# PANELTS: A MULTI-DOMAIN BENCHMARK DATASET FOR PANEL-BASED TIME SERIES FORECASTING

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Many real-world forecasting problems involve multiple related units observed over time, each with its own target variable and associated covariates - a structure known as panel data. Yet existing time series benchmarks are predominantly organised as collections of independent univariate series or single-entity multivariate recordings, leaving panel-structured forecasting systematically underserved. We introduce PanelTS, a multi-domain benchmark dataset designed to close this gap. PanelTS adopts a unified representation that natively supports univariate, multivariate, and panel-based forecasting tasks while preserving cross-unit structure, enabling rigorous evaluation of models that exploit inter-unit dependencies. The benchmark comprises 1,844 datasets, spanning synthetic data with controllable temporal dynamics and real-world datasets drawn from COVID-19 epidemiology, exchange-traded funds (ETFs), currency markets, and equity markets, thereby providing diversity in domain, granularity, and scale. To facilitate reproducibility and systematic comparison, PanelTS provides standardised data formats, task definitions, and evaluation protocols across varying input lengths, prediction horizons, and temporal resolutions. We report baseline results across representative forecasting models to establish reference performance and demonstrate compatibility with existing approaches. PanelTS is publicly released to support the development and fair evaluation of the next generation of panel-aware forecasting methods: https://huggingface.co/datasets/Multiple-Time-Series/PanelTS (data), https://anonymous.4open.science/status/PanelTS-020D (code), https://huggingface.co/spaces/Echo0623/PanelTS-ICLR-Review (demo).

## 1 INTRODUCTION

Forecasting problems in many real-world domains involve collections of related observational units rather than isolated time series. Exchange rates, for example, are shaped by shared macroeconomic and financial conditions, with currencies exhibiting co-movement through trade and capital-market linkages Bodie & Kane (2020); Pastorek (2023). Stock returns are similarly influenced by market-wide and sector-specific factors Fama & French (2015); Hou (2007), while regional disease trajectories may depend on spatial interactions as well as demographic and healthcare-related characteristics Arik et al. (2020); Embury et al. (2022); Mehta et al. (2020). Despite differences across domains, these scenarios share a common panel structure: each unit is associated with its own target time series, while dependencies across units arise through shared covariates, unit-level attributes, and correlated temporal dynamics.

Existing benchmarks for time series forecasting are predominantly organised either as collections of independent univariate series or as multivariate recordings of a single entity Hebrail & Berard (2012); Makridakis et al. (2018); Zhou et al. (2021). Neither is well-suited to panel settings: treating units independently discards potentially informative cross-unit structure, while collapsing all units into a single multivariate representation conflates unit-level dynamics with feature-level variation. Panel data, which track repeated observations of multiple distinct units over time, offer a more faithful representation of these scenarios, yet remain underserved by existing benchmarks. To address this gap, we introduce PanelTS, a multi-domain benchmark dataset for panel-based time series forecasting. PanelTS organizes observational units, their target series, and shared covariates under a unified format that flexibly supports univariate tasks, multivariate tasks scoped to individual units, and panel-based



Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

tasks spanning all units jointly, enabling systematic evaluation across the full spectrum of forecasting regimes, with three key contributions:

- **First benchmark for panel-based forecasting.** We introduce PanelTS, to our knowledge the first benchmark dataset specifically designed for panel-based time series forecasting. PanelTS explicitly preserves relationships across multiple units, supporting systematic evaluation of forecasting methods in panel settings.
- **Diverse real-world and synthetic scenarios.** PanelTS includes five representative scenarios spanning synthetic data generation, COVID-19 dynamics, exchange-traded funds, foreign exchange markets, and stock markets, representing diverse temporal characteristics, cross-unit dependencies, and covariate structures.
- **Unified and reproducible evaluation protocols.** PanelTS provides standardised data schemas, task definitions, and evaluation protocols across different temporal granularities, input lengths, and prediction horizons, enabling consistent and reproducible comparison of forecasting approaches under univariate, multivariate, and panel forecasting settings.

## 2 Related Work

### 2.1 Existing Forecasting Data Paradigms

**Multivariate Time Series** A multivariate time series contains multiple variables observed over time for the same underlying entity. Representative forecasting datasets include ETT Zhou et al. (2021), ILI Centers for Disease Control and Prevention (2025), Electricity Hebrail & Berard (2012), Weather Kolle (2025), Traffic Zhao et al. (2019), and PeMSD Guo et al. (2019). These datasets are widely used for modeling temporal and cross-variable dependencies, but they typically do not keep an explicit unit dimension with shared covariate definitions across units.

**Panel Data** Also referred to as longitudinal or cross-sectional time series data, panel data consist of repeated observations of multiple, often homogeneous, units over time Hsiao (2022). They are widely used in empirical research because they capture both temporal and cross-unit dynamics. Traditional panel-data analysis, however, has primarily focused on inference, causal analysis, and the estimation of unit-specific effects rather than on forecasting tasks. Modern panel-data research also emphasizes challenges such as serial correlation and cross-sectional dependence across units Gao et al. (2024).

**Multi-Compartment Data** Multi-compartment data describe multiple interacting components, each with its own attributes or states. Such data commonly arise in areas including cell biology and pharmacokinetics Nakata et al. (2012); Rescigno (1960). They are related to panel-based forecasting in that multiple entities are observed jointly over time. However, unlike panel data, the entities in multi-compartment systems are often heterogeneous (e.g., liver and kidney compartments) rather than homogeneous units such as regions, countries, or companies. Moreover, multi-compartment systems are typically modeled based on domain-specific mechanistic assumptions that describe interactions between compartments.

### 2.2 Time Series Forecasting Datasets

Table 1 summarizes representative time series forecasting datasets. Existing resources are often organised either as collections of univariate series, such as M4 Makridakis et al. (2018), or as multivariate series for a single entity, such as ETT and Electricity. Other datasets address specific structures or domains. For example, M5 Makridakis et al. (2022) provides hierarchical retail sales data with calendar and price-related features, while MIMIC-III contains rich clinical covariates but is mainly used for clinical outcome prediction rather than general forecasting. These resources are valuable, but they are not primarily designed for forecasting settings where multiple related units are observed over time with shared covariates.

Recent datasets have further expanded the scope of time series evaluation. ProbTS Zhang et al. (2024) emphasizes probabilistic forecasting across horizons, while the traffic benchmark LargeST Liu et al. (2023) and BuildingsBench Emami et al. (2023) focus on large-scale application-specific forecasting. Job-SDF Chen et al. (2024) studies multi-granularity forecasting in labor-market data,

Table 1: Comparison of Time Series Forecasting Datasets.

| Name | Description | Multi domain | Panel based | Shared covariates across units |
|---|---|---|---|---|
| PanelTS (Ours) | Panel-based dataset with shared covariates | ✓ | ✓ | ✓ |
| M4 | Large-scale univariate dataset across multiple domains | ✓ | ✗ | ✗ |
| M5 | Large-scale hierarchical retail sales forecasting dataset | ✗ | ✓ | ✗ |
| ILI | Influenza-like illness multivariate time series dataset | ✗ | ✗ | ✗ |
| ETT | Multivariate transformer temperature measurements | ✗ | ✗ | ✗ |
| Electricity | Multivariate electricity consumption time series | ✗ | ✗ | ✗ |
| Weather | Multivariate weather observation time series | ✗ | ✗ | ✗ |
| Traffic | Multivariate traffic flow and speed measurements | ✗ | ✗ | ✗ |
| PeMSD (4/7/8) | Sensor-network traffic forecasting datasets | ✗ | ✓ | ✗ |
| Solar-Power | Solar power generation time series dataset | ✗ | ✗ | ✗ |
| MIMIC-III | Clinical time series with patient-level variables | ✗ | ✗ | ✓ |

and PSML Zheng et al. (2022) provides multi-scale time series for evaluating models across temporal resolutions. Despite their importance, these datasets are primarily designed for specific domains, scales, or evaluation settings.

In contrast, PanelTS establishes a unified benchmark for panel-based time series forecasting, supporting the construction of univariate, multivariate, and panel-based forecasting tasks within a common data framework. It provides standardised task definitions, evaluation protocols, and temporal forecasting structures that are generally absent from existing panel datasets such as the Penn World Table Growth & Centre (2023) and the Panel Study of Income Dynamics Social Research Center, which are primarily intended for economic and social-science analysis rather than machine learning forecasting research.

## 3 Data Organization of PanelTS

### 3.1 Data Organization for Panel-Based Forecasting

Panel-based forecasting requires multi-unit time series data that preserve the underlying unit structure. Each unit has its own target variable and covariates, while corresponding covariates share common definitions across units whenever applicable. This organization is designed to capture both temporal dynamics (i.e., changes over time) and cross-unit dynamics (i.e., variation across units).

Figure 1 compares univariate, multivariate, and panel-based data organizations. In the univariate setting shown, each unit is represented by a single target series $y_{i,t}$ and is forecast using its own history. Covariates and cross-unit relationships are not explicitly represented. The multivariate setting groups multiple variables $z_{1,t}, \ldots, z_{D,t}$ within one unit, allowing their temporal and inter-variable dependencies to be modeled. This organization describes relationships within a single unit, without an explicit cross-unit structure.

PanelTS differs from these organizations by organizing all units within a shared panel structure. Each unit contains one target variable together with a common set of covariate types shared across units. This organization makes it possible to represent relationships between targets and covariates within units, as well as relationships among corresponding covariates and targets across units. The goal is not to assume that such relationships always improve forecasting performance, but to provide a data organization in which they can be systematically studied.

### 3.2 PanelTS Structure

**Definition.** A PanelTS dataset is represented as

$$\mathcal{D} = \{(y_{i,t}, \mathbf{x}_{i,t}) : i = 1, \ldots, N,\ t = 1, \ldots, T\},$$

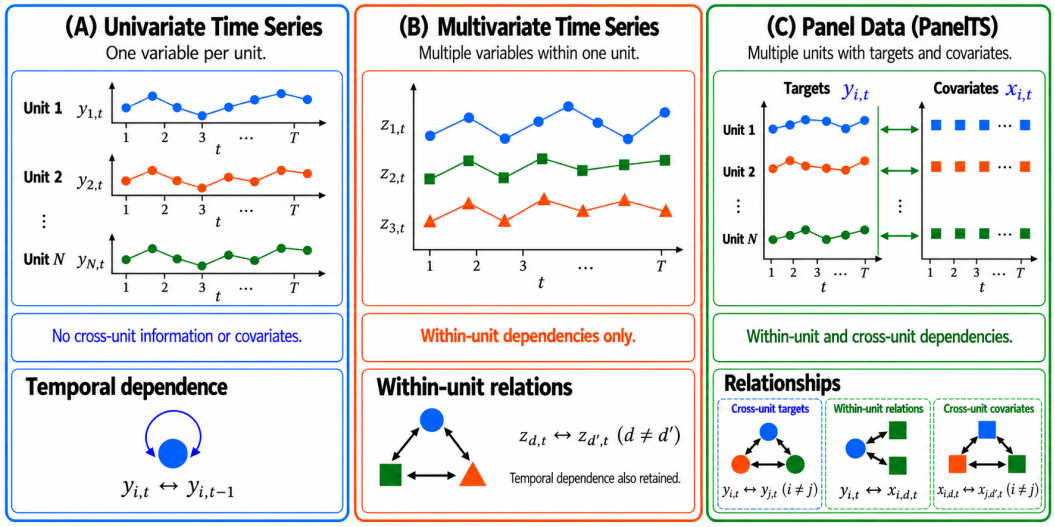



Figure 1: Comparison of data organizations for panel-based forecasting. (A) Univariate forecasting uses each unit's target history independently. (B) Multivariate forecasting models multiple variables within one unit, including their temporal and inter-variable dependencies. (C) PanelTS organizes multiple units with one target per unit and a common set of covariate types. It preserves target–covariate relationships within units and relationships among targets and covariates across units.

where $i$ indexes the observational unit and $t$ indexes time. Here, $y_{i,t}$ denotes the target variable of unit $i$ at time $t$, and $\mathbf{x}_{i,t}$ denotes the $M$ covariates associated with the same unit and time point as:

$$\mathbf{x}_{i,t} = (x_{i,1,t}, \ldots, x_{i,M,t}).$$

Within each dataset, units are aligned on a common timeline. Covariates of the same type may exhibit correlations across units; for example, $x_{i,M,t}$ and $x_{j,M,t}$ may be related for different units $i$ and $j$ at time $t$. PanelTS stores target variables and covariates in separate but linked data files. This design allows users to flexibly select the information required for a given forecasting setup, including target histories alone, target histories with covariates, or multiple related units with shared covariate fields. Consequently, the same data source can support both standard time-series forecasting models and methods specifically designed for panel forecasting.

Although PanelTS is intended for forecasting tasks, its structure also enables the application of panel-data analysis methods, such as pooled regression and fixed-effects models Baltagi (2008); Gao (2007); Wooldridge (2010).[1]

## 4 Datasets in PanelTS

PanelTS consists of five dataset groups, as summarized in Table 2. All groups follow a unified panel-based organization, where multiple units are observed over time, each unit has its own target series, and covariates are consistently defined across units within each group. In total, PanelTS comprises 1,844 datasets and occupies 2.73 GB of storage, covering synthetic, epidemiological, exchange-rate, ETFs (Exchange-Traded Funds), and stock-market panel time series.

### 4.1 Synthetic Data

The synthetic component of PanelTS provides controlled panel-based time series with known target–covariate relationships. This is to facilitate targeted studies as the dataset is fully adjustable. Each

[1]Details are provided in **App. A**. Panel-data analysis is optional and secondary, and is not the focus of this benchmark.

Table 2: Overview of the five dataset groups in PanelTS.

| Dataset group | # Units | Target variable | # Covariates | Description |
|---|---|---|---|---|
| I. Synthetic data | 5<br>10<br>20 | $y$ | $\{5, 10, 20\}$[a] | Synthetic panel-based time series |
| II. Currency exchange rates | 16 | Adjusted close price | 19 | Exchange rates relative to the U.S. dollar |
| III. Exchange-traded funds | 88 | Adjusted close price | 6 | Commodity ETFs |
| IV. COVID-19 | 20<br>79<br>320 | Daily new cases | 17<br>17<br>4 | Aggregated regions by LGAs[b]<br>LGAs<br>Aggregated regions by postcodes[c] |
| V. Stock market | 503<br>82<br>62 | Adjusted close price | 112<br>112<br>112 | S&P 500 stocks<br>S&P 500 Information Technology stocks<br>S&P 500 Health Care stocks |

[a] For the synthetic datasets, we consider all combinations of the number of units and covariates, with both selected from $\{5, 10, 20\}$, resulting in nine synthetic configurations in total. [b] LGAs denote local government areas, the administrative units used by the Victorian Government for regional reporting. [c] Postcodes denote postal-code areas used for mail delivery and small-area geographic reporting by the Victorian Government.

synthetic dataset contains $N$ units observed over a common timeline. For unit $i$ at time $t$, the target value is generated from a covariate-driven additive process:

$$y_{i,t} = c_i + \sum_{m=1}^{M} \phi_m x_{i,m,t} + e_{i,t}, \tag{1}$$

where $c_i$ captures cross-unit dynamics, $x_{i,m,t}$ denotes the $m$-th covariate, $\phi_m$ controls its contribution, and $e_{i,t}$ is a stochastic residual. Each covariate is generated by combining a piecewise trend component, a seasonal component, and a stochastic remainder component. This follows the standard decomposition view Hyndman & Athanasopoulos (2018) and related decomposition methods Bandara et al. (2025); Yang et al. (2021), while allowing the generated covariates to exhibit nonlinear temporal behavior through structural changes, periodic variation, and skewed stochastic fluctuations:

$$x_{i,m,t} = w_g g_{i,m,t} + w_s f_{i,m,t} + w_r r_{i,m,t}, \tag{2}$$

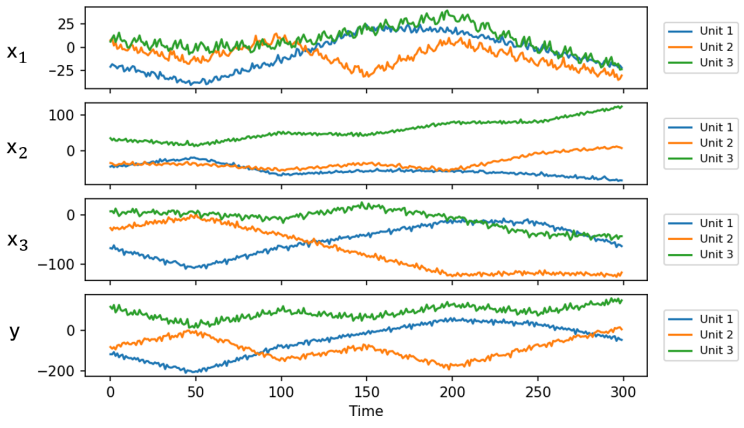



Figure 2: Each unit contains three covariate time series ($x_1$–$x_3$) and a corresponding target series $y$.

where $g_{i,m,t}$, $f_{i,m,t}$, and $r_{i,m,t}$ denote trend, seasonal, and remainder components, respectively, and $w_g + w_s + w_r = 1$. The trend component is generated from piecewise linear segments Chen et al. (2025), the seasonal component from Fourier terms Bloomfield (2004), and the remainder from a Gamma distribution. Although the target is constructed through an additive covariate relationship, its temporal behavior can be non-stationary and nonlinear over time because the covariates contain structural changes, seasonal variation, and stochastic fluctuations. This design provides a controlled setting in which the number of units, number of covariates, and temporal patterns can be adjusted

as data-generating process is fully known. Detailed equations, parameter settings, and additional examples are in **App. B**. Also examples of generated synthetic covariates are shown in App. B.

Figure 2 is an example of the synthetic panel structure. Each unit is associated with three covariate time series ($x_1$–$x_3$) and a target series $y$, where $y$ is generated according to Eq. equation 1. We construct nine synthetic configurations by combining the number of units and the number of covariates, with $(N, M) \in \{5, 10, 20\} \times \{5, 10, 20\}$. Each generated dataset contains 100,000 time points. This length is chosen for practical use and is not a limitation of the generator; $N$ and $M$ can be adjusted according to the application or computational budget. Additional details are provided in **App. C**.

### 4.2 Currency Exchange Rates

This dataset contains daily exchange rates against the U.S. dollar for 16 non-USD currencies associated with G20 economies. Each currency is treated as an observational unit. The sample covers the period from November 11, 2019 to October 31, 2024. The target variable is the adjusted closing exchange rate. For each currency, we also provide 19 macroeconomic covariates describing the economic and financial conditions of the corresponding country or region. These covariates include indicators related to economic activity, prices, monetary conditions, trade, and external balances. The data are collected from public sources including World Bank Open Data, the International Monetary Fund, and Yahoo Finance. Further details are provided in **App. D**.

### 4.3 Exchange-Traded Fund Data

This dataset contains 88 commodity exchange-traded funds (ETFs) traded in the United States, with each ETF treated as an observational unit. We collect daily market data from April 17, 2015 to April 15, 2025. The target variable is the adjusted closing price. For each ETF, we provide six market covariates: opening price, high price, low price, closing price, percentage change, and trading volume. These variables describe daily price movements and trading activity, and can be used as lagged covariates in forecasting tasks to avoid using information unavailable at the prediction time. The data are collected from public financial data sources such as Yahoo Finance. Further details are provided in **App. E**.

### 4.4 COVID-19 Data

This dataset focuses on regional COVID-19 case dynamics in Victoria, Australia. Regions are treated as observational units, and the target variable is the number of daily reported new cases. We collect daily case data from the Victorian Government for the period from January 1, 2022 to September 16, 2022.[2] The original data are reported for 79 local government areas (LGAs) and 704 postcode areas. To support analysis at different spatial granularities, we construct three panel-based datasets with 20, 79, and 320 regional units. The 20-unit setting is obtained by aggregating LGAs into larger regions, the 79-unit setting corresponds to LGAs, and the 320-unit setting is constructed from postcode-level areas. The covariates describe regional demographic, health, and socioeconomic characteristics, including vaccination rates, diabetes prevalence, obesity rates, respiratory disease indicators, and income levels. Regions are aggregated based on geographic adjacency, and covariates are aggregated by summation or averaging depending on their interpretation, such as total population versus prevalence rates. Due to data availability constraints, the full covariate set is not available for all postcode-level units. Further details are provided in **App. F**.

### 4.5 Stock Market Data

This dataset contains three stock-market panels with different market scopes: the full S&P 500 universe with 503 stocks, the Information Technology sector with 82 stocks, and the Health Care sector with 62 stocks. Each stock is treated as an observational unit. For each stock, we collect ten years of daily market data ending on March 21, 2025. The target variable is the adjusted closing price. Daily trading volume is included as a basic market covariate. In addition, we provide 107 financial and market-valuation covariates for each stock, covering balance-sheet indicators, profitability measures,

[2]After September 16, 2022, case numbers were no longer reported daily and are therefore not included in this dataset.

and valuation metrics. These variables describe firm-level characteristics and market conditions, and are intended to support forecasting tasks that incorporate additional covariate information. When used for forecasting, market variables and financial covariates should be aligned according to the forecasting protocol so that only information available at the prediction time is used. The data are collected from public financial data sources such as Yahoo Finance. Detailed variable descriptions are provided in **App. G**.

### 4.6 Task Protocol

All forecasting tasks in PanelTS are constructed in temporal order. Training, validation, and test sets can be split along the time dimension, and each rolling input window only uses observations before the prediction horizon. The target variable and covariates are aligned by the time stamp. For a given forecasting window, users can select the corresponding target and covariate values from the same time range. Models that do not use covariates can be evaluated using the target variable alone.

## 5 Reference Experiments on PanelTS

### 5.1 Evaluation Metrics

**RMSE and MAE.** We report root mean squared error (RMSE) and mean absolute error (MAE) as standard forecasting accuracy metrics. However, when target scales differ substantially across units, directly averaging raw RMSE or MAE can be misleading. For example, in the exchange-rate dataset, the magnitude of the Indonesian Rupiah (IDR) exchange rate is much larger than that of currencies such as the British Pound (GBP), so raw errors for IDR can dominate the average. To reduce scale dominance across units, we also report normalized RMSE and MAE, denoted as $\mathrm{RMSE}^*$ and $\mathrm{MAE}^*$:

$$\mathrm{RMSE}^* = \frac{1}{N}\sum_{i=1}^{N}\frac{\mathrm{RMSE}_i}{\bar{y}_i}, \tag{3}$$

$$\mathrm{MAE}^* = \frac{1}{N}\sum_{i=1}^{N}\frac{\mathrm{MAE}_i}{\bar{y}_i}, \tag{4}$$

where $\mathrm{RMSE}_i$ and $\mathrm{MAE}_i$ denote the error metrics for unit $i$, and $\bar{y}_i$ is the average target value of unit $i$. These normalized metrics are used to make errors more comparable across units with different scales; they are not intended as universally optimal error measures.

**Computational Cost** The reported cost is intended only as a reference runtime under our experimental setup. In panel-based forecasting settings, models may need to process many units, covariates, and rolling evaluation windows, making runtime an important practical consideration.

## 6 Dataset Usability and Reference Results

The purpose of the experiments is to provide reference results and demonstrate the usability of PanelTS with existing forecasting models. We focus on currency exchange rates and COVID-19 as two illustrative scenarios, since they represent different domains, unit counts, covariate structures, and levels of cross-unit dependence.

### 6.1 Baseline Methods

We evaluate representative forecasting methods covering several commonly used modeling paradigms, including classical statistical models, recurrent neural networks, linear neural models, probabilistic models, graph/state-space models, Transformer-based models, and LLM-based forecasting models. The selected methods are ARIMA ArunKumar et al. (2021), VAR Hyndman & Athanasopoulos (2018), ESE Xu et al. (2026), LSTM Feng et al. (2022), DLinear Zeng et al. (2023), NLinear Zeng et al. (2023), Informer Zhou et al. (2021), FiLM Zhou et al. (2022), SCINet Liu et al. (2022a), DeepAR Le Guen & Thome (2020), KVAE Tang & Matteson (2021), TPGNN Liu et al. (2022b), PatchTST Nie et al. (2023), and TimeLLM Jin et al. (2024). These models are included to show that PanelTS can be used with a broad range of existing forecasting pipelines. Several methods, such as VAR, DLinear, Informer, TPGNN, and PatchTST, can be applied to multiple time series or multivariate forecasting settings Qiu et al. (2025).

**Implementation** All experiments use the same input length and prediction horizon within each reported setting. For the reference tables below, we use 100 input steps and a one-step-ahead horizon. Computational cost is measured on a machine equipped with a Ryzen 9 7950X 16-core CPU, 64GB RAM, and an NVIDIA RTX 4090 GPU with 24GB memory.

### 6.2 Results on Representative Scenarios

The reference results are obtained using a rolling-window evaluation protocol. For each evaluation time $t$, models use the previous 100 time steps as input and produce a one-step-ahead forecast for $t+1$. When covariates are used, covariate values are selected from the same input window as the target variable. The window is then moved forward by one time step. For each unit, RMSE and MAE are computed over all rolling forecasts, and the reported values are averaged across units. For exchange rates, $\mathrm{RMSE}^*$ and $\mathrm{MAE}^*$ are additionally computed by normalizing each unit-level error by the average target value of that unit.

Table 3: Reference forecasting results on the currency exchange-rate and COVID-19 datasets. All results use 100 input steps and a one-step-ahead horizon. For exchange rates, $\mathrm{RMSE}^*$ and $\mathrm{MAE}^*$ denote normalized errors scaled by 100. Cost is reported in minutes.

| Model | Exchange rates, 16 units | | | COVID-19, 20 units | | | COVID-19, 79 units | | | COVID-19, 320 units | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\mathrm{RMSE}^*$ | $\mathrm{MAE}^*$ | Cost | RMSE | MAE | Cost | RMSE | MAE | Cost | RMSE | MAE | Cost |
| ARIMA | 1.280 | 0.629 | 0.19 | 79.70 | 79.59 | 0.33 | 77.72 | 67.41 | 1.31 | 7.44 | 6.24 | 5.07 |
| VAR | 1.248 | 0.653 | – | 95.79 | 90.86 | – | 90.85 | 81.74 | – | 8.07 | 7.68 | – |
| ESE | 1.183 | 0.600 | 0.22 | 84.54 | 83.91 | 1.31 | 55.54 | 50.85 | 2.10 | 4.83 | 4.58 | 2.23 |
| LSTM | 1.218 | 0.604 | 5.71 | 78.45 | 74.96 | 9.23 | 61.42 | 57.26 | 39.96 | 6.23 | 5.47 | 149.26 |
| DLinear | 1.212 | 0.603 | 6.33 | 79.47 | 72.41 | 10.40 | 56.16 | 53.95 | 40.98 | 5.93 | 5.10 | 163.57 |
| NLinear | 1.214 | 0.608 | 6.71 | 74.96 | 70.77 | 11.56 | 55.62 | 52.23 | 45.67 | 5.47 | 4.99 | 160.09 |
| Informer | 1.219 | 0.608 | 3.58 | 76.84 | 71.79 | 5.95 | 62.01 | 60.57 | 26.97 | 6.05 | 5.80 | 100.08 |
| FiLM | 1.218 | 0.612 | 6.69 | 76.84 | 73.85 | 11.70 | 56.21 | 51.34 | 48.55 | 5.74 | 4.96 | 181.06 |
| SCINet | 1.184 | 0.603 | 8.82 | 79.75 | 69.94 | 14.18 | 55.86 | 53.56 | 62.27 | 5.94 | 5.59 | 206.06 |
| DeepAR | 1.219 | 0.610 | 5.19 | 83.41 | 74.65 | 9.22 | 56.34 | 54.32 | 37.25 | 6.29 | 5.04 | 130.67 |
| KVAE | 1.204 | 0.604 | 4.39 | 70.85 | 64.44 | 6.80 | 53.12 | 52.71 | 32.41 | 5.29 | 4.94 | 109.62 |
| TPGNN | 1.182 | 0.601 | 6.75 | 78.37 | 69.86 | 10.58 | 59.22 | 58.13 | 42.92 | 6.06 | 5.41 | 158.84 |
| PatchTST | 1.209 | 0.607 | 4.32 | 70.01 | 63.52 | 6.92 | 56.44 | 52.41 | 27.70 | 5.11 | 4.43 | 109.89 |
| TimeLLM | 1.182 | 0.603 | 4.53 | 73.47 | 70.02 | 6.20 | 59.48 | 52.64 | 24.15 | 5.12 | 4.82 | 107.73 |

Most conventional forecasting baselines are not specifically designed for the panel-based setting considered in PanelTS, where multiple related units share covariate definitions. We therefore evaluate these baselines under their standard target-history setting, while covariate-aware panel-based modeling is represented by ESE Xu et al. (2026).

Directly adding shared covariates to standard univariate or multivariate methods is possible, but it does not necessarily provide the most appropriate configuration for these models. As a representative check, **App. H** reports SCINet results with and without covariates in a multivariate setting. The results show that adding covariates improves the multivariate variant, but its accuracy remains substantially worse than repeated unit-wise prediction. Therefore, the main reference table reports conventional baselines under their standard target-history setting.

Table 3 summarizes the reference results on the currency exchange-rate and COVID-19 datasets. The two scenarios illustrate different uses of PanelTS: exchange rates contain financial units with macroeconomic covariates, while COVID-19 contains regional units evaluated at multiple spatial granularities. For exchange rates, TPGNN, TimeLLM, and ESE are among the strongest methods under the normalized metrics reported in the main table. The normalized metrics are used because exchange-rate scales differ substantially across currencies.

For COVID-19, the best-performing method varies across spatial granularities. PatchTST obtains the lowest RMSE and MAE in the 20-unit setting. In the 79-unit setting, KVAE obtains the lowest RMSE, while ESE obtains the lowest MAE. In the 320-unit setting, ESE obtains the lowest RMSE and runtime, while PatchTST obtains the lowest MAE. For runtime, ARIMA is the fastest method in

the exchange-rate, 20-unit COVID-19, and 79-unit COVID-19 settings, whereas ESE is the fastest in the 320-unit COVID-19 setting.

Overall, these results show that PanelTS can be used with a broad range of forecasting models across different domains, unit counts, and spatial granularities. We report a compact set of reference results in the main text for readability, with additional baseline results and settings provided in **App. I**.

## 7 DISCUSSION

PanelTS is motivated by the panel structure underlying many real-world forecasting problems. Time series are often noisy, non-stationary, and subject to structural breaks or shifts in the data-generating process Aue & Horváth (2013); Han et al. (2019); Hyndman & Rostami-Tabar (2025); Kashpruk et al. (2023); Vela et al. (2022). They may also contain missing values and outliers, which can further reduce forecasting stability and reliability Hyndman & Athanasopoulos (2018); Seliem (2022). In practice, these challenges frequently arise across groups of related units rather than within isolated series. For example, disease incidence across neighboring regions may share demographic or healthcare-related factors, while financial assets are often jointly influenced by broader market conditions.

Each unit in the PanelTS datasets is stored together with its target variable and covariates, preserving information about unit-specific characteristics, shared covariate patterns, and cross-unit relationships. This structure is also valuable for evaluating computational efficiency and scalability when forecasting large collections of related units jointly Li et al. (2019); Torres et al. (2021); Waqas & Humphries (2024). As a result, PanelTS more closely reflects real-world forecasting settings, where both temporal dynamics and inter-unit relationships play important roles. Nevertheless, two limitations remain:

**Limited panel-specific baselines.** The reference experiments mainly use models developed for univariate or multivariate time series forecasting. PanelTS is compatible with existing forecasting pipelines, but most of these models do not explicitly use panel-specific information, such as cross-unit dynamics, shared covariates, or cross-unit dependence. The reported results should therefore be treated as reference points rather than a complete evaluation of panel-based forecasting methods.

**Domain coverage.** PanelTS currently includes five forecasting scenarios: synthetic data, COVID-19 dynamics, exchange-traded funds, currency exchange rates, and stock markets. These scenarios provide diversity in domains, numbers of units, and covariate structures, but they are not exhaustive. Future versions could expand to additional panel-based forecasting, such as influenza surveillance, animal disease outbreaks, energy systems, transportation networks, and derivative markets.

Overall, PanelTS complements existing time-series datasets by preserving the unit dimension and organizing related units together with their associated covariates. It provides a reusable benchmark for studying temporal dynamics, cross-unit dependencies, and shared structural patterns across units.

## 8 CONCLUSION

We introduced PanelTS, a multi-domain dataset for panel-based time-series forecasting. Many forecasting problems involve related units observed over time, where the data contain not only the history of each unit, but also unit-specific differences, shared covariates, and potential cross-unit relationships. A panel-based organization is well suited to this setting because it preserves these sources of information, rather than collapsing them into either independent series or a single multivariate sequence. PanelTS follows this design and supports target-only, covariate-aware, and panel-based forecasting tasks from the same data source. The current release includes 1,844 synthetic and real-world datasets covering COVID-19 dynamics, exchange-traded funds, currency exchange rates, and stock markets, together with task construction protocols and reference benchmark results.

**Future Work** We will extend PanelTS in two directions. First, we plan to include more forecasting methods designed for panel-based settings. Second, we will expand the dataset to cover more application domains and data sources.

## AI USE STATEMENT

Generative AI tools were used only for grammar checking and language polishing. They were not used for dataset construction, methodology development, experimental design, code generation, result analysis, or scientific conclusions. All AI-assisted edits were reviewed and approved by the authors, who take full responsibility for the final manuscript.

## ETHICS STATEMENT

This work adheres to the ICLR Code of Ethics. No human subjects or animal experiments were involved. PanelTS is constructed from synthetic data and publicly available real-world datasets covering epidemiology and financial markets. The COVID-19 data are provided at aggregated regional levels, and no personally identifiable information is included. All source datasets are used in accordance with their respective access, licensing, and usage conditions. Dataset sources and relevant documentation are provided to support transparent and responsible use of the benchmark.

## REPRODUCIBILITY STATEMENT

We have made every effort to ensure that PanelTS can be reproduced and used consistently. The dataset, preprocessing procedures, benchmark code, and documentation are provided through the accompanying repository. PanelTS uses a unified panel representation across its synthetic and real-world datasets, with standardized task definitions, input lengths, forecasting horizons, and evaluation protocols.

## REFERENCES

Sercan Arik, Chun-Liang Li, Jinsung Yoon, Rajarishi Sinha, Arkady Epshteyn, Long Le, Vikas Menon, Shashank Singh, Leyou Zhang, Martin Nikoltchev, et al. Interpretable sequence learning for covid-19 forecasting. *Advances in neural information processing systems*, 33:18807–18818, 2020.

K. E. ArunKumar, Dinesh V. Kalaga, Ch. Mohan Sai Kumar, Govinda Chilkoor, Masahiro Kawaji, and Timothy M. Brenza. Forecasting the dynamics of cumulative covid-19 cases (confirmed, recovered and deaths) for top-16 countries using statistical machine learning models: Auto-regressive integrated moving average (arima) and seasonal auto-regressive integrated moving average (sarima). *Applied Soft Computing*, 103:107161 – 107161, 2021.

Alexander Aue and Lajos Horváth. Structural breaks in time series. *Journal of Time Series Analysis*, 34(1):1–16, 2013.

Badi H. Baltagi. *Econometric Analysis of Panel Data*. John Wiley & Sons, 2008.

Kasun Bandara, Rob J Hyndman, and Christoph Bergmeir. Mstl: A seasonal-trend decomposition algorithm for time series with multiple seasonal patterns. *International Journal of Operational Research*, 52(1):79–98, 2025.

Peter Bloomfield. *Fourier analysis of time series: an introduction*. John Wiley & Sons, 2004.

Zvi Bodie and Alex Kane. Investments. 2020.

Centers for Disease Control and Prevention. U.S. Influenza Surveillance System: ILINet data. `https://www.cdc.gov/fluview/overview/fluview-interactive.html?utm_source=chatgpt.com`, 2025. Accessed: 2025-01-xx.

Li Chen, Jiti Gao, and Farshid Vahid. *Nonlinear Trending Time Series: Theory and Practice*. World Scientific, 2025.

Xi Chen, Chuan Qin, Chuyu Fang, Chao Wang, Fuzhen Zhuang, Hengshu Zhu, Hui Xiong, et al. Job-sdf: A multi-granularity dataset for job skill demand forecasting and benchmarking. *Advances in Neural Information Processing Systems*, 37:129329–129356, 2024.

Patrick Emami, Abhijeet Sahu, and Peter Graf. Buildingsbench: A large-scale dataset of 900k buildings and benchmark for short-term load forecasting. *Advances in Neural Information Processing Systems*, 36:19823–19857, 2023.

Jessica Embury, Ming-Hsiang Tsou, Atsushi Nara, and Eyal Oren. A spatio-demographic perspective on the role of social determinants of health and chronic disease in determining a population's vulnerability to covid-19. *Preventing chronic disease*, 19:E38, 2022.

Eugene F Fama and Kenneth R French. A five-factor asset pricing model. *Journal of financial economics*, 116(1):1–22, 2015.

Lin Feng, Ziren Chen, Harold A Lay Jr, Khaled M. Furati, and Abdul-Qayyum M. Khaliq. Data driven time-varying seir-lstm/gru algorithms to track the spread of covid-19. *Mathematical biosciences and engineering : MBE*, 19 9:8935–8962, 2022.

Jiti Gao. *Nonlinear time series: semiparametric and nonparametric methods*. Chapman and Hall/CRC, 2007.

Jiti Gao, Bin Peng, and Yayi Yan. Higher-order expansions and inference for panel data models. *Journal of the American Statistical Association*, 119(548):2760–2771, 2024.

Zhaoxing Gao and Ruey S Tsay. A structural-factor approach to modeling high-dimensional time series and space-time data. *Journal of Time Series Analysis*, 40(3):343–362, 2019.

Groningen Growth and Development Centre. Penn World Table version 10.01, 2023. URL `https://doi.org/10.34894/QT5BCC`.

Shengnan Guo, Youfang Lin, Ning Feng, Chao Song, and Huaiyu Wan. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pp. 922–929, 2019.

Zhongyang Han, Jun Zhao, Henry Leung, King Fai Ma, and Wei Wang. A review of deep learning models for time series prediction. *IEEE Sensors Journal*, 21(6):7833–7848, 2019.

Georges Hebrail and Alice Berard. Individual household electric power consumption data set. *UCI Machine Learning Repository*, 2012.

Kewei Hou. Industry information diffusion and the lead-lag effect in stock returns. *The review of financial studies*, 20(4):1113–1138, 2007.

Cheng Hsiao. *Analysis of panel data*. Number 64. Cambridge university press, 2022.

Rob J Hyndman and George Athanasopoulos. *Forecasting: principles and practice*. OTexts, 2018.

Rob J Hyndman and Bahman Rostami-Tabar. Forecasting interrupted time series. *Journal of the Operational Research Society*, 76(4):790–803, 2025.

S&P Dow Jones Indices. S&p us indices methodology, 2023.

Ming Jin, Shiyu Wang, Lintao Ma, Zhixuan Chu, James Y Zhang, Xiaoming Shi, Pin-Yu Chen, Yuxuan Liang, Yuan-Fang Li, Shirui Pan, and Qingsong Wen. Time-LLM: Time series forecasting by reprogramming large language models. In *International Conference on Learning Representations (ICLR)*, 2024.

Nataliia Kashpruk, Cezary Piskor-Ignatowicz, and Jerzy Baranowski. Time series prediction in industry 4.0: a comprehensive review and prospects for future advancements. *Applied Sciences*, 13 (22):12374, 2023.

Olaf Kolle. Documentation of the weather station on top of the roof of the institute building of the max-planck-institute for biogeochemistry. *Recuperado de https://www. bgc-jena. mpg. de/wetter*, 2025.

Vincent Le Guen and Nicolas Thome. Probabilistic time series forecasting with shape and temporal diversity. *Advances in Neural Information Processing Systems*, 33:4427–4440, 2020.

Shiyang Li, Xiaoyong Jin, Yao Xuan, Xiyou Zhou, Wenhu Chen, Yu-Xiang Wang, and Xifeng Yan. Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting. *Advances in neural information processing systems*, 32, 2019.

Minhao Liu, Ailing Zeng, Muxi Chen, Zhijian Xu, Qiuxia Lai, Lingna Ma, and Qiang Xu. Scinet: Time series modeling and forecasting with sample convolution and interaction. *Advances in Neural Information Processing Systems*, 35:5816–5828, 2022a.

Xu Liu, Yutong Xia, Yuxuan Liang, Junfeng Hu, Yiwei Wang, Lei Bai, Chao Huang, Zhenguang Liu, Bryan Hooi, and Roger Zimmermann. Largest: A benchmark dataset for large-scale traffic forecasting. *Advances in Neural Information Processing Systems*, 36:75354–75371, 2023.

Yijing Liu, Qinxian Liu, Jian-Wei Zhang, Haozhe Feng, Zhongwei Wang, Zihan Zhou, and Wei Chen. Multivariate time-series forecasting with temporal polynomial graph neural networks. *Advances in neural information processing systems*, 35:19414–19426, 2022b.

Spyros Makridakis, Evangelos Spiliotis, and Vassilios Assimakopoulos. The m4 competition: Results, findings, conclusion and way forward. *International Journal of forecasting*, 34(4):802–808, 2018.

Spyros Makridakis, Evangelos Spiliotis, and Vassilios Assimakopoulos. The m5 competition: Background, organization, and implementation. *International Journal of Forecasting*, 38(4): 1325–1336, 2022.

Mihir Mehta, Juxihong Julaiti, Paul Griffin, and Soundar Kumara. Early stage machine learning–based prediction of us county vulnerability to the covid-19 pandemic: machine learning approach. *JMIR public health and surveillance*, 6(3):e19446, 2020.

Yukihiko Nakata, Philipp Getto, Anna Marciniak-Czochra, and Tomás Alarcón. Stability analysis of multi-compartment models for cell production systems. *Journal of biological dynamics*, 6(sup1): 2–18, 2012.

Yuqi Nie, Nam H. Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. In *International Conference on Learning Representations*, 2023.

Daniel Pastorek. Euro area uncertainty and euro exchange rate volatility: Exploring the role of transnational economic policy. *Finance Research Letters*, 58:104351, 2023. ISSN 1544-6123. doi: https://doi.org/10.1016/j.frl.2023.104351. URL https://www.sciencedirect.com/science/article/pii/S1544612323007237.

Xiangfei Qiu, Hanyin Cheng, Xingjian Wu, Jilin Hu, Chenjuan Guo, and Bin Yang. A comprehensive survey of deep learning for multivariate time series forecasting: A channel strategy perspective. *arXiv preprint arXiv:2502.10721*, 2025.

Aldo Rescigno. Synthesis of a multicompartmented biological model. *Biochimica et Biophysica Acta*, 37(3):463–468, 1960.

Muhammad Metwally Seliem. Handlingoutlier data as missing values by imputation methods: application of machine learning algorithms. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 13(1):273–286, 2022.

Social Research Center. Panel study of income dynamics, public use dataset [restricted use data, if appropriate].

Binh Tang and David S Matteson. Probabilistic transformer for time series analysis. *Advances in Neural Information Processing Systems*, 34:23592–23608, 2021.

José F Torres, Dalil Hadjout, Abderrazak Sebaa, Francisco Martínez-Álvarez, and Alicia Troncoso. Deep learning for time series forecasting: a survey. *Big data*, 9(1):3–21, 2021.

Daniel Vela, Andrew Sharp, Richard Zhang, Trang Nguyen, An Hoang, and Oleg S Pianykh. Temporal quality degradation in ai models. *Scientific reports*, 12(1):11654, 2022.

Muhammad Waqas and Usa Wannasingha Humphries. A critical review of rnn and lstm variants in hydrological time series predictions. *MethodsX*, pp. 102946, 2024.

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

Beinan Xu, Andy Song, Jiti Gao, and Feng Liu. Once-for-all: Scalable simultaneous forecasting via equilibrium state estimation, 2026. URL https://doi.org/10.5281/zenodo.20050422. The 43rd International Conference on Machine Learning.

Linxiao Yang, Qingsong Wen, Bo Yang, and Liang Sun. A robust and efficient multi-scale seasonal-trend decomposition. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5085–5089. IEEE, 2021.

Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. Are transformers effective for time series forecasting? In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pp. 11121–11128, 2023.

Jiawen Zhang, Xumeng Wen, Zhenwei Zhang, Shun Zheng, Jia Li, and Jiang Bian. Probts: Benchmarking point and distributional forecasting across diverse prediction horizons. *Advances in Neural Information Processing Systems*, 37:48045–48082, 2024.

Liang Zhao, Olga Gkountouna, and Dieter Pfoser. Spatial auto-regressive dependency interpretable learning based on spatial topological constraints. *ACM Transactions on Spatial Algorithms and Systems (TSAS)*, 5(3):1–28, 2019.

Xiangtian Zheng, Nan Xu, Loc Trinh, Dongqi Wu, Tong Huang, S Sivaranjani, Yan Liu, and Le Xie. A multi-scale time-series dataset with benchmark for machine learning in decarbonized energy grids. *Scientific Data*, 9(1):359, 2022.

Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pp. 11106–11115, 2021.

Tian Zhou, Ziqing Ma, Qingsong Wen, Liang Sun, Tao Yao, Wotao Yin, Rong Jin, et al. Film: Frequency improved legendre memory model for long-term time series forecasting. *Advances in Neural Information Processing Systems*, 35:12677–12690, 2022.

# A Illustrative Examples of Panel Data Analysis

Although the primary function of PanelTS is to support time series forecasting, its panel data organization also permits the application of panel data analysis tools. We provide brief illustrative examples to demonstrate how conventional panel regression models can be applied to PanelTS.

A simple pooled regression treats all observations across units and time as a single dataset, ignoring unit-specific dynamics. For example, a pooled linear regression model can be written as

$$y_{i,t} = \beta_0 + \boldsymbol{\beta}^\top \mathbf{x}_{i,t} + \varepsilon_{i,t}, \tag{5}$$

where $y_{i,t}$ denotes the target variable of unit $i$ at time $t$, $\mathbf{x}_{i,t}$ represents the corresponding covariates, and $\varepsilon_{i,t}$ is an error term. This formulation provides a simple baseline for exploring relationships between covariates and the target variable across all units.

To account for unobserved, time-invariant difference across units, a fixed-effects model can be applied. In this setting, the regression model takes the form

$$y_{i,t} = \alpha_i + \boldsymbol{\beta}^\top \mathbf{x}_{i,t} + \varepsilon_{i,t}, \tag{6}$$

where $\alpha_i$ denotes a unit-specific intercept that captures time invariant characteristics of unit $i$. By controlling for these fixed effects, the model focuses on within-unit variation over time.

# B Synthetic Data Generation

The synthetic data are designed to follow the panel-based organization used throughout PanelTS. Each synthetic dataset contains $N$ units observed on a common time axis. For each unit, the target series is generated from unit-specific intercepts, covariates, and stochastic residuals. The covariates are constructed from trend, seasonal, and remainder components, following the standard time-series decomposition view Hyndman & Athanasopoulos (2018) and related decomposition methods Bandara et al. (2025); Yang et al. (2021). This construction allows the generated covariates, and therefore the target series, to exhibit structural changes, periodic variation, and skewed stochastic fluctuations.

## B.1 Trend Component

The trend component represents the long-term movement of a covariate series. Real-world trends rarely follow a single global linear trajectory; they often contain structural changes or local shifts. We therefore generate the trend using a piecewise linear function Chen et al. (2025). Let the time axis be divided into $P$ consecutive intervals

$$I_p = \{t \mid \tau_{p-1} \le t < \tau_p\}, \quad p = 1, \ldots, P.$$

For unit $i$ and covariate $m$, the trend component is defined as

$$g_{i,m,t} = \sum_{p=1}^{P} \mathbf{1}_{I_p}(t) \left[ g_{i,m,\tau_{p-1}} + \sigma_{i,m,p}(t - \tau_{p-1}) \right], \tag{7}$$

where $\mathbf{1}_{I_p}(t)$ equals 1 if $t \in I_p$ and 0 otherwise. The slope $\sigma_{i,m,p}$ controls the local trend within interval $I_p$, and $g_{i,m,\tau_{p-1}}$ ensures continuity across adjacent intervals. The resulting trend is linear within each segment but can exhibit nonlinear behavior over the full time span.

## B.2 Seasonal Component

The seasonal component captures repeated temporal patterns. We generate seasonality using Fourier terms Bloomfield (2004):

$$f_{i,m,t} = \sum_{k=1}^{K} \left[ \theta_{i,m,k} \cos\left(\frac{2\pi k t}{S}\right) + \psi_{i,m,k} \sin\left(\frac{2\pi k t}{S}\right) \right], \tag{8}$$

where $K$ is the number of harmonics, $S$ is the seasonal period, and $\theta_{i,m,k}$ and $\psi_{i,m,k}$ control the amplitude and phase of the $k$-th harmonic. When $K = 1$, the seasonal component reduces to a simple sinusoidal pattern. Larger values of $K$ allow more complex seasonal shapes, including multi-peak or asymmetric patterns.

### B.3 Remainder Component

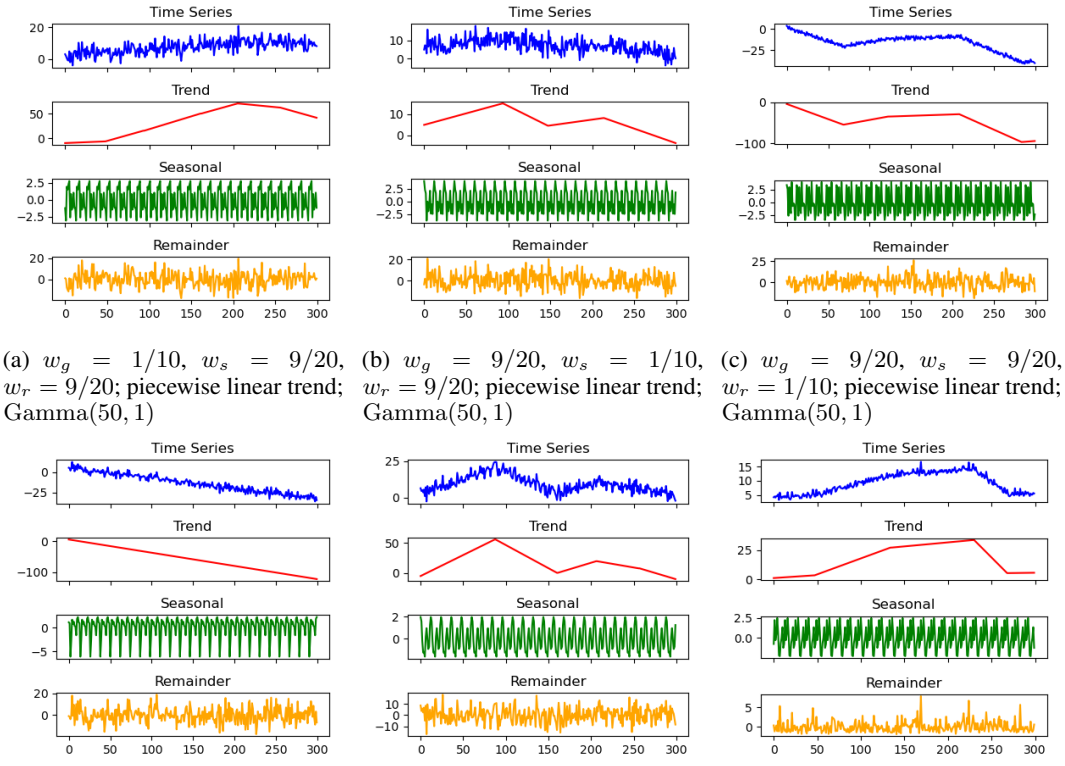
The remainder component represents variation not explained by trend or seasonality and is commonly treated as noise Gao & Tsay (2019); Hyndman & Athanasopoulos (2018). We sample the remainder from a Gamma distribution:

$$r_{i,m,t} \sim \text{Gamma}(\alpha, \beta_t), \tag{9}$$

where $\alpha$ is the shape parameter and $\beta_t$ is the scale parameter. By default, $\beta_t = \beta_0$, corresponding to i.i.d. noise. To allow non-i.i.d. stochastic fluctuations, the scale parameter can vary periodically:

$$\beta_t = \beta_0 \left(1 + \delta \sin\left(\frac{2\pi t}{S_r}\right)\right), \tag{10}$$

where $\delta$ controls the magnitude of the periodic fluctuation and $S_r$ controls its period. The Gamma distribution is used because it can represent skewness and long-tailed variation. When the shape parameter $\alpha$ is large, the Gamma distribution approaches a Gaussian-like shape. The residual term $e_{i,t}$ in Eq. equation 1 is generated using the same distributional form as the remainder component.




(a) $w_g = 1/10$, $w_s = 9/20$, $w_r = 9/20$; piecewise linear trend; $\text{Gamma}(50, 1)$

(b) $w_g = 9/20$, $w_s = 1/10$, $w_r = 9/20$; piecewise linear trend; $\text{Gamma}(50, 1)$

(c) $w_g = 9/20$, $w_s = 9/20$, $w_r = 1/10$; piecewise linear trend; $\text{Gamma}(50, 1)$

(d) $w_g = w_s = w_r = 1/3$; linear trend; $\text{Gamma}(50, 1)$

(e) $w_g = w_s = w_r = 1/3$; piecewise linear trend; $\text{Gamma}(50, 1)$

(f) $w_g = w_s = w_r = 1/3$; piecewise linear trend; $\text{Gamma}(2, 1)$

Figure 3: Examples of synthetic covariate time series generated under different parameter configurations. Each panel shows the generated covariate series together with its trend, seasonal, and remainder components. The examples illustrate how component weights, trend forms, and noise distributions affect the resulting temporal patterns.

Figure 3 illustrates representative covariate series generated by Eq. equation 2. (a)–(c) use piecewise linear trends with different weights assigned to the trend, seasonal, and remainder components. (d)–(f) use equal component weights, while varying the trend form and the Gamma noise parameters. In (a)–(e), the remainder component is generated from $\text{Gamma}(50, 1)$, which gives a nearly symmetric noise pattern. (f) uses $\text{Gamma}(2, 1)$, producing more visibly skewed stochastic fluctuations. These examples show that the generator can produce covariates with different trend strengths, seasonal patterns, and noise characteristics.

### B.4 SYNTHETIC CONFIGURATIONS

For the synthetic component of PanelTS, the number of units $N$ and the number of covariates $M$ are both selected from

$$\{5, 10, 20\}.$$

We consider all combinations of $N$ and $M$, resulting in nine synthetic configurations. These configurations vary the panel size and covariate dimension while keeping the data-generating process controlled and known.

## C DESCRIPTION OF SYNTHETIC DATA

Table 4: Synthetic data description for 5 units with 5 covariate time series.

| Unit | Target Variable | Description | Type |
|---|---|---|---|
| 1 | $y_1$ | $\mathbf{x}_1 = (x_{1,1}, x_{1,2} \ldots, x_{1,5})$ | Float |
| 2 | $y_2$ | $\mathbf{x}_2 = (x_{2,1}, x_{2,2} \ldots, x_{2,5})$ | Float |
| 3 | $y_3$ | $\mathbf{x}_3 = (x_{3,1}, x_{3,2} \ldots, x_{3,5})$ | Float |
| 4 | $y_4$ | $\mathbf{x}_4 = (x_{4,1}, x_{4,2} \ldots, x_{4,5})$ | Float |
| 5 | $y_5$ | $\mathbf{x}_5 = (x_{5,1}, x_{5,2} \ldots, x_{5,5})$ | Float |

Table 4 illustrates an example of a synthetic dataset in a panel data organization with five units and five covariate time series per unit. Each unit is associated with a target variable $y_i$ and a vector of covariate time series $\mathbf{x}_i = (x_{i,1}, x_{i,2}, \ldots, x_{i,5})$, where all variables are observed over a common time span. This structure is consistent across all synthetic datasets.

In addition to the configuration shown in Table 4, we construct multiple synthetic datasets by varying both the number of units and the number of covariates. Specifically, we consider combinations of $(N, M) \in \{5, 10, 20\} \times \{5, 10, 20\}$, resulting in nine synthetic datasets, denoted as 5–5, 5–10, 5–20, 10–5, 10–10, 10–20, 20–5, 20–10, and 20–20. Each synthetic dataset contains 100000 time points, ensuring a long temporal span for evaluating forecasting methods under different dimensional and cross-sectional configurations.

The number of time points in the released datasets is chosen for practical reasons related to data storage and distribution. This choice does not reflect a limitation of the synthetic data generation framework, which can be readily applied to produce time series of greater length when needed.

## D DESCRIPTION OF EXCHANGE RATE DATA

Exchange rates represent the value of one country's currency relative to another and are among the most widely studied time series in financial analysis and forecasting Bodie & Kane (2020).

The G20 includes 19 major countries and 17 major currencies, with Table 5 displaying 16 currencies and their corresponding countries or organizations. Excluding the US dollar, as it is the base currency for all other currencies. We collected daily exchange rates based on the US dollar for each currency from November. 11, 2019 to October. 31, 2024, for a period of 5 years.

Table 6 describes the six columns of collected exchange rate data, including, the date, the opening price, highest price, lowest price, closing price, and the adjusted closing price. The last column, 'Adj Clos', is used as the target.

Table 7 list the economic and financial indicators for G20 countries for the same time periods of exchange rate data. It includes data such as the population, GDP, external debt, gold reserves, current account balance, inflation rate, trade figures (imports and exports), foreign direct investment, and various monetary supply measures. Additionally, we include tourist arrivals, food inflation, and the consumer price index (CPI) as covariates. It should be noted that the covariate data related to the Euro is from the European Union rather than individual countries such as France or Germany. The data is sourced from World Bank Open Data and the International Monetary Fund.

Table 5: G20 non-USD currencies and their issuing countries.

| Code | Country or Region | Currency |
|---|---|---|
| ARS | Argentina | Argentine Peso |
| AUD | Australia | Australian Dollar |
| BRL | Brazil | Brazilian Real |
| CAD | Canada | Canadian Dollar |
| CNY | China | Renminbi (Yuan) |
| EUR | France; Germany; Italy; European Union | Euro (EUR) |
| GBP | United Kingdom | Pound Sterling |
| INR | India | Indian Rupee |
| IDR | Indonesia | Indonesian Rupiah |
| JPY | Japan | Japanese Yen |
| KRW | South Korea | South Korean Won |
| MXN | Mexico | Mexican Peso |
| RUB | Russia | Russian Ruble |
| SAR | Saudi Arabia | Saudi Riyal |
| TRY | Turkey | Turkish Lira |
| ZAR | South Africa | South African Rand |

Table 6: Six columns of exchange rate daily data

| Variable | Description | Data Type |
|---|---|---|
| Date | Trading date | Date |
| Open | Opening price of the day | Float |
| High | Highest price of the day | Float |
| Low | Lowest price of the day | Float |
| Close | Closing price of the day | Float |
| Adj Close | Adjusted closing price | Float |

Table 7: Covariates for 16 countries or regions

| Variable | Description | Data Type |
|---|---|---|
| DateTime | Date | Date |
| Population (million) | Total population in millions | Float |
| GDP (billion USD) | Gross Domestic Product in USD | Float |
| External Debt (million USD) | Total external debt in USD | Float |
| Gold Reserves (tonnes) | Total gold reserves in tonnes | Float |
| Current Account (billion USD) | Current account balance in USD | Float |
| Inflation Rate | Annual inflation rate in percentage | Float |
| Import (million USD) | Total imports in USD | Float |
| Export (million USD) | Total exports in USD | Float |
| Foreign Direct Investment (million USD) | Foreign direct investment in USD | Float |
| Tourist Arrivals | Number of tourist arrivals | Float |
| Food Inflation | Food inflation rate in percentage | Float |
| Consumer Price Index CPI | Consumer Price Index | Float |
| Money Supply M1 (billion USD) | Narrow money supply in USD | Float |
| Money Supply M2 (billion USD) | Broad money supply in USD | Float |
| Money Supply M3 (billion USD) | Total money supply in USD | Float |

## E Description of Exchange Traded Fund Daily Data

Commodity exchange-traded funds (ETFs) are financial instruments traded on stock exchanges that track the performance of commodity prices or commodity-related indices. Unlike equities or

exchange rates, commodity ETFs are primarily influenced by fluctuations in underlying commodity markets, which makes them a distinct application scenario in the dataset.

Table 8: Eight columns of exchange traded fund daily data

| Variable | Description | Data Type |
|---|---|---|
| Date | Trading day | Date |
| Open | Opening price of the ETF | Float |
| High | Highest price of the ETF during the day | Float |
| Low | Lowest price of the ETF during the day | Float |
| Close | Closing price of the ETF | Float |
| Adj Close | Adjusted closing price of the ETF | Float |
| Change | Percentage change in closing price | Float |
| Volume | Trading volume in number of shares | Integer |

Table 8 describes the eight columns of daily data collected for all 88 commodity ETFs, spanning from April. 17, 2015, to April. 15, 2025 recent date (ten years) in the dataset. It contains a total of 2,515 observations and includes the following key variables: opening price, high price, low price, closing price, adjusted closing price, percentage change, and trading volume. The data is structured to provide comprehensive daily information on the ETF's price movements and trading activity.

## F Description of COVID-19 Data

Table 9: Six columns of COVID-19 daily data

| Variable | Description | Data Type |
|---|---|---|
| Date | Date | Date |
| Population | Total population | Float |
| Active | Active cases | Float |
| Rate | Infection rate | Float |
| New | New confirmed cases | Float |
| Band | Restriction band | Float |

The dataset contains daily COVID-19 data for the 20, 79, 320 regions throughout 2022, in six columns, including date, population, active cases, rates, new cases and risk band level (see Table 9). Note that in this case, the number of "new cases" is the target. The associated covariates about the regions in Victoria and other epidemic related data are collected from main resources, government agencies, e.g. the health department and the Australian bureau of statistics, as detailed in Table 10. These covariates provide comprehensive health and socioeconomic information across 20 and 79 regions throughout 2022, including population health, vaccination rates, and economic conditions.

## G Description for Stock Market Data

The stock market dataset is based on companies included in the S&P 500 index, which is widely regarded as a representative indicator of the U.S. equity market and global economic conditions. The dataset includes 503 stocks, each treated as an observational unit. Although the composition of the S&P 500 changes over time Indices (2023), the index provides a balanced trade-off between representativeness and scale.

For historical data of stock, spanning from March 24, 2015, to March 25, 2025 (ten years). The dataset includes daily trading information such as opening price, closing price, highest and lowest prices, percentage change, trading volume, with target variable is adjusted close price.

Table 10: covariates for regions (only valid for multiple units of 20 and 79 regions).

| Covariate | Description | Data Type |
|---|---|---|
| Region | Name of the region | String |
| Type 2 Diabetes Ratio (%) | Percentage of diagnosed Type 2 diabetes cases | Float |
| Multiple Chronic Diseases (%) | Percentage with two or more chronic diseases | Float |
| Chronic Heart Disease (%) | Percentage with chronic heart disease | Float |
| Obesity Rate (%) | Percentage of obese individuals | Float |
| Respiratory Diseases (%) | Percentage with respiratory diseases | Float |
| Asthma Diagnosis (%) | Percentage diagnosed with asthma | Float |
| Daily Smoker (%) | Percentage of daily smokers | Float |
| Vaccine Dose 1 (%) | First dose vaccine coverage | Float |
| Vaccine Dose 2 (%) | Second dose vaccine coverage | Float |
| Vaccine Dose 3 (%) | Third dose vaccine coverage | Float |
| Hospital Employment Ratio | Ratio of hospital-employed individuals | Float |
| Median Weekly Income (AUD) | Median weekly household income in AUD | Float |
| Median Age | Median age of the population | Float |

Table 11: Eight columns of stocks daily data for all multiple units: S&P 500 stocks (503 units), S&P 500 information technology stocks (82 units), and S&P 500 health care stocks (62 units).

| Variable | Data Type | Description |
|---|---|---|
| Date | Trading day | Date |
| Open | Opening price of stock | Float |
| High | Highest price of stock | Float |
| Low | Lowest price of stock | Float |
| Close | Closing price of stock | Float |
| Adj. Close | Adjusted closing price of stock | Float |
| Change | Percentage change in closing price | Float |
| Volume | Trading volume in number of shares | Integer |

Tables 12 and 13 provide a comprehensive set of financial and market-valuation. The first table details income-statement, balance-sheet and cash-flow variables—ranging from top-line figures (e.g. Revenue, Gross Profit) through profitability margins (e.g. EBITDA Margin, Profit Margin) to liquidity and leverage measures (e.g. Current Ratio, Debt/Equity). The second table covers market-based measures such as Market Capitalization, Enterprise Value multiples (e.g. EV/EBITDA), pricing ratios (e.g. P/E, P/FCF), and return metrics (e.g. Dividend Yield, Total Return).

## H Covariate Usage in Conventional Baselines

Most conventional forecasting baselines used in the main reference experiments are not specifically designed for the panel-based setting considered in PanelTS, where multiple related units share covariate definitions. In particular, directly adding shared covariates to a standard multivariate forecasting formulation may change the task structure and does not necessarily lead to the strongest configuration for these models. To examine this issue, we conduct an additional check using SCINet on the currency exchange-rate dataset.

Table 14 compares three SCINet settings. The first setting applies SCINet repeatedly to each currency and combines the resulting unit-wise forecasts. The second setting treats all exchange rates as a multivariate target without using covariates. The third setting further adds covariates to the multivariate formulation. The results show that adding covariates improves the multivariate variant, reducing RMSE from 98.44 to 96.23 and MAE from 21.18 to 18.99. However, this remains substantially worse than repeated unit-wise prediction, which achieves RMSE 5.90 and MAE 5.50. This suggests that simply appending covariates to a standard multivariate forecasting model is not sufficient to exploit the panel-based structure of PanelTS.

Table 12: Covariates for Stock Market Data - Part I

| Variable | Description | Data Type |
|---|---|---|
| Date | Reporting date of the period | date |
| Revenue | Total revenue for the period | Float |
| Revenue Growth | Year-over-year revenue growth rate | Float |
| Cost of Revenue | Cost of goods sold | Float |
| Gross Profit | Revenue minus cost of revenue | Float |
| Selling, General & Administrative Expenses | SG&A expenses | Float |
| Research & Development Expenses | R&D expenses | Float |
| Operating Expenses | Total operating expenses | Float |
| Other Operating Expenses | Other miscellaneous operating expenses | Float |
| Operating Income | Earnings from operations | Float |
| Interest Expense (Income) | Net interest expense or income | Float |
| Other Expense (Income) | Other non-operating expense or income | Float |
| Pretax Income | Income before tax expense | Float |
| Income Tax Expense | Income tax expense | Float |
| Net Income | Net income including all items | Float |
| Preferred Dividends | Dividends to preferred shareholders | Float |
| Net Income to Common Shareholders | Net income attributable to common shareholders | Float |
| Earnings Per Share (Diluted) | Diluted earnings per share | Float |
| Weighted Average Shares Outstanding (Diluted) | Weighted average diluted shares outstanding | Float |
| EBITDA | Earnings before interest, taxes, depreciation, and amortization | Float |
| EBIT | Earnings before interest and taxes | Float |
| Net Income Growth | Year-over-year net income growth rate | Float |
| Shares Outstanding (Basic) | Basic weighted average shares outstanding | Float |
| Shares Outstanding (Diluted) | Diluted weighted average shares outstanding | Float |
| Shares Change | Change in shares outstanding | Float |
| EPS (Basic) | Basic earnings per share | Float |
| EPS Growth | Year-over-year EPS growth rate | Float |
| Free Cash Flow Per Share | Free cash flow divided by shares outstanding | Float |
| Dividend Per Share | Dividends declared per share | Float |
| Dividend Growth | Year-over-year dividend per share growth rate | Float |
| Gross Margin | Gross profit as a percentage of revenue | Float |
| Operating Margin | Operating income as a percentage of revenue | Float |
| Profit Margin | Net income as a percentage of revenue | Float |
| Free Cash Flow Margin | Free cash flow as a percentage of revenue | Float |
| Effective Tax Rate | Income tax expense as a percentage of pre-tax income | Float |
| EBITDA Margin | EBITDA as a percentage of revenue | Float |
| Depreciation & Amortization | Depreciation and amortization expenses | Float |
| EBIT Margin | EBIT as a percentage of revenue | Float |
| Tangible Book Value | Book value excluding intangible assets | Float |
| Tangible Book Value Per Share | Tangible book value per share | Float |
| Total Cash & Cash Equivalents | Total cash and cash equivalents | Float |
| Total Cash Per Share | Cash and cash equivalents per share | Float |
| Total Debt | Total interest-bearing debt obligations | Float |
| Current Ratio | Current assets divided by current liabilities | Float |
| Quick Ratio | (Current assets – inventory) / current liabilities | Float |
| Return on Invested Capital (ROIC) | Return on invested capital | Float |

Therefore, in the main reference table, conventional baselines are reported under their standard target-history setting. This choice reflects their intended use in time series forecasting and avoids forcing covariates into model formulations that are not designed to use them effectively.

# I Additional Reference Results

## I.1 Extended Results on Currency Exchange Rates

This section provides additional exchange-rate results under different input lengths and prediction horizons. Table 15 reports detailed results for the setting with input length $W = 100$ and horizons $H \in \{1, 2, 5, 10\}$, which extends the one-step results shown in the main text. Table 16 summarizes the best reported values across all input lengths $W \in \{20, 50, 100, 200\}$ and horizons $H \in \{1, 2, 5, 10\}$.

Table 16 shows that the best-performing method varies across input lengths, prediction horizons, and evaluation metrics. This illustrates that PanelTS can support multiple forecasting task settings, rather than being tied to a single fixed input length or prediction horizon.

## I.2 Extended Results on COVID-19

This section provides additional COVID-19 forecasting results under different input lengths, prediction horizons, and spatial granularities. Table 17 reports detailed results for the setting with

Table 13: Covariates for Stock Market Data - Part II

| Variable | Description | Data Type |
|---|---|---|
| Market Capitalization | Total market value of equity | Float |
| Enterprise Value | Market cap + total debt – total cash | Float |
| Cash & Equivalents | Cash and cash equivalents on the balance sheet | Float |
| Short-Term Investments | Marketable securities and other ST investments | Float |
| Cash Growth | Year-over-year change in cash and equivalents | Float |
| Receivables | Accounts receivable | Float |
| Inventory | Inventory | Float |
| Other Current Assets | Other current-asset items | Float |
| Total Current Assets | Sum of all current assets | Float |
| Property, Plant & Equipment | Net PP&E | Float |
| Long-Term Investments | Long-term investment holdings | Float |
| Goodwill and Intangibles | Goodwill plus intangible assets | Float |
| Other Long-Term Assets | Other non-current assets | Float |
| Total Long-Term Assets | Sum of all non-current assets | Float |
| Total Assets | Total assets | Float |
| Accounts Payable | Accounts payable | Float |
| Deferred Revenue | Deferred revenue | Float |
| Current Debt | Short-term portion of debt | Float |
| Total Current Liabilities | Sum of all current liabilities | Float |
| Other Current Liabilities | Other current-liability items | Float |
| Long-Term Debt | Long-term debt obligations | Float |
| Total Long-Term Liabilities | Sum of all non-current liabilities | Float |
| Other Long-Term Liabilities | Other non-current-liability items | Float |
| Total Liabilities | Total liabilities | Float |
| Debt Growth | Year-over-year change in total debt | Float |
| Retained Earnings | Accumulated retained earnings | Float |
| Comprehensive Income | Total comprehensive income | Float |
| Shareholders Equity | Total equity attributable to shareholders | Float |
| Net Cash / Debt | Cash minus debt | Float |
| Net Cash / Debt Growth | Year-over-year change in net cash/debt | Float |
| Net Cash Per Share | (Cash – Debt) per share | Float |
| Working Capital | Current assets minus current liabilities | Float |
| Book Value Per Share | Total equity divided by shares outstanding | Float |
| Share-Based Compensation | Stock-based compensation expense | Float |
| Other Operating Activities | Other cash flows from operating activities | Float |
| Operating Cash Flow | Net cash flow from operating activities | Float |
| Operating Cash Flow Growth | Year-over-year change in operating cash flow | Float |
| Capital Expenditures | Capital expenditures | Float |
| Acquisitions | Cash spent on acquisitions | Float |
| Change in Investments | Change in investment holdings | Float |
| Other Investing Activities | Other cash flows from investing activities | Float |
| Investing Cash Flow | Net cash flow from investing activities | Float |
| Dividends Paid | Cash dividends paid | Float |
| Share Issuance / Repurchase | Cash from issuing or repurchasing shares | Float |
| Debt Issued / Paid | Net cash from debt issued or repaid | Float |
| Other Financing Activities | Other cash flows from financing activities | Float |
| Financing Cash Flow | Net cash flow from financing activities | Float |
| Net Cash Flow | Total net cash flow Float | Float |
| Free Cash Flow | Free cash flow | Float |
| Free Cash Flow Growth Market Cap Growth | Year-over-year change in market capitalization | Float |
| EV / EBIT | Enterprise value divided by EBIT | Float |
| EV / EBITDA | Enterprise value divided by EBITDA | Float |
| Price / Earnings (P/E) | Share price divided by earnings per share | Float |
| PEG Ratio | Price/earnings to growth ratio | Float |
| Price/Earnings to Growth | Alternate PEG calculation | Float |
| Price / Sales (P/S) | Share price divided by sales per share | Float |
| Price / Book (P/B) | Share price divided by book value per share | Float |
| PB Ratio (alt.) | Alternate notation for P/B | Float |
| Price / Free Cash Flow (P/FCF) | Share price divided by free cash flow per share | Float |
| Price / Operating Cash Flow (P/OCF) | Share price divided by operating cash flow per share | Float |
| Debt / Equity | Total debt divided by total equity | Float |
| Dividend Yield | Annual dividend per share divided by share price | Float |
| Payout Ratio | Dividends per share divided by earnings per share | Float |
| Buyback Yield | Share repurchase amount divided by market cap | Float |
| Total Return | Combined price appreciation and dividends | Float |

Table 14: Effect of multivariate prediction and covariate usage on currency exchange-rate forecasting using SCINet as a representative conventional baseline. Results are reported using raw RMSE, MAE, and runtime in seconds.

| Setting | RMSE | MAE | Time (sec) |
|---|---|---|---|
| SCINet, repeated unit-wise prediction | 5.90 | 5.50 | 685.2 |
| SCINet, multivariate targets only | 98.44 | 21.18 | 121.2 |
| SCINet, multivariate with covariates | 96.23 | 18.99 | 129.5 |

Table 15: Extended currency exchange-rate results with input length $W = 100$ and prediction horizons $H \in \{1, 2, 5, 10\}$. Each cell lists results for horizons 1/2/5/10. $\mathrm{RMSE}^*$ and $\mathrm{MAE}^*$ denote normalized errors scaled by 100. Cost is reported in minutes.

| Model | RMSE | MAE | $\mathrm{RMSE}^*$ | $\mathrm{MAE}^*$ | Cost |
|---|---|---|---|---|---|
| ARIMA | 6.82 / 7.84 / 10.84 / 15.46 | 5.68 / 7.02 / 10.06 / 15.14 | 1.28 / 1.54 / 2.11 / 2.96 | 0.63 / 0.83 / 1.25 / 1.81 | 0.19 / 0.24 / 0.29 / 0.31 |
| VAR | 6.82 / 7.84 / 10.84 / 15.46 | 5.92 / 7.10 / 9.93 / 15.03 | 1.25 / 1.48 / 2.11 / 2.99 | 0.65 / 0.81 / 1.26 / 1.83 | – |
| ESE | 6.01 / 7.60 / 10.69 / 15.31 | 5.52 / 6.75 / 9.87 / 14.49 | 1.18 / 1.50 / 2.08 / 2.91 | 0.60 / 0.82 / 1.24 / 1.81 | 0.22 / 0.21 / 0.23 / 0.23 |
| LSTM | 6.17 / 7.71 / 10.77 / 15.37 | 5.67 / 6.87 / 9.87 / 14.71 | 1.21 / 1.48 / 2.09 / 2.92 | 0.60 / 0.81 / 1.24 / 1.80 | 5.71 / 6.81 / 8.46 / 9.28 |
| DLinear | 5.88 / 7.55 / 10.59 / 15.18 | 5.41 / 6.68 / 9.71 / 14.21 | 1.21 / 1.43 / 2.09 / 2.84 | 0.60 / 0.80 / 1.25 / 1.78 | 6.33 / 8.25 / 8.93 / 9.88 |
| NLinear | 5.96 / 7.58 / 10.60 / 15.06 | 5.46 / 6.79 / 9.85 / 14.17 | 1.21 / 1.47 / 2.06 / 2.85 | 0.61 / 0.81 / 1.23 / 1.78 | 6.71 / 8.30 / 9.08 / 9.85 |
| Informer | 6.01 / 7.67 / 10.60 / 15.29 | 5.46 / 6.79 / 9.85 / 14.17 | 1.22 / 1.46 / 2.10 / 2.92 | 0.61 / 0.81 / 1.25 / 1.80 | 3.58 / 4.92 / 5.42 / 6.12 |
| FiLM | 5.93 / 7.72 / 10.76 / 15.34 | 5.54 / 7.02 / 9.73 / 14.45 | 1.22 / 1.46 / 2.07 / 2.90 | 0.61 / 0.80 / 1.24 / 1.81 | 6.69 / 9.15 / 9.36 / 10.44 |
| SCINet | 5.90 / 7.60 / 10.47 / 15.30 | 5.50 / 6.94 / 9.58 / 14.87 | 1.18 / 1.46 / 2.06 / 2.90 | 0.60 / 0.80 / 1.23 / 1.79 | 8.82 / 10.49 / 12.37 / 13.41 |
| DeepAR | 6.08 / 7.62 / 10.62 / 15.20 | 5.61 / 6.86 / 9.90 / 14.53 | 1.22 / 1.48 / 2.11 / 2.91 | 0.61 / 0.81 / 1.25 / 1.81 | 5.19 / 6.08 / 7.63 / 8.33 |
| KVAE | 6.05 / 7.65 / 10.81 / 15.24 | 5.59 / 6.94 / 9.85 / 14.56 | 1.20 / 1.49 / 2.11 / 2.93 | 0.60 / 0.81 / 1.25 / 1.82 | 4.39 / 5.60 / 7.12 / 8.02 |
| TPGNN | 6.09 / 7.53 / 10.64 / 15.24 | 5.63 / 6.76 / 9.75 / 14.88 | 1.18 / 1.48 / 2.10 / 2.96 | 0.60 / 0.81 / 1.26 / 1.81 | 6.75 / 7.78 / 8.90 / 9.47 |
| PatchTST | 6.04 / 7.45 / 10.63 / 15.32 | 5.57 / 6.72 / 9.63 / 14.70 | 1.21 / 1.46 / 2.04 / 2.91 | 0.61 / 0.80 / 1.23 / 1.80 | 4.32 / 5.41 / 6.06 / 6.64 |
| TimeLLM | 5.91 / 7.57 / 10.49 / 14.96 | 5.52 / 6.67 / 9.46 / 13.84 | 1.18 / 1.51 / 2.08 / 2.71 | 0.60 / 0.81 / 1.23 / 1.82 | 4.53 / 5.62 / 6.24 / 6.97 |

Table 16: Summary of extended currency exchange-rate results across input lengths and prediction horizons. Each row corresponds to one task setting. The table reports the best reported raw errors, normalized errors, and fastest runtime for each setting. When multiple methods share the same best value after rounding, tied methods are listed.

| Input length $W$ | Horizon $H$ | Best RMSE | Best MAE | Best $\mathrm{RMSE}^*$ | Best $\mathrm{MAE}^*$ | Fastest |
|---|---|---|---|---|---|---|
| 20 | 1 | SCINet: 6.42 | SCINet: 5.62 | SCINet: 1.17 | SCINet: 0.62 | ARIMA: 0.18 |
| 20 | 2 | TPGNN: 7.96 | PatchTST: 6.99 | FiLM: 1.48 | PatchTST: 0.83 | ESE/ARIMA: 0.21 |
| 20 | 5 | DLinear: 11.74 | DLinear: 10.15 | LSTM/PatchTST: 2.17 | DLinear: 1.29 | ESE: 0.22 |
| 20 | 10 | DLinear: 17.92 | DLinear: 15.68 | DLinear: 3.36 | DLinear: 1.92 | ESE: 0.23 |
| 50 | 1 | TPGNN: 6.11 | PatchTST: 5.29 | DLinear: 1.23 | ESE: 0.60 | ARIMA: 0.18 |
| 50 | 2 | PatchTST: 7.74 | SCINet: 6.64 | LSTM: 1.49 | LSTM/DLinear: 0.79 | ESE: 0.21 |
| 50 | 5 | SCINet/TimeLLM: 10.93 | NLinear/Informer: 9.70 | DLinear: 2.11 | DLinear/SCINet: 1.22 | ESE: 0.23 |
| 50 | 10 | TimeLLM: 15.82 | TimeLLM: 13.47 | SCINet: 3.00 | TimeLLM: 1.79 | ESE: 0.23 |
| 100 | 1 | DLinear: 5.88 | DLinear: 5.41 | ESE/SCINet: 1.18 | ESE/SCINet: 0.60 | ARIMA: 0.19 |
| 100 | 2 | PatchTST: 7.45 | TimeLLM: 6.67 | DLinear: 1.43 | DLinear/SCINet: 0.80 | ESE: 0.21 |
| 100 | 5 | SCINet: 10.47 | TimeLLM: 9.46 | PatchTST: 2.04 | SCINet/PatchTST: 1.23 | ESE: 0.23 |
| 100 | 10 | TimeLLM: 14.96 | TimeLLM: 13.84 | TimeLLM: 2.71 | DLinear/NLinear: 1.78 | ESE: 0.23 |
| 200 | 1 | Informer: 5.83 | TPGNN: 5.19 | TimeLLM: 1.19 | ESE: 0.59 | ARIMA: 0.19 |
| 200 | 2 | TimeLLM: 7.46 | PatchTST: 6.51 | TimeLLM: 1.45 | ESE/TimeLLM: 0.79 | ESE: 0.21 |
| 200 | 5 | ESE: 10.30 | TimeLLM: 9.21 | TimeLLM: 2.03 | TimeLLM: 1.19 | ESE: 0.24 |
| 200 | 10 | FiLM: 14.90 | PatchTST: 13.86 | TimeLLM: 2.77 | TimeLLM: 1.73 | ESE: 0.24 |

input length $W = 100$ and prediction horizons $H \in \{1, 2, 5, 10\}$. Each cell lists results for horizons 1/2/5/10. Table 18 summarizes the best reported values across all input lengths $W \in \{10, 20, 50, 100\}$ and horizons $H \in \{1, 2, 5, 10\}$.

Table 18 shows that the best-performing method varies across input lengths, prediction horizons, and spatial granularities. This illustrates that the COVID-19 component of PanelTS can support multiple forecasting settings rather than a single fixed evaluation configuration.

Table 17: Extended COVID-19 forecasting results with input length $W = 100$ and prediction horizons $H \in \{1, 2, 5, 10\}$. Each cell lists results for horizons 1/2/5/10. Results are reported separately for the 20-unit, 79-unit, and 320-unit settings.

| Model | 20 units | | 79 units | | 320 units | |
|---|---|---|---|---|---|---|
| | RMSE | MAE | RMSE | MAE | RMSE | MAE |
| ARIMA | 79.70 / 82.07 / 93.27 / 98.31 | 79.59 / 86.79 / 88.08 / 102.05 | 77.72 / 80.34 / 85.53 / 87.48 | 67.41 / 73.38 / 79.06 / 82.51 | 7.44 / 7.72 / 8.23 / 9.20 | 6.24 / 6.77 / 7.35 / 7.98 |
| VAR | 95.79 / 95.86 / 114.49 / 117.09 | 90.86 / 91.64 / 103.24 / 112.44 | 90.85 / 93.56 / 98.09 / 112.45 | 81.74 / 83.38 / 91.26 / 100.95 | 8.07 / 8.41 / 9.57 / 10.45 | 7.68 / 7.83 / 8.70 / 9.79 |
| ESE | 84.54 / 92.19 / 94.74 / 97.34 | 83.91 / 88.29 / 92.14 / 95.02 | 55.54 / 59.72 / 61.91 / 69.27 | 50.85 / 53.45 / 55.90 / 64.62 | 4.83 / 5.24 / 5.89 / 5.55 | 4.58 / 4.78 / 5.04 / 5.30 |
| LSTM | 78.45 / 84.85 / 95.02 / 99.83 | 74.96 / 76.35 / 89.23 / 95.77 | 61.42 / 66.73 / 74.49 / 75.18 | 57.26 / 59.12 / 68.40 / 64.53 | 6.23 / 6.77 / 7.56 / 7.94 | 5.47 / 5.62 / 6.56 / 7.00 |
| DLinear | 79.47 / 82.09 / 85.64 / 92.78 | 72.41 / 72.86 / 77.40 / 88.74 | 56.16 / 57.09 / 60.76 / 63.74 | 53.95 / 53.81 / 57.57 / 58.37 | 5.93 / 6.12 / 6.39 / 6.90 | 5.10 / 5.13 / 5.44 / 5.78 |
| NLinear | 74.96 / 81.63 / 85.26 / 89.84 | 70.77 / 74.01 / 80.04 / 81.92 | 55.62 / 60.84 / 63.34 / 63.35 | 52.23 / 54.71 / 59.30 / 59.71 | 5.47 / 5.96 / 6.24 / 6.61 | 4.99 / 5.22 / 5.62 / 5.80 |
| Informer | 76.84 / 78.13 / 85.99 / 93.52 | 71.79 / 77.42 / 80.70 / 90.27 | 62.01 / 64.41 / 74.67 / 73.26 | 60.57 / 65.27 / 70.23 / 67.35 | 6.05 / 6.30 / 6.53 / 7.41 | 5.80 / 6.27 / 6.48 / 7.26 |
| FiLM | 76.84 / 83.54 / 84.06 / 98.40 | 73.85 / 76.50 / 79.11 / 82.39 | 56.21 / 61.18 / 65.31 / 68.25 | 51.34 / 53.08 / 56.94 / 56.29 | 5.74 / 6.21 / 6.06 / 7.37 | 4.96 / 5.07 / 5.65 / 5.49 |
| SCINet | 79.75 / 82.18 / 83.74 / 96.06 | 69.94 / 69.80 / 76.46 / 88.44 | 55.86 / 56.96 / 68.90 / 69.10 | 53.56 / 53.10 / 58.84 / 63.85 | 5.94 / 6.06 / 6.27 / 7.17 | 5.59 / 5.50 / 6.14 / 7.09 |
| DeepAR | 83.41 / 94.20 / 97.04 / 99.18 | 74.65 / 91.80 / 87.61 / 84.99 | 56.34 / 66.58 / 65.39 / 67.14 | 54.32 / 63.03 / 64.01 / 62.18 | 6.29 / 6.32 / 7.34 / 7.37 | 5.04 / 6.13 / 5.95 / 6.48 |
| KVAE | 70.85 / 85.97 / 93.53 / 98.45 | 64.44 / 76.52 / 84.68 / 97.92 | 53.12 / 60.00 / 64.77 / 70.75 | 52.71 / 58.67 / 62.54 / 68.02 | 5.29 / 5.93 / 6.59 / 7.60 | 4.94 / 5.82 / 5.21 / 7.46 |
| TPGNN | 78.37 / 96.12 / 89.21 / 96.70 | 69.86 / 81.05 / 79.55 / 90.85 | 59.22 / 72.72 / 76.22 / 65.08 | 58.13 / 67.01 / 74.43 / 62.77 | 6.06 / 7.50 / 7.77 / 8.02 | 5.41 / 6.30 / 6.93 / 6.99 |
| PatchTST | 70.01 / 72.20 / 84.02 / 92.65 | 63.52 / 69.19 / 79.02 / 85.67 | 56.44 / 61.24 / 66.97 / 78.95 | 52.41 / 56.65 / 57.05 / 73.31 | 5.11 / 5.32 / 6.41 / 6.80 | 4.43 / 4.81 / 5.93 / 6.10 |
| TimeLLM | 73.47 / 81.53 / 77.49 / 93.36 | 70.02 / 78.06 / 79.67 / 83.26 | 59.48 / 65.81 / 62.13 / 71.49 | 52.64 / 54.82 / 59.23 / 69.72 | 5.12 / 7.22 / 6.10 / 6.67 | 4.82 / 5.21 / 5.99 / 6.42 |

Table 18: Summary of extended COVID-19 results across input lengths, prediction horizons, and spatial granularities. Each row corresponds to one input length and prediction horizon. The table reports the best reported RMSE and MAE for the 20-unit, 79-unit, and 320-unit settings. When multiple methods share the same best value after rounding, tied methods are listed.

| Input length $W$ | Horizon $H$ | 20 units | | 79 units | | 320 units | |
|---|---|---|---|---|---|---|---|
| | | Best RMSE | Best MAE | Best RMSE | Best MAE | Best RMSE | Best MAE |
| 10 | 1 | NLinear: 15.01 | KVAE: 13.43 | NLinear: 43.28 | LSTM: 40.02 | TimeLLM: 4.04 | NLinear: 3.24 |
| 10 | 2 | SCINet: 15.86 | SCINet: 13.87 | NLinear: 47.48 | LSTM: 41.36 | NLinear/PatchTST: 4.52 | SCINet: 3.33 |
| 10 | 5 | PatchTST: 19.05 | PatchTST: 15.13 | KVAE: 52.46 | TimeLLM: 43.82 | NLinear: 4.72 | SCINet: 3.62 |
| 10 | 10 | SCINet: 20.48 | FiLM: 17.00 | NLinear: 52.43 | FiLM: 47.94 | NLinear: 4.96 | FiLM: 3.70 |
| 20 | 1 | NLinear: 34.78 | KVAE: 31.47 | KVAE: 48.14 | KVAE: 47.23 | ESE: 4.76 | TimeLLM: 4.33 |
| 20 | 2 | SCINet: 36.66 | SCINet: 31.96 | DLinear: 55.80 | SCINet: 50.10 | ESE: 5.12 | TimeLLM: 4.52 |
| 20 | 5 | SCINet: 37.20 | SCINet: 34.36 | DLinear: 58.90 | PatchTST: 50.94 | TimeLLM: 5.29 | FiLM: 4.61 |
| 20 | 10 | NLinear: 49.79 | PatchTST: 43.36 | NLinear: 63.18 | FiLM: 56.51 | ESE: 5.49 | ESE: 5.11 |
| 50 | 1 | TPGNN: 53.74 | KVAE: 45.96 | KVAE: 52.41 | FiLM: 48.60 | ESE: 4.73 | KVAE: 4.19 |
| 50 | 2 | TimeLLM: 51.62 | KVAE: 48.42 | DLinear: 56.70 | FiLM: 50.10 | ESE: 5.01 | FiLM/PatchTST: 4.65 |
| 50 | 5 | TimeLLM: 55.95 | PatchTST: 52.09 | TimeLLM: 56.24 | ESE: 54.99 | TimeLLM: 5.07 | FiLM: 4.80 |
| 50 | 10 | KVAE: 65.42 | FiLM: 52.49 | DLinear: 64.59 | FiLM: 54.18 | ESE: 5.42 | FiLM: 5.09 |
| 100 | 1 | PatchTST: 70.01 | PatchTST: 63.52 | KVAE: 53.12 | ESE: 50.85 | ESE: 4.83 | PatchTST: 4.43 |
| 100 | 2 | PatchTST: 72.20 | PatchTST: 69.19 | SCINet: 56.96 | FiLM: 53.08 | ESE: 5.24 | ESE: 4.78 |
| 100 | 5 | TimeLLM: 77.49 | SCINet: 76.46 | DLinear: 60.76 | ESE: 55.90 | ESE: 5.89 | ESE: 5.04 |
| 100 | 10 | NLinear: 89.84 | NLinear: 81.92 | NLinear: 63.35 | FiLM: 56.29 | ESE: 5.55 | ESE: 5.30 |