Calendar and Session Representations in Mixed-Calendar Multivariate Forecasting

Khudran M. Alzhrani
Computers Department-College of Engineering and Computing in Al-Qunfudhah, Umm Al-Qura University,
Al-Qunfudhah, Makkah, Saudi Arabia

***Abstract*—Multivariate forecasting with variables observed on different temporal calendars requires a choice about how their histories are represented in a common input sequence. Prior work has largely treated mixed-calendar alignment as a preprocessing step or focused on modeling irregular observations, leaving direct comparisons of alternative temporal representations under identical forecast targets and evaluation dates less examined. This study compares a *Calendar* representation, which preserves the complete daily grid, with a *Session* representation, which retains only observed financial trading sessions. Epidemiological and financial data from Saudi Arabia and Germany are evaluated using identical forecast targets and target dates. Under *fixed-duration matching*, both representations use the same preceding 28 calendar days, whereas under *fixed-sequence-length matching*, Calendar uses 20 calendar days and Session uses the 20 most recent trading sessions. Ridge regression, a multilayer perceptron, and a gated recurrent unit are used to examine whether representation effects vary across model classes, while additional variants evaluate observation indicators and elapsed-time information. Ridge consistently favors Calendar across the evaluated empirical settings, MLP generally favors Calendar with variation across countries and matching rules, while GRU effects are smaller in magnitude and most 95% uncertainty intervals span zero. The choice between fixed-duration and fixed-sequence-length matching changes some representation effects, while adding observation indicators or elapsed-time information does not consistently improve performance. A target-specific naive baseline outperforms all core learned configurations, so the learned-model comparisons are interpreted primarily as analyses of representation sensitivity. In controlled simulation, representation differences are small under regular scheduled closures but become more Calendar-favoring when additional observation gaps are introduced. Overall, the results show that the effects of temporal representation vary with the model, the construction of the historical input, and the way observation gaps arise.**

Keywords*—Mixed-calendar forecasting; temporal representation; multivariate time-series forecasting; temporal alignment; irregular time series; observation gaps

I. Introduction

Combining variables that follow different observation calendars creates a fundamental question about which dates and observations should be retained in the model input. When one process is observed every calendar day and another only at selected sessions, preserving the complete calendar and retaining only trading-session dates expose a model to different histories. In the setting considered here, epidemiological variables are recorded daily, whereas financial observations are updated only on trading sessions. A calendar-preserving representation retains epidemiological observations that occur while the market is closed but must represent the absence of new financial observations on those dates. A session-based representation avoids those closed-market positions, but in doing so removes epidemiological observations recorded on non-trading dates and compresses unequal calendar intervals into adjacent sequence positions. The two representations therefore differ in both the information they retain and the temporal history they present to the model. This is more than a change in how the input data are formatted.

The COVID-19 period provides a practical setting in which this mixed-calendar problem is explicit. Studies combining pandemic indicators with financial-market data have had to reconcile epidemiological observations recorded on the daily calendar with market observations available only on trading dates [1], [2]. In those studies, temporal alignment was a methodological step used to address the economic question of interest, rather than the focus of a direct comparison between alternative temporal representations. This leaves a distinct forecasting question: when variables following different observation calendars are modeled jointly, what are the consequences of preserving the complete calendar rather than retaining only the dates on which the financial series is observed?

The issue extends beyond this particular application. Research on financial time series has shown that consecutive trading observations can represent unequal amounts of elapsed calendar time and that non-trading periods can affect temporal dependence and subsequent forecasting behavior [3], [4], [5]. Economic sentiment can also shift during non-trading weekends and be associated with subsequent stock returns [6]. More broadly, irregular-time-series research has developed methods for representing observation availability, unequal temporal spacing, and asynchronous multivariate observations. Much of this literature addresses how irregular observations and their timing should be represented or modeled. A related but less directly examined question arises when more than one defensible temporal representation can be constructed before model fitting: whether preserving the common calendar or retaining only trading-session dates changes forecasting behavior when prediction targets and evaluation dates are held fixed. The answer may depend on how the model processes its input, whether fixed-duration or fixed-sequence-length matching is used, and how the observation gaps arise.

This study examines that question through a controlled comparison of *Calendar* and *Session* representations for multivariate forecasting with epidemiological and financial variables. The empirical analysis uses Saudi Arabia and Germany as two settings with different regular trading weeks and observed-session calendars.

The study makes three contributions:

- It formulates the Calendar–Session choice in mixed-calendar multivariate forecasting as a temporal-representation problem and directly compares preserving the complete calendar with retaining only observed trading sessions while keeping forecast targets and target dates identical across representations.
- It separately evaluates the Calendar–Session representation choice, the addition of observation indicators and elapsed-time information, and the choice between fixed-duration and fixed-sequence-length matching. The resulting design examines whether representation effects vary across model classes, prediction targets, countries, and matching rules.
- It uses controlled simulation to examine whether Calendar–Session representation effects depend on how observation gaps arise. The simulation compares regular scheduled closures with irregular holiday gaps, extended closures, random gaps, and state-dependent gaps.

The remainder of this study is organized as follows. Section II reviews related work on mixed-calendar alignment and irregular multivariate time series. Section III describes the data, temporal representations, forecasting design, evaluation procedure, controlled simulation, and sensitivity analyses. Section IV presents the empirical, simulation, and robustness results. Section V concludes the study.

II. Related Work

A. Non-Trading Periods and Mixed-Calendar Alignment

Financial time series provide a direct example of the distinction between sequence position and elapsed calendar time. Lyócsa and Molnár [3] showed that weekends and holidays create non-equidistant gaps between daily financial observations and that accounting for these gaps can affect estimated temporal dependence and out-of-sample volatility forecasts. Díaz-Mendoza and Pardo [4] further distinguished among weekends, holidays, and longer non-trading interruptions when modeling range-based volatility, highlighting that both the occurrence and duration of market closures can be relevant. Lyócsa and Todorova [5] examined information arising during trading and non-trading periods and found that volatility accumulated outside regular trading hours can contribute to subsequent volatility forecasts. Lau [6] similarly considered changes in economic sentiment over weekends and their association with subsequent stock returns. Collectively, this literature shows that consecutive trading observations can be separated by different amounts of calendar time and that information accumulated during non-trading periods can remain relevant to subsequent market behavior.

The representation problem becomes different when financial variables are combined with variables that continue to be observed while the market is closed. Pandemic-era financial studies provide concrete examples. O'Donnell et al. [1] aligned daily COVID-19 infections with stock-market observations by adding cases reported over weekends to the next trading day. Brueckner and Vespignani [2] used a comparable rule in their analysis of Australian stock-market performance, assigning infections reported on non-trading days to the following trading date. Such synchronization procedures are appropriate to the substantive questions addressed in those studies, but they do not directly compare alternative temporal representations while keeping the downstream forecasting targets and evaluation dates fixed.

B. Irregular Multivariate Time-Series Representation and Modeling

A broader irregular-time-series literature has examined how observation availability and unequal temporal spacing should be represented for statistical and machine-learning models. Li and Marlin [7] formulate irregular sampling from a missing-data perspective, emphasizing the relationship between observed values, observation indices, and the representation supplied to a learning algorithm. GRU-D [8] explicitly incorporates missingness masks and elapsed times into recurrent processing and uses learned decay to reflect the time since previous observations. These approaches are particularly relevant to the distinction considered here between changing the temporal index itself and augmenting a representation with explicit information about observation availability or elapsed time.

Another line of research models evolution over unequal intervals directly rather than treating consecutive observations as equally spaced discrete steps. Continuous-time formulations include latent ordinary differential equation models, continuous GRU-based dynamics, neural controlled differential equations, and continuous recurrent units [9], [10], [11], [12]. These methods provide different mechanisms for incorporating irregular observation times into the model dynamics.

Recent irregular multivariate time-series forecasting research has also reconsidered how asynchronous observations should be temporally organized for learning. Several methods seek to operate on irregular observations or construct compact representations without relying on conventional dense global synchronization [13], [14], [15], [16], [17]. In contrast, Zhou et al. [18] revisit canonical pre-alignment, in which variables are placed on a shared global timestamp structure, while addressing the computational cost introduced by sequence expansion.

How observation gaps arise provides a further distinction. Vanderschueren et al. [19] study informative sampling in a setting where observation times can depend on evolving characteristics or outcomes, showing why observation timing cannot always be treated as independent of the underlying process. Time-IMM [20] likewise distinguishes irregularity according to its generating cause, including trigger-based, constraint-based, and artifact-based mechanisms, rather than treating all irregular observation patterns as equivalent. These perspectives are relevant to the distinction between predictable market closures and gaps arising through less regular processes.

Across these strands of literature, prior work has adopted synchronization rules for mixed-calendar data, adjusted domain-specific forecasting models for non-equidistant observations, incorporated observation availability and elapsed time into predictive models, or developed specialized approaches for irregular multivariate sequences. The present study focuses specifically on the temporal-representation choice that precedes model fitting. It holds the forecasting targets and evaluation dates fixed while varying the temporal representation supplied to the model, compares fixed-duration matching with fixed-sequence-length matching, and evaluates representation effects across model classes and different ways observation gaps arise.

III. Methodology

A. Problem Formulation and Data

This study considers multivariate forecasting when variables combined in a common temporal model follow different observation calendars. Let

$$\mathcal{T} = \{t_1, t_2, \ldots, t_T\} \tag{1}$$

denote the ordered calendar dates in the study period, and let

$$\mathcal{S} \subset \mathcal{T} \tag{2}$$

denote the subset of dates with an observed financial trading session. Epidemiological observations are available on the daily calendar $\mathcal{T}$, whereas observed financial values are available only on $\mathcal{S}$. The study compares a *Calendar* representation that retains the complete daily grid with a *Session* representation that retains only observed trading-session dates.

The empirical evaluation uses epidemiological and financial data for Saudi Arabia and Germany over the common period from 2 March 2020 through 4 May 2023. Each country series contains 1,159 consecutive calendar dates. The forecasting variables are daily reported COVID-19 cases, daily reported deaths, a financial-level series, and financial log returns. The two countries are analyzed separately under the same forecasting and evaluation protocol. Table I summarizes the datasets and observation calendars.

Saudi epidemiological observations were obtained from the Saudi Arabia Coronavirus Disease (COVID-19) Situation dataset published by the King Abdullah Petroleum Studies and Research Center (KAPSARC) [21]. The Saudi financial series was constructed from a fixed panel of 184 Saudi-listed equities retrieved from Yahoo Finance through the `yfinance` package [22], [23]. Let $A^{\mathrm{cf}}_{k,t}$ denote the adjusted closing value for equity k, carried forward from its most recent available company-level observation, and let $\mathcal{K}_t$ denote the equities for which this value is available on date t. The aggregate financial level is:

$$P^{\mathrm{SA}}_t = \frac{1}{|\mathcal{K}_t|} \sum_{k \in \mathcal{K}_t} A^{\mathrm{cf}}_{k,t}. \tag{3}$$

A Saudi date is classified as an observed session when at least one retained equity records trading activity on that date. This session definition is independent of the aggregate construction, which can contain carried-forward company-level values for constituents without a new observation on the same date.

German epidemiological observations were derived from the Robert Koch Institute SARS-CoV-2 dataset [24] and aggregated nationally by reporting date (*Meldedatum*). The German financial series is the DAX index obtained through `yfinance` using the symbol `^GDAXI` [22], [23]. Observed DAX dates were validated against the Xetra trading calendars for 2020–2023 [25].

For either country, let P_t denote the financial level on an observed session $t \in \mathcal{S}$. For Calendar inputs, the most recently observed market-level value is carried forward across non-session dates,

$$P^{\mathrm{cf}}_t = P_{s(t)}, \qquad s(t) = \max\{s \in \mathcal{S} : s \leq t\}. \tag{4}$$

No backward filling is used. The observed-session indicator is $O_t = \mathbb{I}(t \in \mathcal{S})$. For an observed session t, the financial log return relative to the preceding observed session t^- is

$$R_t = \log P_t - \log P_{t^-}, \tag{5}$$

and $Q_t = \mathbb{I}(R_t \text{ is available})$ denotes return availability.

B. Forecasting Task and Temporal Representations

The prediction task is fixed-horizon multi-target forecasting. For target start date t, all representations predict four outcomes over the subsequent seven-calendar-day interval

$$\mathcal{H}_t = \{t, t+1, \ldots, t+6\}. \tag{6}$$

All historical inputs precede t. The epidemiological targets are the log-transformed seven-day sums

$$y^{(\mathrm{cases})}_t = \log\left(1 + \sum_{\tau \in \mathcal{H}_t} C_\tau\right), \tag{7}$$

$$y^{(\mathrm{deaths})}_t = \log\left(1 + \sum_{\tau \in \mathcal{H}_t} D_\tau\right), \tag{8}$$

where, C_τ and D_τ denote daily reported cases and deaths.

Let

$$\mathcal{O}_t = \{\tau \in \mathcal{H}_t : Q_\tau = 1\} \tag{9}$$

denote the dates in the forecast interval with an available observed-session return. The financial targets are cumulative log return and the root sum of squared observed-session log returns,

TABLE I. EMPIRICAL DATASETS AND OBSERVATION CALENDARS USED IN THE FORECASTING EXPERIMENT

Country	Epidemiological data	Financial series	Regular trading week	Calendar days	Observed sessions	Non-session dates
Saudi Arabia	Daily national COVID-19 case and death data derived from the KAPSARC Saudi Arabia COVID-19 dataset	Constructed adjusted-close aggregate from 184 Saudi-listed equities	Sunday–Thursday	1,159	790	369
Germany	Daily national COVID-19 case and death data derived from the RKI SARS-CoV-2 dataset	DAX index aligned with the Xetra trading calendar	Monday–Friday	1,159	810	349

$$y_t^{(\mathrm{return})} = \sum_{\tau \in \mathcal{O}_t} R_\tau, \tag{10}$$

$$y_t^{(\mathrm{volatility})} = \sqrt{\sum_{\tau \in \mathcal{O}_t} R_\tau^2}. \tag{11}$$

An empty sum is defined as zero. These four targets are predicted jointly by the learned models.

Eligible target starts require a complete seven-day forecast interval, at least 28 preceding calendar days, and at least 20 preceding observed sessions. Eligible dates are sampled at a seven-calendar-day stride, producing 161 common target starts per country. The target dates and target values are identical across all temporal representations and both historical-input matching rules.

The four base input channels are log-transformed cases, log-transformed deaths, the carried-forward financial level on the daily grid, and the financial return. Five representation conditions are constructed from these channels, as summarized in Table II and illustrated in Fig. 1.

C0 retains the base channels at every calendar position. C1 additionally supplies O_t and Q_t, and C2 further includes δ_t, the number of calendar days since the most recent observed financial value. At Calendar positions that are not trading sessions, the market level is carried forward, while the financial return remains unavailable and is set to zero only after standardization. The Session representations retain only dates $s_1 < \cdots < s_L$ with $s_j \in \mathcal{S}$. Consequently, epidemiological observations occurring on non-session dates are not included as separate input positions in S0 or S1.

Let $\nu(t)$ denote the integer calendar-day index of date t. S1 adds the inter-session gap g_j, defined as the number of calendar days between consecutive observed sessions,

$$g_j = \begin{cases} 0, & j = 1, \\ \nu(s_j) - \nu(s_{j-1}), & j > 1. \end{cases} \tag{12}$$

C0 and S0 form the principal Calendar–Session comparison. C1–C0, C2–C1, and S1–S0 examine the incremental effects of observation indicators and elapsed-time information. C2–S1 compares Calendar and Session when both representations explicitly include elapsed-time information.

C. Historical-Input Matching and Preprocessing

Two historical-input matching rules are used because fixed-duration and fixed-sequence-length matching hold different aspects of the historical input constant. Under *fixed-duration matching*, both the Calendar and Session representations are constructed from the same preceding 28-calendar-day interval,

$$\mathcal{L}_t^{(D)} = \{t-28, t-27, \ldots, t-1\}. \tag{13}$$

Calendar therefore contains 28 daily positions, whereas Session retains only the observed sessions within those 28 calendar days.

Under *fixed-sequence-length matching*, Calendar uses the 20 calendar days immediately preceding t,

$$\mathcal{L}_t^{(L,C)} = \{t-20, t-19, \ldots, t-1\}, \tag{14}$$

whereas Session uses the 20 most recent observed sessions,

$$\mathcal{L}_t^{(L,S)} = \{s_{j-19}, \ldots, s_j\}, \qquad s_j < t. \tag{15}$$

The two representations therefore have the same sequence length but can span different amounts of calendar time. The C0–S0 contrast consequently evaluates the overall difference between the two practical representation choices. It does not isolate the effect of temporal indexing alone, because doing so would require retained information and historical coverage to be held fixed as well.

Fixed-duration Session inputs are left-padded to 28 positions for models that require inputs of equal length. Ridge regression and the multilayer perceptron use flattened arrays that retain these padding positions. Recurrent models process only the valid Session positions and do not update their hidden states on padding.

Fig. 1 summarizes the observation calendars, the Calendar and Session representation conditions, the two historical-input matching rules, and the common seven-calendar-day forecast target.

All fitted transformations are estimated separately within each temporal fold using the training partition only. Continuous input channels are standardized as

$$\widetilde{x}_{t,j} = \frac{x_{t,j} - \mu_j^{(\mathrm{tr})}}{\sigma_j^{(\mathrm{tr})}}, \tag{16}$$

where, financial-level statistics are estimated from positions corresponding to observed sessions and return statistics from positions with $Q_t = 1$. Carried-forward financial levels use the

TABLE II. TEMPORAL REPRESENTATION CONDITIONS USED IN THE MAIN FORECASTING EXPERIMENT

Code	Label	Temporal index	Non-session dates retained	Observation indicators	Elapsed-time information
C0	Calendar	Daily calendar	Yes	None	None
C1	Calendar + Indicators	Daily calendar	Yes	Observed session; return availability	None
C2	Calendar + Indicators + Elapsed Time	Daily calendar	Yes	Observed session; return availability	Days since the most recent observed financial value
S0	Session	Observed trading sessions	No	None	None
S1	Session + Gap	Observed trading sessions	No	None	Calendar days between consecutive observed sessions

scaling parameters estimated from observed financial levels. Unavailable returns and padded positions are set to zero after standardization, while binary indicators remain coded as 0 or 1.

Elapsed-time variables are transformed using

$$\phi(u) = \frac{\log\left(1 + \max(u, 0)\right)}{\log\left(1 + \max(M_u^{(\mathrm{tr})}, 1)\right)}, \tag{17}$$

where, $M_u^{(\mathrm{tr})}$ is the maximum nonnegative training value for the corresponding timing variable. The recurrent elapsed-step variable used by the decay-based GRU is transformed by the same function using its own maximum estimated from the training data. The four forecasting targets are standardized independently using means and standard deviations estimated from the training partition only; these fitted parameters are then applied unchanged to the corresponding validation and test targets.

D. Forecasting Models

The model set combines simple forecasting baselines with linear, feed-forward, and recurrent models. The *training-mean baseline* predicts the training-partition mean of each target. The *previous-seven-day baseline* applies the corresponding target-construction rule to the seven days immediately preceding the forecast interval. The *target-specific naive baseline* uses the preceding seven-day values for cases, deaths, and volatility and predicts zero cumulative return.

Ridge regression is fitted to flattened transformed inputs, with the regularization parameter selected from

$$\alpha \in \{0.01, 0.1, 1, 10, 100, 1000\} \tag{18}$$

using the mean standardized-target RMSE on the validation partition. The multilayer perceptron uses hidden layers of 64, 32, and 64 ReLU units, with dropout probability 0.05 after the first hidden layer and a four-unit linear output. The recurrent model is a single-layer gated recurrent unit (GRU) with hidden dimension 32 [26], followed by a linear projection from the final valid hidden state to the four targets.

A GRU with learned hidden-state decay is additionally evaluated for the C2 and S1 representations, which explicitly include elapsed-time information. The model follows the general use of elapsed-time-dependent decay in recurrent processing of irregular observations [8], while applying decay only to the previous hidden state. Let Δ_j denote the elapsed calendar interval associated with recurrent step j. A nonnegative hidden-unit-specific decay rate is learned as

$$\boldsymbol{\lambda} = \mathrm{softplus}(\boldsymbol{\theta}), \tag{19}$$

and the transformed elapsed interval $\widetilde{\Delta}_j$ defines

$$\boldsymbol{\gamma}_j = \exp\left(-\widetilde{\Delta}_j \boldsymbol{\lambda}\right), \tag{20}$$

$$\overline{\mathbf{h}}_{j-1} = \boldsymbol{\gamma}_j \odot \mathbf{h}_{j-1}. \tag{21}$$

The recurrent update is

$$\mathbf{h}_j = \mathrm{GRUCell}\left(\widetilde{\mathbf{x}}_j, \overline{\mathbf{h}}_{j-1}\right). \tag{22}$$

For Calendar sequences, $\Delta_1 = 0$ and $\Delta_j = 1$ for $j > 1$; for Session sequences, $\Delta_j = g_j$.

The neural models minimize mean squared error across the four standardized targets. Optimization uses Adam [27] with learning rate 10^{-3}, weight decay 10^{-5}, full-batch training, and gradient clipping at 5. Validation performance is evaluated after each epoch using the mean RMSE across the four standardized targets. Training is limited to 60 epochs with early-stopping patience of six, and the checkpoint with the lowest validation score is retained. Neural experiments are repeated using ten random seeds,

$$\{11, 22, 33, 44, 55, 66, 77, 88, 99, 111\}, \tag{23}$$

whereas Ridge regression and the non-learned baselines are deterministic and are evaluated once per temporal fold.

E. Out-of-Sample Evaluation and Statistical Analysis

The primary empirical evaluation uses four expanding-window out-of-sample folds. Training observations precede validation observations, and validation observations precede test observations within every fold. Validation and test blocks each contain 12 target starts separated by seven calendar days,

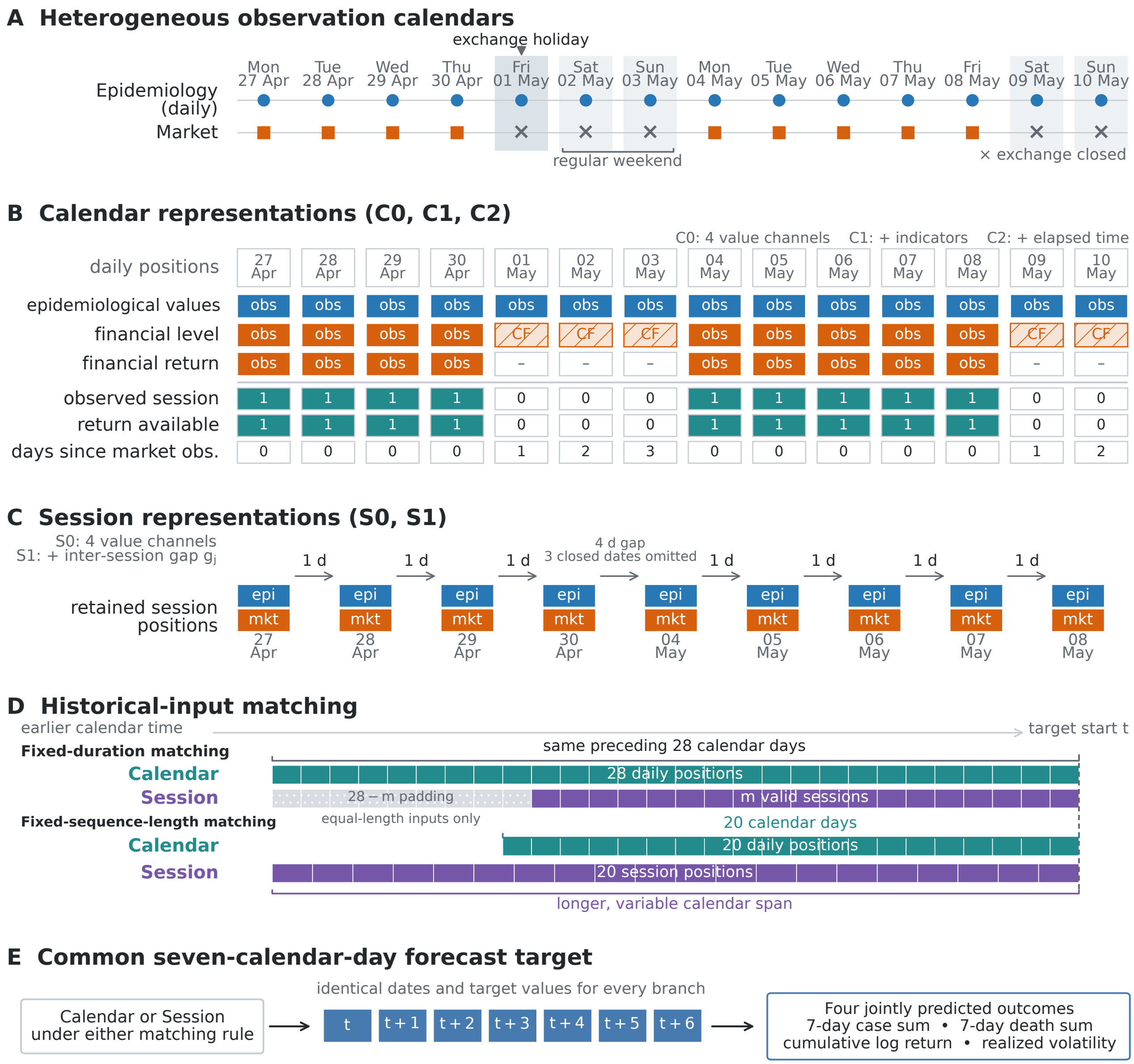

Fig. 1. Conceptual construction of the Calendar and Session representations and their common forecasting task. (A) An illustrative German calendar segment shows daily epidemiological observations and observed market sessions, including the 1 May 2020 exchange holiday and regular weekend closures. (B) Calendar representations retain every calendar date. C0 contains four value channels, C1 adds the observed-session and return-availability indicators, and C2 adds the number of calendar days since the most recent observed financial level. CF denotes a carried-forward financial level, and dashes denote unavailable financial returns on non-session dates. (C) Session representations retain only observed trading sessions, with S1 adding the number of calendar days between consecutive sessions to S0. (D) Fixed-duration matching uses the same preceding 28 calendar days, whereas fixed-sequence-length matching uses 20 positions and allows the Session representation to span a longer and variable calendar interval. Left padding under fixed-duration matching is used only for models that require inputs of equal length. (E) Every combination of representation and matching rule predicts the same four outcomes over an identical seven-calendar-day forecast interval.

so their seven-day target intervals do not overlap. A 35-calendar-day exclusion interval separates the end of the training target periods from validation and the end of the validation target periods from testing. The same temporal schedule is used for both countries, both historical-input matching rules, and all representation conditions. The resulting fold schedule is summarized in Table III.

The training window expands across folds from the earliest eligible target. Transformations are fitted using training data only, Ridge regularization is selected on validation data, and neural validation performance determines early stopping and checkpoint selection. The test partition is evaluated after these choices are fixed.

For target j, standardized RMSE is

TABLE III. PRIMARY EXPANDING-WINDOW OUT-OF-SAMPLE EVALUATION SCHEDULE. VALIDATION AND TEST PERIODS SHOW THE SPAN FROM THE BEGINNING OF THE FIRST TARGET INTERVAL TO THE END OF THE FINAL TARGET INTERVAL IN EACH BLOCK

Fold	Training target starts	Last training target end	Validation target period	Test target period
0	53	2021-04-04	2021-05-10–2021-08-01	2021-09-06–2021-11-28
1	77	2021-09-19	2021-10-25–2022-01-16	2022-02-21–2022-05-15
2	102	2022-03-13	2022-04-18–2022-07-10	2022-08-15–2022-11-06
3	127	2022-09-04	2022-10-10–2023-01-01	2023-02-06–2023-04-30

$$\mathrm{RMSE}_j^{(\mathrm{std})} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\widetilde{y}_{i,j} - \widehat{\widetilde{y}}_{i,j}\right)^2}, \tag{24}$$

and the aggregate forecasting score is the unweighted mean across the four targets,

$$\overline{\mathrm{RMSE}}_{\mathrm{std}} = \frac{1}{4}\sum_{j=1}^{4}\mathrm{RMSE}_j^{(\mathrm{std})}. \tag{25}$$

For model m, country c, and matching rule r, the principal representation contrast is

$$\Delta_{m,c,r} = \overline{\mathrm{RMSE}}_{\mathrm{std},m,c,r,\mathrm{C0}} - \overline{\mathrm{RMSE}}_{\mathrm{std},m,c,r,\mathrm{S0}}. \tag{26}$$

Throughout the study, the Calendar–Session difference denotes the Calendar score minus the Session score, as in Eq. 26; negative values indicate lower Calendar error and positive values indicate lower Session error. Point estimates are calculated directly from the original paired test predictions rather than from bootstrap averages.

Dependence-aware uncertainty is estimated with a circular moving-block bootstrap [28], [29]. Within each fold, the 12 ordered test target starts are resampled in circular blocks of four consecutive dates. With the seven-day target stride, this block length is used as a pragmatic local-dependence scale intended to reflect dependence associated with the 28-day historical-input window. It was not empirically optimized, and the correspondence is only approximate under fixed-sequence-length matching because the elapsed calendar span varies. Empirical contrasts use 1,000 bootstrap replicates while preserving the pairing between compared representations. For neural models, seed identifiers are resampled with replacement, and the same sampled seeds are retained across paired representations. The bootstrap resamples evaluation target starts within the already fitted folds and therefore conditions on the fitted expanding-window training histories; models are not refitted within each bootstrap replicate. The 2.5th and 97.5th percentiles of the resulting distribution define the reported 95% uncertainty intervals.

The same prediction-level resampling procedure is used for target-specific C0–S0 comparisons and for the effects of adding indicators or elapsed-time features in C1–C0, C2–C1, and S1–S0. Learned models are compared with the target-specific naive baseline by calculating the difference in aggregate score within each paired bootstrap replicate. Differences for individual targets are also summarized from the pooled original test predictions to determine whether the overall advantage of the naive baseline is driven by particular targets.

Model dependence is assessed by directly comparing each model's Calendar–Session effect with the corresponding GRU effect within the same bootstrap replicate. For model a relative to GRU,

$$I_{a-\mathrm{GRU},c,r} = \Delta_{a,c,r} - \Delta_{\mathrm{GRU},c,r}, \tag{27}$$

with $a \in \{\mathrm{Ridge}, \mathrm{MLP}\}$. The sampled target dates are shared across the two model contrasts, and neural seed samples are paired where applicable. This directly evaluates whether representation sensitivity differs by model class.

The effect of the matching rule is evaluated through

$$I_{\mathrm{match},m,c} = \Delta_{m,c,D} - \Delta_{m,c,L}, \tag{28}$$

where, D and L denote fixed-duration and fixed-sequence-length matching, respectively. The same target dates and bootstrap resampling are used for both terms. Seed- and fold-level summaries of whether the C0–S0 difference favors Calendar or Session are used descriptively to assess stability and are not treated as independent inferential replications. The reported 95% intervals are not adjusted for multiplicity across the full set of model, target, representation, and design comparisons. They are therefore interpreted as uncertainty summaries, with emphasis on effect magnitudes and recurring patterns across related conditions rather than on whether individual intervals exclude zero.

F. Controlled Simulation Study

A controlled simulation study examines whether Calendar–Session differences change with how observation gaps arise. Each simulation contains 730 calendar days and a latent binary regime $S_t \in \{0, 1\}$, initialized at $S_0 = 0$, with transition probabilities

$$\Pr(S_t \neq S_{t-1} \mid S_{t-1} = 0) = 0.035, \tag{29}$$
$$\Pr(S_t \neq S_{t-1} \mid S_{t-1} = 1) = 0.070. \tag{30}$$

A continuous latent process is initialized at $Z_0 = 0$ and evolves according to

$$Z_t = 0.88Z_{t-1} + 0.7S_t + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}\left(0, (0.35\sigma)^2\right), \tag{31}$$

where, $\sigma \in \{0.6, 1.0, 1.4\}$ controls the noise level.

Synthetic epidemiological and financial channels are generated from the latent process as

$$C_t^* = \max\left\{0, \exp\left(2.3 + 0.35Z_t + \eta_t^{(c)}\right) - 1\right\}, \tag{32}$$

$$D_t^* = \max\left\{0, 0.02C_t^* + \eta_t^{(d)}\right\}, \tag{33}$$

$$M_t = 0.12Z_t + \eta_t^{(m)}, \tag{34}$$

$$P_t^* = 100 \exp\left(\sum_{u=0}^{t} 0.01M_u\right), \tag{35}$$

where,

$\eta_t^{(c)} \sim \mathcal{N}(0, (0.25\sigma)^2)$, $\eta_t^{(d)} \sim \mathcal{N}(0, (0.40\sigma)^2)$, and $\eta_t^{(m)} \sim \mathcal{N}(0, (0.45\sigma)^2)$.

Two regular weekly observation schedules are considered: Monday–Friday and Sunday–Thursday. Five gap mechanisms are evaluated. The *scheduled* condition includes regular closures only. *Irregular holiday* gaps are represented by additional independent gaps with probability 0.025, while *extended closures* consist of two additional closures lasting 5–10 calendar days. *Random gaps* are generated independently with probability 0.10. *State-dependent gaps* are generated with probability 0.18 when $S_t = 1$. The state-dependent mechanism is informative by construction because S_t also enters the latent process Z_t. The mechanisms are not calibrated to equalize the total number, frequency, or duration of omitted observations. Their contrasts therefore reflect differences in both gap structure and information loss.

After imposing each observation schedule, financial values are retained only on observed dates and the corresponding Calendar and Session representations are reconstructed using the same definitions as in the empirical experiment. One hundred simulation seeds are evaluated for each weekly-schedule and gap-mechanism combination, with three noise levels represented across seeds. Forecasting uses fixed-duration matching, a 28-day historical interval, a seven-day forecast interval, and a seven-day stride.

The simulation target is the mean future latent value,

$$\bar{Z}_t^{(7)} = \frac{1}{7} \sum_{\tau \in \mathcal{H}_t} Z_\tau. \tag{36}$$

Forecast dates are divided chronologically into the first 60% for model fitting, an intermediate 20% excluded from fitting, and the final 20% for testing. Ridge regression with $\alpha = 1$ predicts the continuous latent target.

Simulation analyses focus on the Calendar–Session differences in latent-target RMSE for C0–S0 and C2–S1. Comparisons are paired by simulation seed, weekly schedule, gap mechanism, and noise level. Because the same simulation seeds are reused across observation conditions, uncertainty is estimated with 2,000 seed-cluster bootstrap replicates, sampling simulation seeds with replacement while retaining all associated scenario observations.

To test whether the way observation gaps are generated changes the Calendar–Session difference relative to predictable scheduled closures, each additional gap condition k is compared directly with the scheduled reference:

$$A_k = \Delta_k - \Delta_{\text{scheduled}}, \tag{37}$$

where, Δ_k is the Calendar–Session difference in latent-target RMSE under gap condition k. These comparisons between each gap condition and scheduled closures are matched by simulation seed, weekly schedule, and noise level and use the same seed-cluster bootstrap procedure.

G. Sensitivity Analyses and Reproducibility

Three sensitivity analyses assess whether the main representation patterns depend on neural training duration or overlap among successive target dates. First, the C2 and S1 representations, which include elapsed-time information, are evaluated with the standard GRU and the GRU with learned hidden-state decay using a longer training budget of 200 epochs, early-stopping patience of 15, and five seeds. Second, a 14-calendar-day target stride reduces overlap among successive historical windows. Third, a 35-calendar-day stride produces non-overlapping combined historical-input and target intervals under fixed-duration matching.

The stride-based analyses evaluate both historical-input matching rules using C0, C2, S0, and S1 with Ridge and GRU models; the decay-based GRU is additionally evaluated for C2 and S1. These analyses are kept separate from the primary seven-day-stride experiment and are used to assess the stability of the main representation findings. A reproduction package contains the processed datasets, analysis code, experimental configurations, prediction-level outputs, simulation results, and sensitivity-analysis outputs.

IV. Results

A. Forecasting Context and Baseline Performance

The two matching rules included different amounts of historical calendar time and different numbers of observed sessions. Under fixed-sequence-length matching, the elapsed span from the earliest of the 20 Session positions to the forecast target averaged 28.99 calendar days in Germany and 29.25 days in Saudi Arabia, with medians of 28 days and ranges of 28–34 and 28–35 days, respectively. Under fixed-duration matching, the common 28-day historical interval contained an average of 19.57 observed sessions in Germany and 19.13 in Saudi Arabia, with respective ranges of 16–20 and 15–20 sessions. Thus, fixed-sequence-length matching varied the calendar span covered by the historical input, whereas fixed-duration matching varied the number of Session positions.

The simple forecasting baselines also established a demanding reference level for the learned models. As shown in Table IV, the target-specific naive baseline achieved the lowest aggregate mean standardized RMSE in both countries, with values of 0.541 in Germany and 0.514 in Saudi Arabia. The corresponding lowest observed learned-model values were 0.798–0.799 in Germany and 0.669–0.686 in Saudi Arabia, depending on the matching rule. Across the complete primary analysis, the target-specific naive baseline retained the lower aggregate error in all 68 core learned-configuration comparisons, with every paired moving-block-bootstrap interval for the difference between the learned model and the naive baseline above zero. The same ordering extended to the target level, where the target-specific naive forecast had lower observed standardized RMSE in all 272 learned-configuration–target comparisons. The strong naive-baseline performance may reflect the short forecast horizon and persistence in some targets, while the zero cumulative-return forecast also provides a demanding reference for a noisy return target. These mechanisms were not isolated experimentally and are therefore treated only as plausible explanations.

B. Model-Dependent Calendar–Session Representation Effects

The primary C0–S0 comparison revealed a marked dependence on model class. Fig. 2 presents the Calendar–Session differences in mean standardized RMSE, with negative values indicating lower Calendar error. Ridge showed a Calendar-favoring effect in every country and matching-rule combination. The differences ranged from -0.031 to -0.046 in Germany and from -0.052 to -0.076 in Saudi Arabia, with all four 95% uncertainty intervals below zero. MLP followed the same overall profile, although the Saudi estimate under fixed-duration matching was smaller (-0.010) and its interval crossed zero. Under fixed-sequence-length matching, the MLP differences reached -0.028 in Germany and -0.050 in Saudi Arabia.

GRU showed a different pattern. Its aggregate Calendar–Session point estimates were small in magnitude: the German estimates were slightly Session-favoring under both matching rules ($+0.010$ and $+0.009$), whereas the Saudi estimates were slightly Calendar-favoring (-0.005 and -0.008). Most corresponding 95% uncertainty intervals spanned zero, so these estimates provide limited evidence of a consistent directional representation effect for GRU under the primary training protocol. The direction of the effect also varied across folds and neural seeds in the German setting.

Direct comparisons between model-specific representation effects quantified the differences between models visible in Fig. 2. Table V reports the difference between each model's C0–S0 effect and the corresponding GRU effect. The Ridge-minus-GRU differences ranged from -0.041 to -0.068, and all four uncertainty intervals remained below zero. MLP also produced a more Calendar-favoring effect than GRU in both German settings and in Saudi Arabia under fixed-sequence-length matching. The Saudi MLP-minus-GRU difference under fixed-duration matching was smaller at -0.005, with an interval of $[-0.026, 0.013]$. The empirical representation effect is therefore most pronounced for Ridge, intermediate and setting-dependent for MLP, and smaller in estimated magnitude for GRU, for which the uncertainty intervals also provide weaker evidence of a consistent directional effect under the primary training protocol.

C. Target and Design Dependence of Representation Effects

Target-specific analysis showed that Calendar–Session differences varied across prediction targets, models, and countries. Fig. 3 summarizes the target-specific standardized-RMSE differences. In Germany, the clearest Ridge effect under fixed-duration matching occurred for deaths (-0.085, 95% interval $[-0.134, -0.033]$), while GRU showed the opposite direction for cases ($+0.040$, $[0.000, 0.083]$). Under fixed-sequence-length matching, the German Ridge differences for cases and deaths were -0.081 and -0.093, respectively, whereas the GRU case difference remained Session-favoring at $+0.043$.

Saudi Arabia showed stronger Calendar-favoring effects for several epidemiological targets, particularly under fixed-sequence-length matching. Ridge differences for cases and deaths were -0.082 and -0.064, while the corresponding MLP differences were -0.051 and -0.127. The pattern also extended beyond the epidemiological outcomes: Saudi Ridge produced a Calendar-favoring realized-volatility difference of -0.109 under fixed-sequence-length matching, with a 95% interval of $[-0.207, -0.017]$. The Calendar–Session effects therefore differed across individual targets and did not follow a simple division between epidemiological and financial outcomes.

The difference between the fixed-duration and fixed-sequence-length C0–S0 effects was close to zero for GRU in Germany (0.0004) and Saudi Arabia (0.003). In Saudi Arabia, the difference between the two matching rules was larger for Ridge (0.024, 95% interval $[0.006, 0.045]$) and MLP (0.040, $[0.017, 0.073]$), indicating a stronger Calendar-favoring effect under fixed-sequence-length matching. The corresponding German differences for Ridge and MLP were 0.014 and 0.010, with both uncertainty intervals spanning zero. The matching rule therefore changed representation effects mainly for Ridge and MLP in the Saudi setting.

Of the 36 C1–C0, C2–C1, and S1–S0 comparisons, 34 uncertainty intervals included zero. The only two intervals that excluded zero showed small increases in mean standardized RMSE: $+0.010$ after adding Calendar observation indicators to Ridge under German fixed-duration matching and $+0.003$ after adding the Session gap to Ridge under Saudi fixed-duration matching. Adding observation indicators or elapsed-time variables therefore changed some individual configurations but did not consistently improve performance across models, countries, or matching rules.

D. Observation-Gap Mechanisms in Controlled Simulation

Under regular scheduled closures alone, the Calendar–Session differences in latent-target RMSE were close to zero for both representation pairs: -0.003 with a 95% seed-cluster-bootstrap interval of $[-0.080, 0.076]$ for C0–S0 and -0.006 with an interval of $[-0.077, 0.078]$ for C2–S1. Fig. 4 shows the change produced by each additional gap condition relative to scheduled closures.

Compared with scheduled closures, each additional gap condition shifted the Calendar–Session difference in a more Calendar-favoring direction. For C0–S0, the changes relative to scheduled closures were -0.141 for irregular holiday gaps, -0.135 for extended closures, -0.307 for random gaps, and -0.245 for state-dependent gaps. C2–S1, in which both representations include elapsed-time information, showed the same broad ordering, with respective changes of -0.197, -0.113, -0.357, and -0.251. All eight seed-cluster-bootstrap uncertainty intervals were below zero. Random gaps produced the largest change for both representation pairs, followed

TABLE IV. Aggregate Forecasting Performance of Simple Baselines and the Lowest Observed Learned Configurations. Values are Mean Standardized RMSE; Lower Values Indicate Lower Forecasting Error. The Learned Minima are Descriptive Test-Set Summaries Rather than Independently Validated Model-Selection Estimates. C2 Denotes Calendar + Indicators + Elapsed Time

Country	Target-specific naive	Previous 7-day aggregate	Training mean	Lowest learned, Fixed duration	Lowest learned, Fixed sequence length
Germany	0.541	0.612	1.032	0.798 (GRU / Session)	0.799 (GRU / Session)
Saudi Arabia	0.514	0.586	1.352	0.686 (Ridge / C2)	0.669 (Ridge / C2)

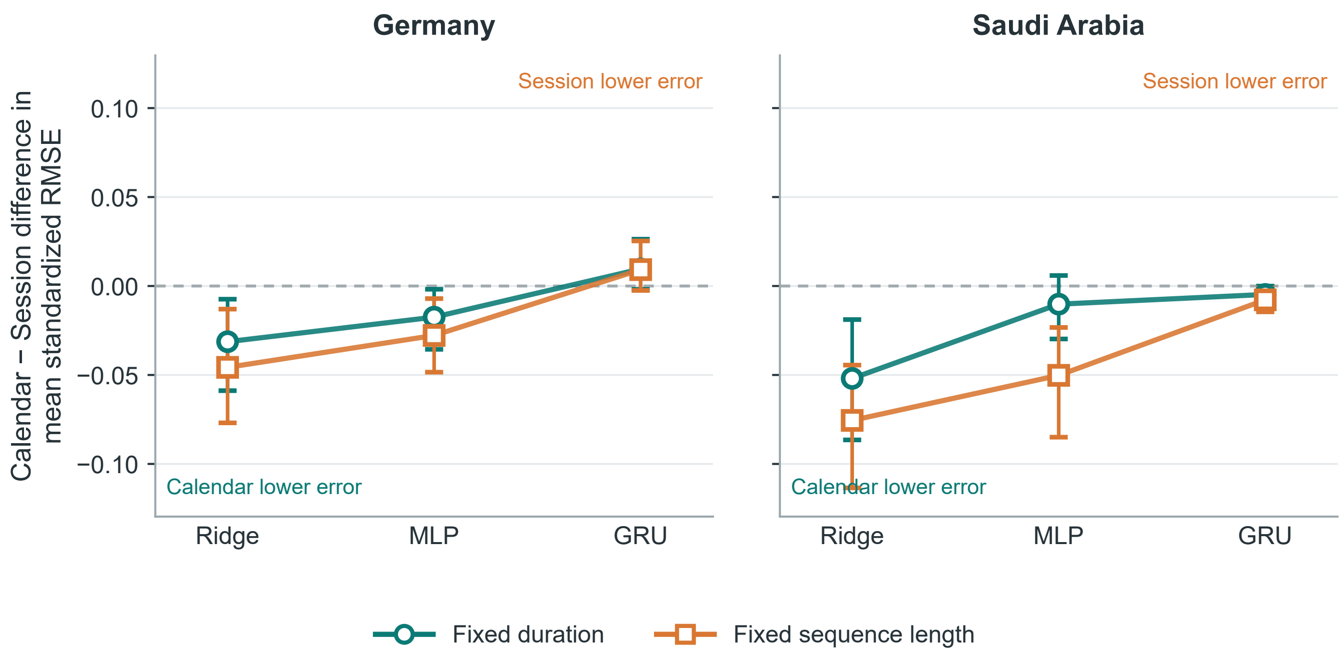

Fig. 2. Model-dependent Calendar–Session representation effects in the primary empirical experiment. Points show the difference in mean standardized RMSE between C0 Calendar and S0 Session, calculated from the original test predictions, and error bars show 95% moving-block-bootstrap uncertainty intervals. Negative values indicate lower Calendar error and positive values indicate lower Session error. Lines connect estimates obtained under the same historical-input matching rule to emphasize changes in representation sensitivity across model classes.

TABLE V. Direct Differences between the C0–S0 Effects of Ridge or MLP and the Corresponding GRU Effect. Negative Values Indicate a More Calendar-Favoring Effect for Ridge or MLP than for GRU. Intervals are 95% Moving-Block-Bootstrap Uncertainty Intervals

		Ridge minus GRU		MLP minus GRU	
Country	Matching rule	Difference from GRU	95% interval	Difference from GRU	95% interval
Germany	Fixed duration	-0.041	[-0.077, -0.011]	-0.027	[-0.052, -0.007]
Germany	Fixed sequence length	-0.055	[-0.097, -0.019]	-0.037	[-0.064, -0.017]
Saudi Arabia	Fixed duration	-0.047	[-0.081, -0.016]	-0.005	[-0.026, 0.013]
Saudi Arabia	Fixed sequence length	-0.068	[-0.100, -0.036]	-0.042	[-0.072, -0.018]

by state-dependent gaps. These results show that the specified additional-gap scenarios produced more Calendar-favoring representation differences than regular scheduled closures, with the largest observed changes under random and state-dependent gaps.

E. Robustness and Sensitivity

Longer neural training preserved the cross-country direction of the C2–S1 recurrent comparisons, in which both representations include elapsed-time information. With the 200-epoch training budget, the C2–S1 differences for the standard GRU and the GRU with hidden-state decay ranged from $+0.028$ to $+0.033$ in Germany, indicating lower Session error, and from -0.028 to -0.040 in Saudi Arabia, indicating lower Calendar error. Values within each country remained similar under fixed-duration and fixed-sequence-length matching.

The reduced-overlap stride analyses showed greater variation as the number of available test target dates decreased. At a 14-day stride, the GRU C0–S0 differences remained small, at approximately $+0.009$ in Germany and between -0.004 and

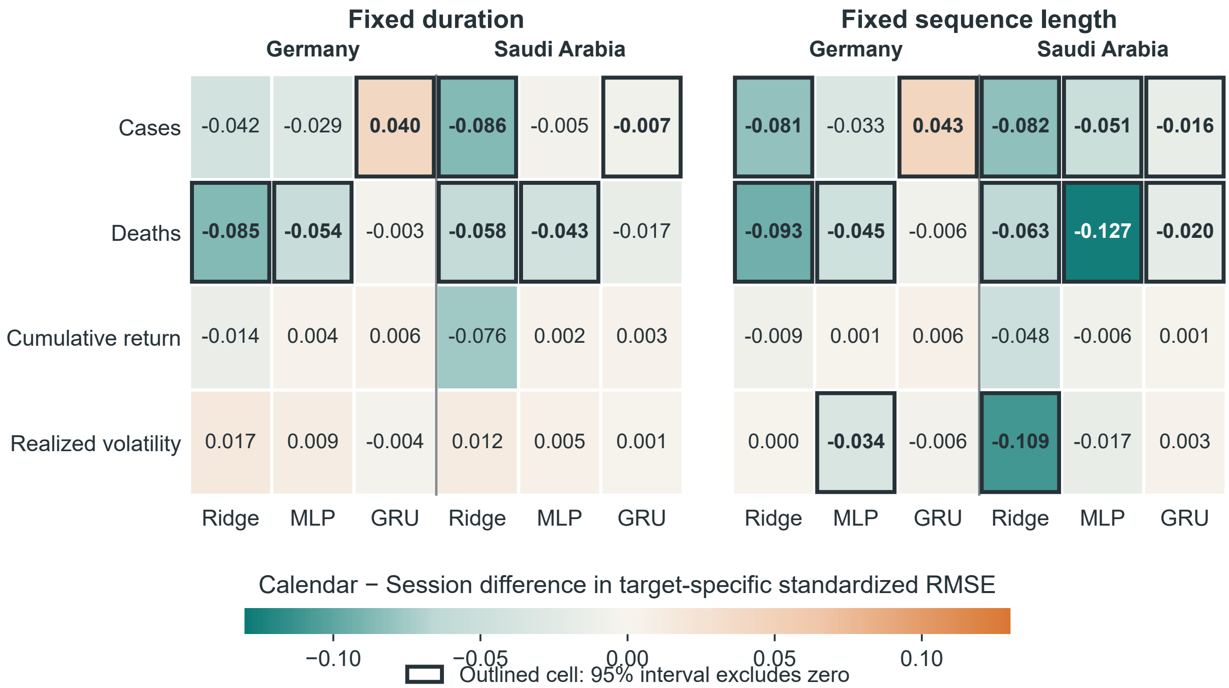

Fig. 3. Target-specific C0–S0 Calendar–Session differences in standardized RMSE under fixed-duration and fixed-sequence-length matching. Cell values and colors encode the signed Calendar–Session difference, with negative values indicating lower Calendar error and positive values indicating lower Session error. Outlined cells have 95% uncertainty intervals that exclude zero.

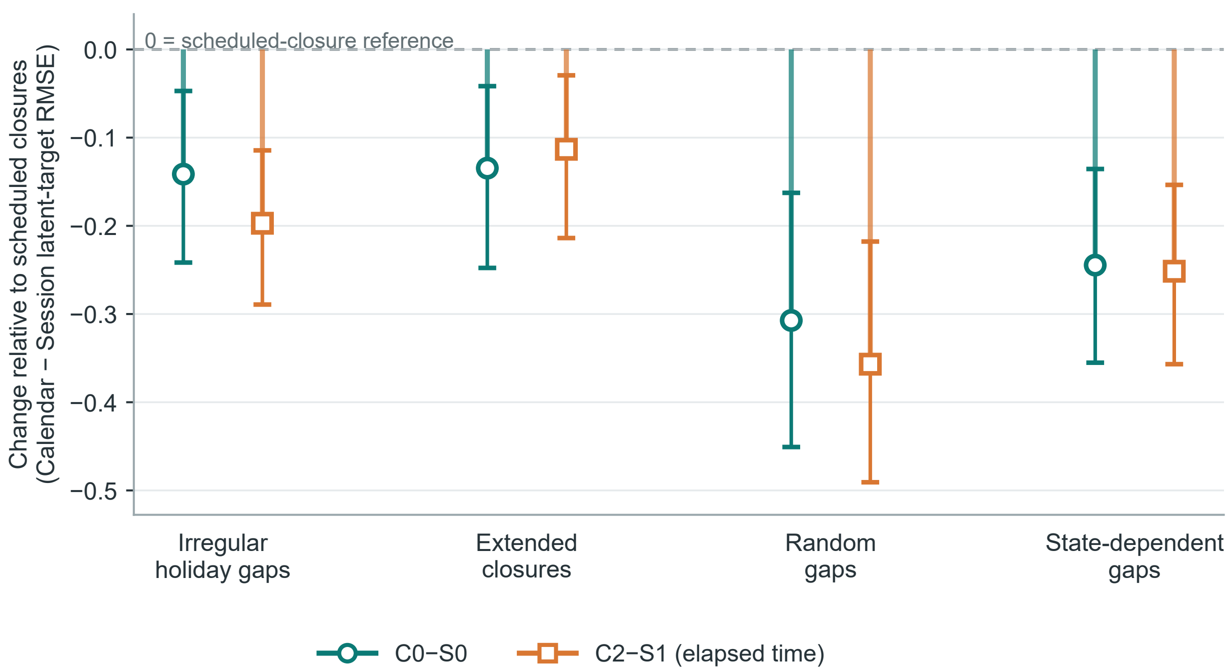

Fig. 4. Changes in Calendar–Session latent-target RMSE differences under additional observation-gap conditions relative to regular scheduled closures. Values show the Calendar–Session RMSE difference for each gap condition minus the corresponding difference under scheduled closures. Negative values indicate a shift toward lower Calendar error relative to the scheduled reference. Error bars show 95% seed-cluster-bootstrap uncertainty intervals.

-0.009 in Saudi Arabia. Ridge remained Calendar-favoring for several conditions, including Saudi fixed-duration matching at -0.045, while the Saudi estimate under fixed-sequence-length matching shifted to $+0.048$. At the 35-day stride, each analysis contained only 12 test target dates, and the point estimates became more variable; Saudi Ridge, for example, produced C0–S0 differences of -0.299 under fixed-duration matching and -0.130 under fixed-sequence-length matching. The stride analyses therefore support greater precision for the primary seven-day-stride estimates while showing that the small GRU differences remain of modest magnitude across several reduced-overlap settings.

Optimization diagnostics provide additional context for the recurrent-model results. Among the primary C0 and S0 GRU runs, 40–42.5% of the German runs and 65% of the Saudi runs reached the maximum training length of 60 epochs. Median training duration was 53–57 epochs for the German C0/S0 configurations and 60 epochs for the corresponding Saudi configurations. Combined with the long-training sensitivity results, these training-duration results support interpreting the model-dependent representation effects as empirical behavior under the specified training protocol.

V. Conclusion

This study shows that the choice between preserving the full calendar and retaining only observed sessions should not be treated as a neutral preprocessing step in mixed-calendar forecasting. Neither representation is universally preferable, and the observed differences depend on how the forecasting model uses the input sequence and, in some settings, on whether fixed-duration or fixed-sequence-length matching is used. Because the target-specific naive baseline outperformed all core learned configurations, the learned-model results should be interpreted primarily as evidence of sensitivity to temporal representation, not as evidence of forecasting superiority. The controlled simulation further shows that representation effects can differ across the specified observation-gap scenarios. The empirical analysis is limited to Saudi Arabia and Germany in a COVID-19 and financial setting, so the direction and magnitude of the observed representation effects should not be assumed to transfer unchanged to other domains, observation calendars, or forecasting horizons. The broader implication is that studies combining variables observed on different calendars should explicitly justify how dates are indexed and how historical inputs are constructed rather than treating alignment as a routine preprocessing choice.

Funding

This research work was funded by Umm Al-Qura University, Saudi Arabia under grant number: 26UQU4340018GSSR04

Acknowledgment

The author extends his appreciation to Umm Al-Qura University, Saudi Arabia for funding this research work through grant number: 26UQU4340018GSSR04

References

[1] N. O'Donnell, D. Shannon, and B. Sheehan, "Immune or at-risk? stock markets and the significance of the COVID-19 pandemic," *Journal of Behavioral and Experimental Finance*, vol. 30, p. 100477, 2021.

[2] M. Brueckner and J. Vespignani, "COVID-19 infections and the performance of the stock market: An empirical analysis for australia," *Economic Papers: A Journal of Applied Economics and Policy*, vol. 40, no. 3, pp. 173–193, 2021.

[3] Š. Lyócsa and P. Molnár, "The effect of non-trading days on volatility forecasts in equity markets," *Finance Research Letters*, vol. 23, pp. 39–49, 2017.

[4] A.-C. Díaz-Mendoza and A. Pardo, "Holidays, weekends and range-based volatility," *The North American Journal of Economics and Finance*, vol. 52, p. 101124, 2020.

[5] Š. Lyócsa and N. Todorova, "Trading and non-trading period realized market volatility: Does it matter for forecasting the volatility of US stocks?" *International Journal of Forecasting*, vol. 36, no. 2, pp. 628–645, 2020.

[6] J. Lau, "Economic sentiment shifts over weekends and their impact on stock returns," *Economics Letters*, vol. 264, p. 112963, 2026.

[7] S. C.-X. Li and B. Marlin, "Learning from irregularly-sampled time series: A missing data perspective," in *Proceedings of the 37th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, vol. 119. PMLR, 2020, pp. 5937–5946.

[8] Z. Che, S. Purushotham, K. Cho, D. Sontag, and Y. Liu, "Recurrent neural networks for multivariate time series with missing values," *Scientific Reports*, vol. 8, no. 1, p. 6085, 2018.

[9] Y. Rubanova, R. T. Q. Chen, and D. K. Duvenaud, "Latent ordinary differential equations for irregularly-sampled time series," in *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[10] E. De Brouwer, J. Simm, A. Arany, and Y. Moreau, "GRU-ODE-Bayes: Continuous modeling of sporadically-observed time series," in *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[11] P. Kidger, J. Morrill, J. Foster, and T. Lyons, "Neural controlled differential equations for irregular time series," in *Advances in Neural Information Processing Systems*, vol. 33, 2020, pp. 6696–6707.

[12] M. Schirmer, M. Eltayeb, S. Lessmann, and M. Rudolph, "Modeling irregular time series with continuous recurrent units," in *Proceedings of the 39th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, vol. 162. PMLR, 2022, pp. 19 388–19 405.

[13] W. Zhang, C. Yin, H. Liu, X. Zhou, and H. Xiong, "Irregular multivariate time series forecasting: A transformable patching graph neural networks approach," in *Proceedings of the 41st International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, vol. 235. PMLR, 2024, pp. 60 179–60 196.

[14] B. Li, Y. Luo, Z. Liu, J. Zheng, J. Lv, and Q. Ma, "HyperIMTS: Hypergraph neural network for irregular multivariate time series forecasting," in *Proceedings of the 42nd International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, vol. 267. PMLR, 2025, pp. 35 502–35 518.

[15] Y. Luo, B. Zhang, Z. Liu, and Q. Ma, "Hi-patch: Hierarchical patch GNN for irregular multivariate time series," in *Proceedings of the 42nd International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, vol. 267. PMLR, 2025, pp. 41 494–41 519.

[16] Z. Hu, J. Wu, H. Xu, M. Liao, N. Feng, B. Gao, S. Lai, and Y. Yue, "IMTS is worth time $\times$ channel patches: Visual masked autoencoders for irregular multivariate time series prediction," in *Proceedings of the 42nd International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, vol. 267. PMLR, 2025, pp. 24 805–24 822.

[17] X. Liu, X. Qiu, X. Wu, Z. Li, C. Guo, J. Hu, and B. Yang, "Rethinking irregular time series forecasting: A simple yet effective baseline," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 40, no. 28, 2026, pp. 23 873–23 881.

[18] Z. Zhou, Y. Huang, Y. Wang, Y. Wu, J. Kwok, and Y. Liang, "Revitalizing canonical pre-alignment for irregular multivariate time series forecasting," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 40, no. 34, 2026, pp. 29 115–29 123.

[19] T. Vanderschueren, A. Curth, W. Verbeke, and M. Van Der Schaar, "Accounting for informative sampling when learning to forecast treatment outcomes over time," in *Proceedings of the 40th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, vol. 202. PMLR, 2023, pp. 34 855–34 874.

[20] C. Chang, J. Hwang, Y. Shi, H. Wang, W. Wang, W.-C. Peng, and T.-F. Chen, "Time-IMM: A dataset and benchmark for irregular multimodal multivariate time series," in *Advances in Neural Information Processing Systems*, vol. 38, 2025, datasets and Benchmarks Track.

[21] King Abdullah Petroleum Studies and Research Center (KAPSARC), "Saudi arabia coronavirus disease (COVID-19) situation," KAPSARC Data Portal, 2023. [Online]. Available: https://datasource.kapsarc.org/explore/assets/saudi-arabia-coronavirus-disease-covid-19-situation/

[22] R. Aroussi, "yfinance: Download market data from Yahoo! Finance's API," Python package, version 0.2.66, accessed: June 13, 2026. [Online]. Available: https://github.com/ranaroussi/yfinance

[23] Yahoo Finance, "Yahoo finance," Online financial market data service, accessed: June 13, 2026. [Online]. Available: https://finance.yahoo.com/

[24] Robert Koch-Institut, "SARS-CoV-2 infektionen in deutschland," 2026, data set.

[25] Deutsche Börse AG, "Xetra trading calendar archive," Deutsche Börse Cash Market. [Online]. Available: https://www.cashmarket.deutsche-boerse.com/cash-en/trading/trading-calendar-and-trading-hours/trading-calendar

[26] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder–decoder for statistical machine translation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar: Association for Computational Linguistics, 2014, pp. 1724–1734.

[27] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations (ICLR)*, 2015, arXiv:1412.6980.

[28] H. R. Künsch, "The jackknife and the bootstrap for general stationary observations," *The Annals of Statistics*, vol. 17, no. 3, pp. 1217–1241, 1989.

[29] D. N. Politis and J. P. Romano, "A circular block-resampling procedure for stationary data," in *Exploring the Limits of Bootstrap*, R. LePage and L. Billard, Eds. John Wiley & Sons, 1992, pp. 263–270.