[2,*] **Qiang Wang**[1] **Xinyu Xi**[1] **Shuyu Lu**[1]
Anthony K. H. Tung[1] **Zhiyong Huang**[1]
[1]Department of Computer Science, National University of Singapore
[2]School of Intelligence Science and Engineering, Harbin Institute of Technology (Shenzhen)
{yihao_ang, yifan_bao, atung, huangzy}@comp.nus.edu.sg
{qwang, xinyu_xi, lushuyu}@u.nus.edu, huangqiang@hit.edu.cn

Abstract

Cryptocurrency markets exhibit extreme volatility, non-stationarity, and complex inter-asset dependencies, posing significant challenges for generating realistic synthetic data–a crucial need for risk management, backtesting, and strategy development. While recent Time Series Generation (TSG) models, such as Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and diffusion methods, have shown promise, they often fall short in capturing crypto-specific dynamics, generalizing effectively, and aligning synthetic data with trading objectives. To address these challenges, we propose **CMS-VAE**, a **VAE**-based framework tailored for **C**rypto **M**arket **S**imulation. CMS-VAE employs a dilated CNN architecture to model long-range temporal dependencies and cross-asset correlations, and introduces the Ensemble Financial Performance Loss (EFPL), which integrates strategy-aware supervision over diverse strategies to produce strategy-consistent and risk-aligned synthetic data. Extensive experiments across generative fidelity, predictive modeling, and statistical arbitrage show that CMS-VAE consistently outperforms state-of-the-art baselines. It achieves up to 96.8% lower prediction errors and 1.4× improvements in the Sharpe ratio. These results position CMS-VAE as an effective and efficient tool for high-fidelity crypto market simulation.

1 Introduction

The rise of DeFi and digital assets has pushed cryptocurrencies into the mainstream, but their extreme volatility, structural non-stationarity, fragmented liquidity, and strong cross-asset co-movements strain conventional models and complicate risk assessment, pricing, and trading. High-fidelity Time Series Generation (TSG) is a promising remedy [2], yet state-of-the-art generators, GANs [26, 23, 25, 6, 21], VAEs [7, 15, 18], diffusion [27, 9, 5, 16, 17], and mixed/flow models [1, 28, 12, 22], are typically validated on generic benchmarks and struggle in crypto due to (1) weak modeling of inter-asset dependencies critical for systemic moves, (2) training/sampling inefficiency (notably in GANs/diffusion) that hinders time-sensitive deployment, and (3) lack of strategy-aware objectives that tie synthetic data to financial utility [3, 2].

Interestingly, among these approaches, VAE-based models have demonstrated competitive (sometimes superior) performance in both generation fidelity and efficiency, especially under conditions of limited data or irregular sampling [2, 3]. This success stems from VAEs' structured latent spaces, stable training dynamics, and explicit likelihood modeling, which collectively mitigate issues such as mode

*Corresponding author.

39th Conference on Neural Information Processing Systems (NeurIPS 2025).

collapse and adversarial instability prevalent in GANs. However, existing variants (e.g., **TimeVAE**, discrete-latent **TimeVQVAE**, Koopman-based **KoVAE**) are not tailored to crypto's volatility/regime shifts and are commonly evaluated with generic, finance-agnostic metrics [3].

To address these challenges, we introduce **CMS-VAE**, a VAE tailored for **C**rypto **M**arket **S**imulation that integrates domain-specific priors with finance-aligned objectives: (1) A dilated 1D-CNN encoder–decoder assets as channels, efficiently capturing short-term fluctuations, long-horizon dependencies, and cross-asset co-movements. (2) An Ensemble Financial Performance Loss (EFPL) augments ELBO with portfolio-level risk-return measures, guiding the latent space toward financially meaningful structure. (3) Strategy-aware supervision evaluates EFPL across diverse trading strategies, ensuring robustness across regimes and investment hypotheses. Comprehensive experiments demonstrate that **CMS-VAE** delivers strong generative fidelity, superior forecasting and ranking accuracy, and fee-robust statistical-arbitrage gains over state-of-the-art baselines on crypto benchmarks [3], offering a practical capability–compute balance for high-stakes financial modeling.

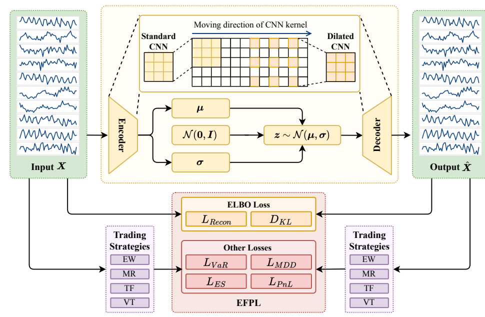


Figure 1: Architecture of CMS-VAE.

2 Related Work

Synthetic TSG methods span GANs, VAEs, diffusion models, and mixed-type hybrids, each trading off fidelity, stability, and efficiency [3, 2]. GANs capture realistic dynamics but suffer from instability and mode collapse [26, 23, 21, 8, 2]. VAEs train stably yet underfit volatility and cross-dependencies [7, 15, 18, 2, 3]. Diffusion models achieve high fidelity at heavy sampling cost [27, 9, 24, 3]. Mixed-types combine strengths with added complexity [28, 12]. In finance, models such as Quant-GAN [25], TailGAN [6], and Sig-GAN [19] target stylized facts like volatility clustering and jumps, with recent work probing what patterns architectures truly learn [14, 11]. Yet instability, computational overhead, and weak transfer across trading objectives remain, especially in crypto markets [3].

3 Preliminaries

Problem Formulation. We investigate the generation of synthetic cryptocurrency market data for risk modeling and trading. This task is to simulate 24-hour log-return sequences by learning a generator $\boldsymbol{g} : \mathbb{Z} \to \mathbb{R}^{24\times 3}$ that maps a latent noise vector $\boldsymbol{z}$ to realistic sequences resembling historical data.

Variational Autoencoders (VAEs). To model this generative process, we adopt the framework of VAEs [13]. Let $\boldsymbol{X}$ denote the training set of time series, where each $\boldsymbol{x} \in \boldsymbol{X}$ is assumed to arise from a latent variable model $\boldsymbol{p}(\boldsymbol{x}|\boldsymbol{z})$ with $\boldsymbol{z}$ drawn from a lower-dimensional latent space. As illustrated in the yellow block of Figure 1, a VAE consists of an encoder $\boldsymbol{q_\phi}(\boldsymbol{z}|\boldsymbol{x})$ that maps an input $\boldsymbol{x}$ to a Gaussian distribution parameterized by mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\sigma}$, and a decoder $\boldsymbol{p_\theta}(\boldsymbol{x}|\boldsymbol{z})$ that reconstructs the time series from latent code $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\sigma})$. The decoder outputs $\boldsymbol{p_\theta}(\boldsymbol{x}|\boldsymbol{z}) = \mathcal{N}(\boldsymbol{\eta_\theta}(\boldsymbol{z}), \boldsymbol{\Sigma}^2\boldsymbol{I})$, with parameters $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$ jointly optimized. Training maximizes the Evidence Lower Bound (ELBO), which balances reconstruction fidelity and latent regularization:

$$L_{\text{ELBO}} = L_{\text{Recon}} + D_{\text{KL}} = \mathbb{E}_{\boldsymbol{q_\phi}(\boldsymbol{z}|\boldsymbol{x})}[\log \boldsymbol{p_\theta}(\boldsymbol{x}|\boldsymbol{z})] - D_{\text{KL}}(\boldsymbol{q_\phi}(\boldsymbol{z}|\boldsymbol{x})||\boldsymbol{p}(\boldsymbol{z})).$$

Here, the first term measures how well $\boldsymbol{x}$ is reconstructed, while the KL divergence ensures that $\boldsymbol{q_\phi}(\boldsymbol{z}|\boldsymbol{x})$ remains close to the prior $\boldsymbol{p}(\boldsymbol{z})$, typically $\mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$. Once trained, new synthetic sequences can be generated by sampling $\boldsymbol{z}$ from the prior and decoding it via $\boldsymbol{p_\theta}(\boldsymbol{x}|\boldsymbol{z})$.

4 CMS-VAE

We propose **CMS-VAE** (Figure 1), a crypto-tailored VAE that (1) uses a dilated 1D-CNN to capture long-range temporal and cross-asset structure, (2) augments ELBO with an Ensemble Financial

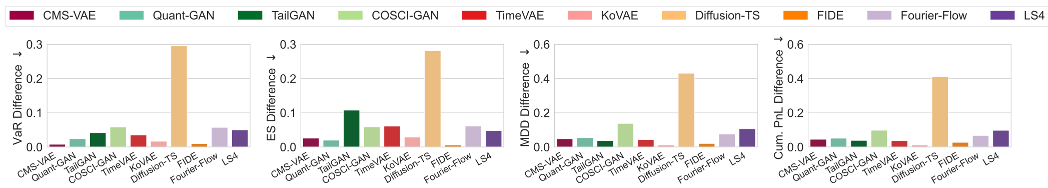


Figure 2: Generative fidelity benchmarking.

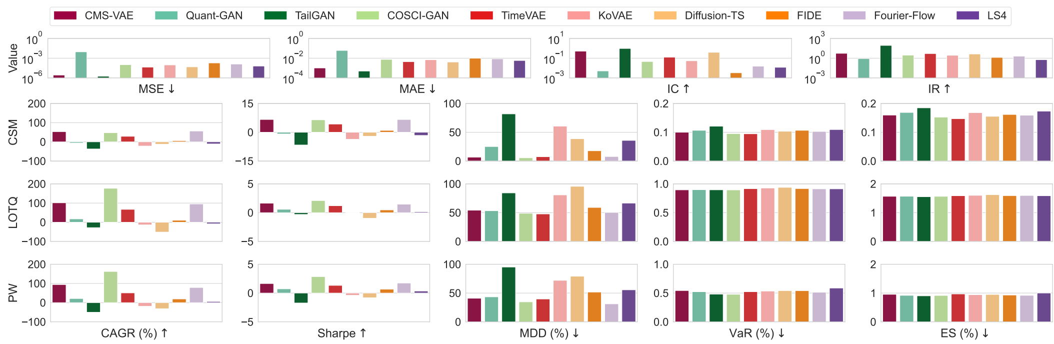


Figure 3: Predictive utility benchmarking.

Performance Loss (EFPL) to enforce portfolio-level risk–return alignment, and (3) applies strategy-aware supervision via EFPL across diverse trading strategies. Together, these components generate synthetic multi-asset series that maintain statistical fidelity while enhancing trading utility.

Dilated CNN Architecture. RNNs capture sequence dependence but scale poorly, Transformers provide global context at high compute cost, and vanilla CNNs are efficient but limited in receptive field [10, 4]. CMS-VAE addresses this by stacking dilated 1D convolutions, expanding context without excessive depth. Treating assets as channels enables learning inter-asset correlations, while dilation balances local variability with long-range dependencies. The design preserves temporal coherence, controls parameter growth, and suits volatile, high-frequency crypto data.

Ensemble Financial Performance Loss (EFPL). Standard ELBO lacks financial semantics, often producing statistically plausible but trading-misaligned samples. EFPL remedies this by penalizing discrepancies between real and synthetic Value-at-Risk (VaR), Expected Shortfall (ES), Maximum Drawdown (MDD), and Cumulative Profit-and-Loss (PnL), aggregated across multiple strategies with tunable weights. The combined loss is:

$$L_{\text{EFPL}} = w_{\text{Recon}} L_{\text{Recon}} + w_{\text{KL}} D_{\text{KL}} + w_{\text{VaR}} L_{\text{VaR}} + w_{\text{ES}} L_{\text{ES}} + w_{\text{MDD}} L_{\text{MDD}} + w_{\text{PnL}} L_{\text{PnL}},$$

where $L_{\text{VaR}} = \sum_{j=1}^{M} w_j^{\text{VaR}} \cdot \|\boldsymbol{x}_j^{\text{VaR}} - \hat{\boldsymbol{x}}_j^{\text{VaR}}\|$, $\boldsymbol{x}_j^{\text{VaR}}$ and $\hat{\boldsymbol{x}}_j^{\text{VaR}}$ are the VaR values for real and synthetic series, with analogous terms for ES, MDD, and PnL. This ensures generated sequences align with risk–return profiles relevant to practice.

Strategy-aware Supervision. To avoid overfitting to generic statistics [23, 7, 27], CMS-VAE applies EFPL under a portfolio of canonical strategies, i.e., Equal-Weight (EW), Mean-Reversion (MR), Trend-Following (TF), Volatility-Trading (VT) [3]; Robustness is evaluated on held-out strategies, i.e., Relative Strength Index (RSI), Simple-Mean (SM), BreakOut (BO), where performance gains reflect genuine transfer of risk-return structure [3]. The modular design of EFPL further allows seamless extension to additional strategies.

5 Experiments

Setup. We evaluate **CMS-VAE** on both crypto-specific and general TSG tasks across four dimensions: generative fidelity, predictive utility, statistical arbitrage, and ablation. The datasets include 24h log-return windows of the top three **cryptocurrencies** [20] and **CTBench** [3], which provides high-frequency returns for 450+ Binance assets across bull, bear, and sideways regimes. We benchmark against 9 state-of-the-art baselines: **Quant-GAN** [25], **TailGAN** [6], **COSCI-GAN** [23], **TimeVAE** [7], **KoVAE** [18], **Diffusion-TS** [27], **FIDE** [9], **Fourier-Flow** [1], and **LS4** [28]).

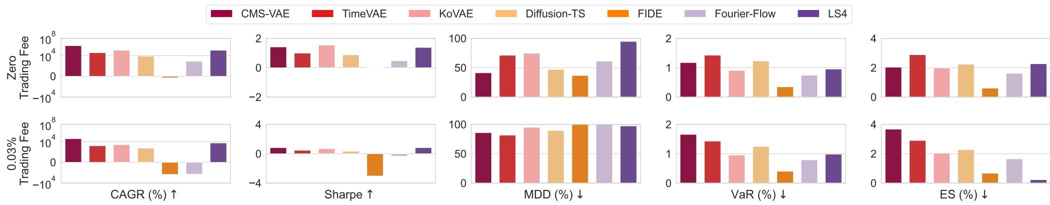
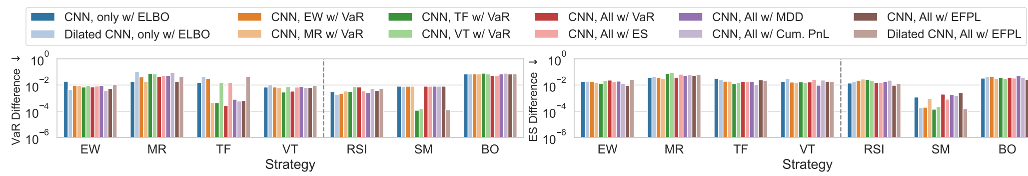
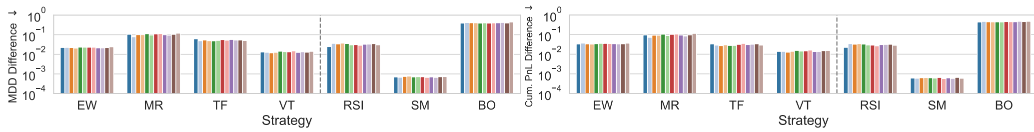


Figure 4: Statistical arbitrage benchmarking.



(a) VaR Difference.

(b) ES Difference.



(c) MDD Difference.

(d) Cum. PnL Difference.

Figure 5: Evaluation of four risk metrics across seven trading strategies for CMS-VAE model variants.

Evaluation metrics follow financial and predictive requirements. is measured by absolute gaps in portfolio-level indicators: **VaR**, **ES**, **MDD**, and **Cumulative PnL** [20]. Predictive utility and market-neutral arbitrage follow CTBench [3], using Mean Squared Error (**MSE**) and Mean Absolute Error (**MAE**) for accuracy, Information Coefficient (**IC**) and Information Ratio (**IR**) for rank consistency, and Compound Annual Growth Rate (**CAGR**), Sharpe Ratio (or simply **Sharpe**), **MDD**, **VaR**, and **ES** for trading performance. All experiments are conducted on a workstation equipped with an Intel® Xeon® Platinum 8480C @3.80GHz, 64 GB RAM, and an NVIDIA H100 GPU.

Generative Fidelity. In Figure 2, CMS-VAE achieves the smallest overall discrepancies, reducing VaR gaps by 77.6% vs. TimeVAE and >86% vs. COSCI-GAN/Fourier-Flow, while maintaining competitive MDD and PnL. Although FIDE slightly outperforms on ES, CMS-VAE delivers a more balanced risk profile by jointly modeling tails and profitability, enabled by EFPL's alignment with risk–return metrics and the dilated CNN's capture of volatility clustering.

Predictive Utility. For predictive utility (Figure 3), CMS-VAE consistently surpasses baselines, cutting MSE by up to 97.3% and boosting rank-based IC/IR by more than 1.6× and nearly 2×, respectively Unlike distribution-focused diffusion models, EFPL preserves predictive signals, translating into robust downstream trading: returns are ∼3.2× higher than TailGAN across CSM, LOTQ, and PW strategies with controlled risk, while some baselines behave erratically.

Statistical Arbitrage. In statistical arbitrage (Figure 4), CMS-VAE delivers dominant performance under both zero fees and 0.03% costs, with more than 11× higher CAGR than LS4 and ∼12× vs. KoVAE, and remains robust under fees. Though risk metrics (MDD, VaR, ES) are moderately higher, Sharpe improves by up to $+0.54$ over Diffusion-TS and $+0.03$ over LS4, with smaller fee-induced degradation, indicating fee-robust tradable signals where distribution-matching models falter.

Ablation Study. Figure 5 confirms the role of both architecture and loss design of CMS-VAE. Dilated CNNs reduce risk-metric gaps under ELBO training, with further gains from EFPL at minimal parameter cost. ELBO-only models underperform EFPL-trained ones; strategy-specific EFPL variants excel in their regimes, while multi-strategy EFPL achieves the best overall trade-off. Crucially, generalization to held-out strategies (BO, SM, RSI) remains strong, showing that EFPL captures transferable market dynamics beyond training objectives.

6 Conclusions

In this work, we present CMS-VAE, a crypto-tailored VAE for cryptocurrency market simulation. By combining the dilated CNNs with a novel EFPL loss and strategy-aware supervision, CMS-VAE effectively captures long-range dependencies, inter-asset correlations, and financial risk profiles. Empirical results show that CMS-VAE achieves superior generative fidelity across key financial metrics, strong predictive utility, and high returns with robust Sharpe ratios. These results underscore its value for backtesting and strategy development, while reduced variability supports more stable models and better decision-making in volatile markets.

Acknowledgments

This research is supported by the Ministry of Education, Singapore, under its MOE AcRF TIER 1 Grant (T1 251RES2517). Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not reflect the views of the Ministry of Education, Singapore.

References

[1] Ahmed M. Alaa, Alex James Chan, and Mihaela van der Schaar. Generative time-series modeling with fourier flows. In *ICLR*, 2021.

[2] Yihao Ang, Qiang Huang, Yifan Bao, Anthony KH Tung, and Zhiyong Huang. Tsgbench: Time series generation benchmark. *Proc. VLDB Endow.*, 17(3):305–318, 2023.

[3] Yihao Ang, Qiang Wang, Qiang Huang, Yifan Bao, Xinyu Xi, Anthony KH Tung, Chen Jin, and Zhiyong Huang. Ctbench: Cryptocurrency time series generation benchmark. *arXiv preprint arXiv:2508.02758*, 2025.

[4] Shaojie Bai, J Zico Kolter, and Vladlen Koltun. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*, 2018.

[5] Andrea Coletta, Sriram Gopalakrishnan, Daniel Borrajo, and Svitlana Vyetrenko. On the constrained time-series generation problem. In *NeurIPS*, pages 61048–61059, 2023.

[6] Rama Cont, Mihai Cucuringu, Renyuan Xu, and Chao Zhang. Tail-gan: Learning to simulate tail risk scenarios. *arXiv preprint arXiv:2203.01664*, 2022.

[7] Abhyuday Desai, Cynthia Freeman, Zuhui Wang, and Ian Beaver. Timevae: A variational auto-encoder for multivariate time series generation. *arXiv preprint arXiv:2111.08095*, 2021.

[8] Yousef El-Laham and Svitlana Vyetrenko. Styletime: Style transfer for synthetic time series generation. In *ICAIF*, pages 489–496, 2022.

[9] Asadullah Hill Galib, Pang-Ning Tan, and Lifeng Luo. Fide: Frequency-inflated conditional diffusion model for extreme-aware time series generation. In *NeurIPS*, pages 114434–114457, 2024.

[10] Y. Hu et al. Fintsb: A comprehensive benchmark for financial time series forecasting. In *arXiv:2502.18834*, 2025.

[11] Riasat Ali Istiaque, Chi Seng Pun, and Yuli Song. Simulating asset prices using conditional time-series gan. In *ICAIF*, pages 770–778, 2024.

[12] Jinsung Jeon, Jeonghak Kim, Haryong Song, Seunghyeon Cho, and Noseong Park. GT-GAN: general purpose time series synthesis with generative adversarial networks. In *NeurIPS*, pages 36999–37010, 2022.

[13] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[14] Sohyeon Kwon and Yongjae Lee. Can gans learn the stylized facts of financial time series? In *ICAIF*, pages 126–133, 2024.

[15] Daesoo Lee, Sara Malacarne, and Erlend Aune. Vector quantized time series generation with a bidirectional prior model. In *AISTATS*, pages 7665–7693, 2023.

[16] Yang Li, Han Meng, Zhenyu Bi, Ingolv T. Urnes, and Haipeng Chen. Population aware diffusion for time series generation. In *AAAI*, pages 18520–18529, 2025.

[17] Ilan Naiman, Nimrod Berman, Itai Pemper, Idan Arbiv, Gal Fadlon, and Omri Azencot. Utilizing image transforms and diffusion models for generative modeling of short and long time series. In *NeurIPS*, pages 121699–121730, 2024.

[18] Ilan Naiman, N Benjamin Erichson, Pu Ren, Michael W Mahoney, and Omri Azencot. Generative modeling of regular and irregular time series data via koopman vaes. In *ICLR*.

[19] Hao Ni, Lukasz Szpruch, Marc Sabate-Vidales, Baoren Xiao, Magnus Wiese, and Shujian Liao. Sig-wasserstein gans for time series generation. In *ICAIF*, pages 1–8, 2021.

[20] Hao Ni, Lukasz Szpruch, Jiajie Tao, and Yang Long. Crypto Market Simulation for Risk Estimation. `https://hackathon.deepintomlf.ai/competitions/40`, 2024. Antalpha ICAIF 2024 Hackathon.

[21] Hengzhi Pei, Kan Ren, Yuqing Yang, Chang Liu, Tao Qin, and Dongsheng Li. Towards generating real-world time series data. In *ICDM*, pages 469–478, 2021.

[22] Jian Qian, Bingyu Xie, Biao Wan, Minhao Li, Miao Sun, and Patrick Yin Chiang. Timeldm: Latent diffusion model for unconditional time series generation. *arXiv preprint arXiv:2407.04211*, 2024.

[23] Ali Seyfi, Jean-François Rajotte, and Raymond T. Ng. Generating multivariate time series with common source coordinated GAN (COSCI-GAN). In *NeurIPS*, pages 32777–32788, 2022.

[24] Zhuohan Wang and Carmine Ventre. A financial time series denoiser based on diffusion models. In *ICAIF*, pages 72–80, 2024.

[25] Magnus Wiese, Robert Knobloch, Ralf Korn, and Peter Kretschmer. Quant gans: deep generation of financial time series. *Quantitative Finance*, 20(9):1419–1440, 2020.

[26] Jinsung Yoon, Daniel Jarrett, and Mihaela van der Schaar. Time-series generative adversarial networks. In *NeurIPS*, pages 5509–5519, 2019.

[27] Xinyu Yuan and Yan Qiao. Diffusion-ts: Interpretable diffusion for general time series generation. In *ICLR*, 2024.

[28] Linqi Zhou, Michael Poli, Winnie Xu, Stefano Massaroli, and Stefano Ermon. Deep latent state space models for time-series generation. In *ICML*, pages 42625–42643, 2023.