

[← Go to DBLP homepage \(/group?id=DBLP.org\)](#)

# A Spatio-Temporal Diffusion Model for Missing and Real-Time Financial Data Inference

[\(https://doi.org/10.1145/3627673.3679806\)](https://doi.org/10.1145/3627673.3679806)

*Yupeng Fang (/profile?id=~Yupeng\_Fang1),  
Ruirui Liu (/profile?id=~Ruirui\_Liu1),  
Huichou Huang (/profile?id=~Huichou\_Huang1),  
Peilin Zhao (/profile?id=~Peilin\_Zhao2),  
Qingyao Wu ([https://dblp.org/search/pid/api?q=author:Qingyao\\_Wu:](https://dblp.org/search/pid/api?q=author:Qingyao_Wu:))*

Published: 01 Jan 2024, Last Modified: 13 May 2025 CIKM 2024 Everyone Revisions (/revisions?id=T8jLfskEVa) BibTeX CC BY-SA 4.0 (<https://creativecommons.org/licenses/by-sa/4.0/>)

**Abstract:** Missing values and unreleased figures are common but highly important for backtesting and real-time analysis in the financial industry, yet underexploited in the existing literature. In this paper, we focus on the issue of empirical asset pricing, where the cross-section of future asset returns is a function of lagged firm characteristics that vary in time frequencies and missing ratios. Most of the existing imputation methods cannot fully capture the complex and evolving spatio-temporal relations among firm-level characteristics. In particular, these methods fail to explicitly consider the spatial relations and feature structure in the stock network where we have to process granular data of thousands of stocks and hundreds of characteristics for each stock. To address these challenges, we propose a spatio-temporal diffusion model (STDM) that gradually recovers the masked financial data conditioning on high-dimensional stock-and-characteristics historical data. We propose characteristic-specific projection to construct characteristic-level features at both ends of the STDM, meanwhile maintaining firm-level features in the middle of the STDM to largely reduce the computational memory. Moreover, along with the temporal attention, we design a spatial graph convolutional network, making it computationally efficient and effective to learn time-varying spatio-temporal interdependence across firms. We further employ an implicit sampler that greatly accelerates the inference procedure so that the STDM is able to produce high-quality point and density estimates of missing and real-time firm characteristics within a few steps. We evaluate our model on the most comprehensive open-source dataset 'OSAP' and generate state-of-the-art performance in extensive experiments.

[About OpenReview \(/about\)](/about)  
[Hosting a Venue \(/group?id=OpenReview.net/Support\)](/group?id=OpenReview.net/Support)  
[All Venues \(/venues\)](/venues)  
[Sponsors \(/sponsors\)](/sponsors)

[Frequently Asked Questions  
\(https://docs.openreview.net/getting-started/frequently-asked-questions\)](https://docs.openreview.net/getting-started/frequently-asked-questions)  
[Contact \(/contact\)](/contact)  
[Terms of Use \(/legal/terms\)](/legal/terms)  
[Privacy Policy \(/legal/privacy\)](/legal/privacy)

[OpenReview \(/about\)](/about) is a long-term project to advance science through improved peer review with legal nonprofit status. We gratefully acknowledge the support of the [OpenReview Sponsors \(/sponsors\)](/sponsors). © 2025 OpenReview