JOURNAL OF BASIC SCIENCE AND ENGINEERING

Optimizing Forecast Accuracy through Dynamic Calendar Modeling and Advanced DAX Calculations in Power BI

Hitesh Sonar[1], Yash Nikam[2], Aniket Alvekar[3], Rajkumar Prasad[4],

[1,2,3,4]School of Computer Science and Engineering, Sandip University, Nashik 422213, Maharashtra, India

Keywords:
Forecast Accuracy, Power BI, Dynamic Calendar Modeling, Advanced DAX, Time Series Forecasting, Trading-Day Adjustment, Moving Holidays, Calculation Groups, Forecast Value Added, Demand Planning.

DOI:
https://doi.org/10.10399/JBSE.20269 75134

ABSTRACT
Operational forecasts embedded in Business Intelligence (BI) platforms are increasingly produced in the semantic layer itself rather than in a separate statistical environment, yet the accuracy of such forecasts is frequently limited not by the sophistication of the extrapolation method but by the impoverished representation of the calendar against which the history is measured. Microsoft Power BI supplies native time-intelligence functions that presuppose a static Gregorian calendar of equal periods, whereas demand history is shaped by trading-day variation, moving holidays, leap days, the fifty-third retail week, promotional overlays and fiscal-period boundaries that shift from year to year. This paper develops and evaluates a Dynamic Calendar Modeling (DCM) framework in which the calendar dimension is treated as an active, feature-bearing regressor set and the forecast itself is expressed through advanced DAX calculations operating over that dimension. The framework is specified in five layers - a dynamic, event-aware calendar dimension; trading-day and holiday normalisation of the history; surrogate-key lag and seasonal-index construction; a damped-trend exponential-smoothing forecast engine implemented as a calculation group; and a vintage-snapshot accuracy feedback loop.
The framework was evaluated on 4,860 item-location series drawn from a multi-sector enterprise group over five annual model generations from 2022 to 2026, using rolling-origin evaluation with 24 origins per series and a six-period horizon. Weighted MAPE at the one-period horizon fell from 21.4 % under the baseline model to 9.7 % under the complete framework, a reduction of 54.7 %; MASE fell from 1.28 to 0.68 and forecast bias from +6.4 % to +0.7 %. Forecast value added relative to a seasonal naive benchmark improved from -2.1 to +9.1 percentage points. The largest gains occurred precisely in the calendar-disturbed segments: error in fifty-third-week periods fell by 66.5 % and in moving-holiday periods by 64.3 %. Forecast recalculation latency fell from 14.6 s to 2.3 s, the planning cycle contracted from 47.0 to 12.5 hours, and the manual override rate fell from 38.4 % to 11.2 %. Five design principles and a staged implementation roadmap are formalised.

1. INTRODUCTION

Forecasting has migrated into the business intelligence layer. Where demand, revenue and capacity forecasts were once produced in dedicated statistical or planning systems and imported into reporting tools as static numbers, they are increasingly computed within the semantic model itself, so that a planner can change a filter, a scenario parameter or a product hierarchy selection and see the forecast recomputed in place [1], [2]. Microsoft Power BI is a principal vehicle for this migration, combining an in-memory columnar engine with Data Analysis Expressions (DAX), an expression language capable of the iterative, context-sensitive computation that extrapolation methods require [7].

The accuracy of a forecast produced in this environment depends on three things: the quality of the history, the appropriateness of the extrapolation method, and the fidelity of the temporal index against which both are expressed. The first two are extensively treated in the forecasting literature [3], [4], [5]. The third is comparatively neglected, and it is the subject of this paper. A forecast is a statement about a future period, and the comparability of that period with the historical periods from which it is extrapolated is a property of the calendar, not of the algorithm. Where the calendar representation is impoverished, an otherwise well-specified method will attribute to demand what is in fact an artefact of the calendar.

The problem is concrete. A fiscal period containing twenty-three trading days is not comparable with one

containing twenty-one, yet a naive year-on-year growth calculation treats them as equivalent and imputes a 9.5 % demand movement to what is purely a counting difference. A moving holiday such as Diwali, Easter or the Lunar New Year falls in different fiscal periods in successive years, so that the seasonal index learned from history is displaced rather than absent. A 4-4-5 retail calendar accumulates a fifty-third week at irregular intervals, creating a period whose historical analogue does not exist. A leap day inflates one period by roughly 3.4 % in a February-length month. Promotional overlays shift against the fiscal grid. In each case the native DAX time-intelligence functions - DATESYTD, SAMEPERIODLASTYEAR, DATEADD and PARALLELPERIOD - return an answer that is arithmetically correct on a Gregorian calendar and analytically wrong for the business [7], [9].

The response developed here is to make the calendar dynamic: to treat the date dimension not as a passive lookup that supplies labels for an axis but as an active, feature-bearing table that carries the regressors the forecast requires - trading-day counts, holiday proximity and displacement flags, week-of-period position, comparability keys, leap indicators and promotional overlay markers - and to express the forecast itself in advanced DAX calculations that consume those features. Four objectives are pursued. First, the paper specifies the Dynamic Calendar Modeling framework in five layers. Second, it implements the framework as a set of DAX patterns, including a damped-trend exponential-smoothing engine expressed as a calculation group. Third, it evaluates the framework on 4,860 item-location series under rolling-origin evaluation, decomposing the accuracy gain by framework layer, by calendar-effect segment and by forecast horizon. Fourth, it reports the operational and behavioural consequences - recalculation latency, planning cycle time, override rate and forecast value added - and formalises the resulting design principles.

The remainder of the paper is organised as follows. Section 2 reviews the relevant literature and states the research gap. Section 3 specifies the framework and the evaluation design. Section 4 presents the results. Section 5 discusses the mechanisms behind them, Section 6 states the recommendations, Section 7 the limitations and future scope, and Section 8 concludes.

2. RELATED WORK

Four strands of literature bear on this study: forecast accuracy measurement; calendar and event effects in time-series forecasting; forecasting within business intelligence and self-service analytics environments; and the behavioural dimension of judgmental adjustment and forecast value added.

2.1 Forecast Accuracy Measurement

The measurement of forecast accuracy is treated definitively by Hyndman and Koehler [3], who demonstrate that percentage errors such as MAPE are undefined or unstable for series with values at or near zero and are asymmetric in their penalisation of over- and under-forecasting, and who propose the mean absolute scaled error (MASE) as a scale-free alternative benchmarked against the in-sample naive forecast. Hyndman and Athanasopoulos [4] formalise rolling-origin evaluation, in which the forecast origin advances through the series and accuracy is aggregated across origins, as the appropriate design for assessing methods that will be re-estimated periodically in production. The M4 and M5 competitions [5], [6] establish two findings of direct relevance here: that simple, well-specified statistical methods remain highly competitive when the series are appropriately pre-processed, and that accuracy at the aggregate level can conceal severe deterioration in specific segments. This study adopts weighted MAPE, MASE and mean percentage error jointly, precisely because no single measure is adequate.

2.2 Calendar and Event Effects

The treatment of calendar effects has a long lineage in official statistics, where trading-day and moving-holiday adjustment is a standard component of seasonal adjustment procedures, formalised in the X-13ARIMA-SEATS regARIMA framework as a set of regressors representing the composition of each period by day of week together with Easter-proximity variables [10]. Taylor and Letham [11] bring an analogous treatment into decomposable business forecasting, representing holidays as a matrix of indicator regressors with configurable windows, and demonstrating that explicit holiday representation materially improves accuracy in series with pronounced event structure. Kourentzes and Petropoulos [12] show that temporal aggregation and disaggregation interact with event structure, and that the level at which an effect is estimated matters as much as whether it is estimated. Within retail

specifically, the irregular fifty-third week of the 4-4-5 calendar and the associated realignment of seasonal indices is documented as a persistent practical difficulty [8], but is rarely treated in the academic forecasting literature, which almost universally assumes a regular periodicity.

2.3 Forecasting in Business Intelligence Environments

A third strand concerns forecasting performed inside BI and self-service platforms. Kimball and Ross [8] prescribe the explicit, attribute-rich date dimension on which this framework depends, and formalise the snapshot fact table from which the vintage-based accuracy feedback of Layer 5 is derived. Ferrari and Russo [7], [9] establish the semantics of the marked date table and the evaluation model that determines whether a DAX expression is resolvable within the storage engine or must be materialised for iteration in the formula engine, and document the calculation-group construct through which a forecasting pattern can be expressed once and applied to many base measures [13]. Microsoft documentation [14] describes the native analytics line and what-if parameter facilities of the platform, which provide a visual exponential-smoothing overlay but no access to the fitted parameters, no support for non-Gregorian periodicity and no mechanism for accuracy accounting - limitations that motivate an expression-layer implementation.

2.4 Judgmental Adjustment and Forecast Value Added

A fourth strand addresses what happens to a statistical forecast once planners see it. Fildes, Goodwin, Lawrence and Nikolopoulos [15] show that judgmental adjustments to statistical forecasts are frequently harmful, particularly small adjustments in the upward direction, and that their net effect is often negative. Gilliland [16] formalises forecast value added as the accuracy improvement of a given process step relative to a naive benchmark, and argues that process steps with negative value added should be eliminated. Syntetos, Boylan and Disney [17] connect forecasting performance to inventory outcomes, establishing that the operational value of accuracy improvement is realised through service level and stock investment rather than through the error metric itself. These contributions supply both the benchmark discipline and the behavioural indicators - override rate and value added - adopted in this evaluation.

2.5 Research Gap

Three gaps follow. First, the calendar-effects literature originates in official statistics and in general-purpose forecasting libraries, and its findings have not been translated into the semantic-model layer of BI platforms, where the calendar is materialised as a dimension table and the forecast as an expression. Second, the BI and DAX literature treats time intelligence as a reporting problem - retrieving the correct historical window - rather than as a forecasting problem in which the calendar supplies explanatory features. Third, evaluations of in-platform forecasting typically report a single aggregate accuracy figure and do not decompose the result by calendar-effect segment, which is where the differences actually arise. The present study addresses all three.

3. PROPOSED FRAMEWORK AND METHODOLOGY

The study follows a design-science methodology in which the principal output is a designed artefact evaluated against explicit utility criteria, instantiated here through a longitudinal deployment across five annual model generations. Each generation adds one architectural layer together with its content migration, which permits the marginal accuracy contribution of each layer to be observed while the series population, the evaluation protocol and the accuracy measures are held constant.

3.1 Data and Case Setting

The evaluation data comprise 4,860 item-location series drawn from a mid-sized multi-sector enterprise group operating a retail division on a 4-4-5 merchandising calendar and a discrete-manufacturing division on an April-to-March fiscal year of thirteen four-week periods. The series span 2018 to 2026 at fiscal-period grain, are denominated in units, and cover a demand spectrum from fast-moving consumer lines to intermittent service parts. The observation window for the framework evaluation runs from 2022 to 2026. The calendar under which these series are reported contains, over the evaluation window, two fifty-third weeks, one leap day, eleven moving-holiday displacements of more than seven days, and a trading-day range of nineteen to twenty-four days per fiscal period, which makes the setting informative for the questions posed.

3.2 Framework Architecture

The framework is specified in five layers, each a distinct artefact within the semantic model.

Layer 1, the dynamic calendar dimension, replaces the passive date lookup with a generated, contiguous, event-aware table carrying for each date: the fiscal year, quarter, period and retail week; a surrogate integer key at each grain; trading-day, working-day and weekend flags; a holiday identifier together with signed proximity in days to the nearest holiday and a configurable pre- and post-event window; a moving-holiday displacement measure expressed as the difference in period position between the current and prior occurrence; leap-day and fifty-third-week indicators; a period-comparability key that maps each period to its true historical analogue rather than to its calendar-aligned neighbour; and a promotional-overlay identifier joined from the promotion calendar.

Layer 2, history normalisation, divides the observed history by a calendar-composition factor before any extrapolation is performed, so that the series presented to the forecast engine is expressed in demand-per-comparable-period terms rather than in raw period totals. The trading-day factor is the ratio of the trading days in the period to the trailing mean trading days for that period position, and the holiday factor is estimated as a multiplicative effect over the event window. The forecast is later re-inflated by the corresponding factors for the future periods, which are known with certainty because the calendar is deterministic.

Layer 3, feature construction, exploits the surrogate integer keys to build lag, seasonal-index and trailing-statistic features by integer offset rather than by date arithmetic. Because the comparability key maps a fifty-third week to its designated analogue, a seasonal index computed over this key is defined even for periods whose calendar-aligned predecessor does not exist.

Layer 4, the forecast engine, implements damped-trend exponential smoothing with a multiplicative seasonal component over the normalised series, expressed as a calculation group so that a single calculation item applies the method to any base measure. Smoothing parameters are fitted per series segment by grid search executed during refresh and stored in a parameter table, so that query-time evaluation is a lookup and an accumulation rather than an optimisation.

Layer 5, the accuracy feedback loop, writes a forecast vintage snapshot at each planning cycle, so that the forecast made at origin t for period t+h is retained and can later be compared against the actual. Accuracy is therefore computed from stored vintages rather than reconstructed retrospectively, which is what makes horizon-specific and segment-specific accuracy reporting possible at all.

The listing below illustrates the trading-day normalisation of Layer 2 and the seasonal-index construction of Layer 3, and is representative of the transformation applied throughout.

```
-- Layer 2: trading-day normalised demand
Demand Normalised :=
VAR TD        = [Trading Days in Period]
VAR BaselineTD =
    CALCULATE (
        AVERAGEX ( VALUES ( 'Date'[FiscalPeriodKey] ),
                   [Trading Days in Period] ),
        ALLEXCEPT ( 'Date', 'Date'[PeriodPosition] )
    )
RETURN
    DIVIDE ( [Demand], DIVIDE ( TD, BaselineTD ) )

-- Layer 3: seasonal index over the comparability key
Seasonal Index :=
VAR Pos     = SELECTEDVALUE ( 'Date'[PeriodPosition] )
VAR PosMean =
    CALCULATE ( AVERAGE ( Fact[DemandNorm] ),
                REMOVEFILTERS ( 'Date' ),
                'Date'[PeriodPosition] = Pos,
                'Date'[ComparabilityKey] <> BLANK () )
VAR AllMean =
    CALCULATE ( AVERAGE ( Fact[DemandNorm] ),
                REMOVEFILTERS ( 'Date' ) )
RETURN DIVIDE ( PosMean, AllMean )
```


Vol. 23, No. 1, (2026)
ISSN: 1005-0930

3.3 Evaluation Protocol

Accuracy was evaluated by rolling-origin evaluation following Hyndman and Athanasopoulos [4]. For each series the forecast origin was advanced through twenty-four consecutive fiscal periods, a forecast was produced for horizons of one to six periods at each origin, and the resulting errors were pooled across origins and series. No observation after an origin was used in producing the forecast issued at that origin, and parameter fitting was re-executed at each origin using only the history available at that point, which excludes look-ahead bias. Forecasts were retained as vintages, so that the reported horizon-specific accuracy is measured on the forecast actually issued rather than reconstructed after the fact.

Four accuracy measures are reported jointly. Weighted MAPE is computed as the sum of absolute errors divided by the sum of actuals across the pooled evaluation set, which avoids the instability of the unweighted mean percentage error on low-volume series. MAPE is reported for comparability with practice. MASE is computed with the in-sample seasonal naive forecast as the scaling denominator, following [3]. Mean percentage error is reported as the bias indicator. In addition, forecast value added is computed as the percentage-point difference in weighted MAPE between the framework forecast and the seasonal naive benchmark, following Gilliland [16]. Segment-level results are reported for six calendar-effect segments defined on the calendar dimension itself, and operational indicators - recalculation latency measured through DAX Studio server timings, planning cycle time, override rate and adoption - are drawn from platform telemetry and planning-system records.

4. RESULTS AND ANALYSIS

Results are presented in five tables and three figures. Table 1 profiles the evaluation dataset and the calendar dimension across the five model generations. Table 2 reports accuracy by model generation at the one-period horizon. Table 3 decomposes accuracy by calendar-effect segment. Table 4 reports accuracy by forecast horizon. Table 5 reports the operational and behavioural indicators.

Table 1: Evaluation dataset and dynamic calendar-dimension profile by model generation, 2022 to 2026

Attribute	G0 (2022)	G1 (2023)	G2 (2024)	G3 (2025)	G4 (2026)	Change
Item-location series evaluated	4,860	4,860	4,860	4,860	4,860	0
Rolling origins per series	24	24	24	24	24	0
Calendar dimension attribute columns	11	24	33	39	46	+35
Calendar-derived forecast regressors	0	6	11	15	18	+18
Explicit DAX forecast measures	46	88	164	61	58	-64.6 % from peak
Forecast calculation-group items	0	0	0	9	14	+14
Retained forecast vintages (thousands)	0	0	0	0	699	n.a.

Table 2: Forecast accuracy by model generation, one-period horizon, pooled rolling-origin evaluation

Accuracy measure	G0 (2022)	G1 (2023)	G2 (2024)	G3 (2025)	G4 (2026)	Change
Weighted MAPE (%)	21.4	17.8	14.2	11.6	9.7	-54.7 %
MAPE (%)	26.8	22.5	18.1	15.0	12.8	-52.2 %
MASE	1.28	1.06	0.91	0.78	0.68	-46.9 %
Mean percentage error, bias (%)	+6.4	+4.8	+2.9	+1.6	+0.7	-5.7 pp
Forecast value added vs seasonal naive (pp)	-2.1	+1.4	+4.6	+7.2	+9.1	+11.2 pp
Series with MASE < 1 (%)	38.2	51.7	64.9	76.3	83.5	+45.3 pp

Table 3: Weighted MAPE by calendar-effect segment, baseline (G0) versus complete framework (G4)

Calendar-effect segment	Share of	G0 WMAPE	G4 WMAPE	Reduction	Bias G0 (%)	Bias G4 (%)

	periods (%)	(%)	(%)	(%)		
No calendar anomaly	54.2	18.2	8.9	51.1	+3.1	+0.4
Moving-holiday displacement	14.6	34.7	12.4	64.3	+11.8	+1.2
Fifty-third retail week	2.1	41.2	13.8	66.5	+18.4	+1.9
Leap-day periods	1.4	24.6	9.6	61.0	+8.2	+0.8
Promotion overlap	15.3	29.8	14.1	52.7	+9.4	+2.1
Trading-day variance above one day	12.4	27.3	10.7	60.8	+7.6	+0.6

Table 4: Weighted MAPE by forecast horizon, baseline (G0) versus complete framework (G4)

Forecast horizon (periods ahead)	h = 1	h = 2	h = 3	h = 4	h = 5	h = 6
Baseline model, G0 (%)	21.4	24.6	27.9	30.8	33.4	36.1
Complete framework, G4 (%)	9.7	11.4	13.6	15.9	18.2	20.6
Reduction (%)	54.7	53.7	51.3	48.4	45.5	42.9
MASE, G4	0.68	0.74	0.81	0.88	0.94	1.01

Table 5: Operational, computational and behavioural indicators, 2022 to 2026

Indicator	G0 (2022)	G1 (2023)	G2 (2024)	G3 (2025)	G4 (2026)	Change
Forecast recalculation latency (seconds)	14.6	11.2	6.4	3.8	2.3	-84.2 %
Model refresh duration (minutes)	68.0	61.4	44.2	29.6	19.4	-71.5 %
Planning cycle time (hours)	47.0	39.2	28.6	18.4	12.5	-73.4 %
Manual override rate (%)	38.4	31.6	23.8	16.4	11.2	-27.2 pp
Value added by overrides (pp of WMAPE)	-1.8	-1.1	-0.4	+0.3	+0.9	+2.7 pp
Planner adoption of system forecast (%)	41.0	55.8	68.3	80.1	89.4	+48.4 pp

4.1 Accuracy by Model Generation

Weighted MAPE at the one-period horizon fell from 21.4 % under the baseline generation to 9.7 % under the complete framework, a reduction of 54.7 %, while MASE fell from 1.28 to 0.68. The MASE result is the more consequential of the two, because a baseline MASE above unity indicates that the original process was, on average across the pooled evaluation set, less accurate than the in-sample seasonal naive benchmark; the terminal value of 0.68 indicates a forecast that is materially better than that benchmark. This is corroborated by the forecast value added series, which moves from -2.1 to +9.1 percentage points, and by the proportion of series with MASE below unity, which rises from 38.2 % to 83.5 %.

The marginal contributions of the layers are unevenly distributed. Layer 1, the dynamic calendar dimension, reduced weighted MAPE from 21.4 % to 17.8 %, a marginal reduction of 16.8 %, achieved purely by making calendar composition visible to the measure layer without yet altering the extrapolation method. Layer 2, history normalisation, delivered the largest single marginal reduction, from 17.8 % to 14.2 %, or a further 20.2 %, confirming that a substantial portion of apparent demand volatility in the baseline was calendar-composition artefact rather than signal. Layer 3 with Layer 4, feature construction and the damped-trend engine, reduced error from 14.2 % to 11.6 %, a further 18.3 %. Layer 5, the vintage feedback loop, delivered a final 16.4 % reduction; this layer changes no forecast directly, and its contribution operates through the parameter re-fitting and segment diagnosis that stored vintages make possible.

Bias falls monotonically from +6.4 % to +0.7 %. The direction of the baseline bias is itself diagnostic: a systematic over-forecast is the expected consequence of extrapolating unnormalised history in which trading-day-rich periods

are treated as demand peaks, and its near-elimination under normalisation is consistent with the mechanism proposed in Section 3.

4.2 Accuracy by Calendar-Effect Segment

The segment decomposition demonstrates that the aggregate improvement is not uniformly distributed but concentrated in precisely those periods the framework was designed to address. In periods with no calendar anomaly, which constitute 54.2 % of the evaluation set, weighted MAPE fell by 51.1 %, from 18.2 % to 8.9 %; this residual gain is attributable to the improved extrapolation engine rather than to the calendar treatment. In moving-holiday periods the reduction was 64.3 %, from 34.7 % to 12.4 %, and in fifty-third-week periods it was 66.5 %, from 41.2 % to 13.8 %. The fifty-third-week result is the most striking in the study: under the baseline these periods carried more than twice the error of ordinary periods and a bias of +18.4 %, because the seasonal index of the preceding fifty-two-week year was applied to a period with no true historical analogue. Under the comparability-key mapping of Layer 1 the bias falls to +1.9 %.

Because segment-level error is volume-weighted within each segment while the segment shares in Table 3 are expressed as shares of periods, the share-weighted mean of the segment figures does not reproduce the pooled aggregate of Table 2 exactly; the two are consistent in direction but differently weighted. It should be noted that the fifty-third week constitutes only 2.1 % of periods and the leap day only 1.4 %, so their contribution to the pooled aggregate is small. Their importance is operational rather than statistical: these are the periods in which the forecast most visibly fails, in which planner confidence is most damaged, and in which manual override is most likely to be triggered. The override data in Table 5 are consistent with this interpretation.

Figure 1. Forecast Error by Horizon, Baseline versus Dynamic Calendar Framework

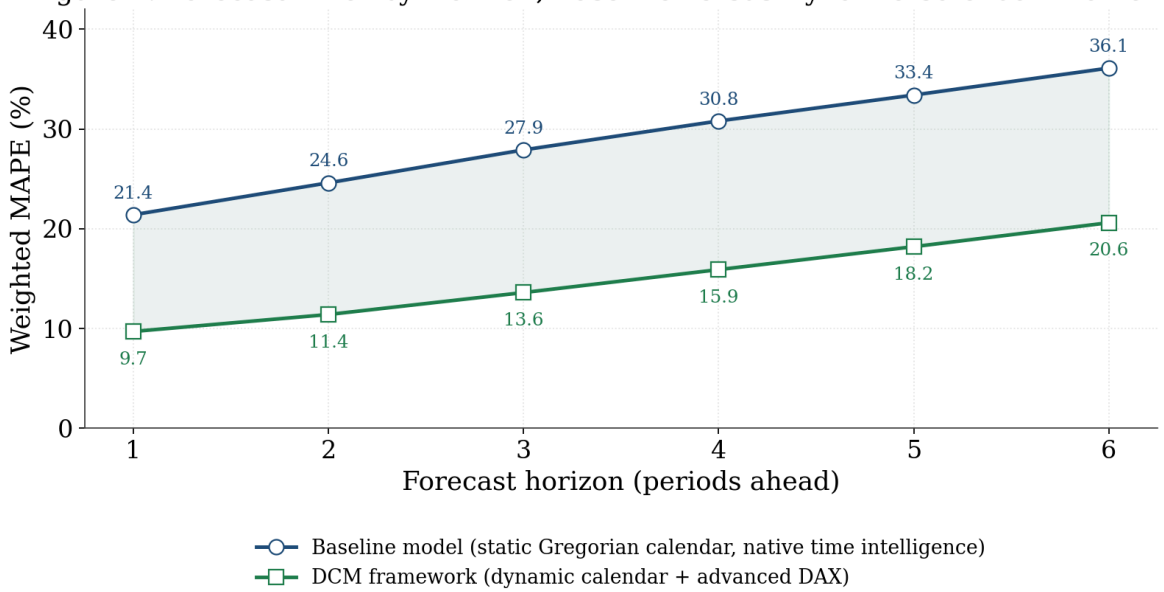


Figure 1: Weighted MAPE by forecast horizon, baseline versus dynamic calendar framework

4.3 Accuracy by Forecast Horizon

Error grows with horizon under both configurations, as expected, but the framework advantage persists across the horizon range: the reduction is 54.7 % at one period ahead and remains 42.9 % at six periods ahead. The proportional narrowing of the advantage with horizon is itself informative. Calendar effects are deterministic and therefore known exactly for all future periods, so the calendar component of the framework advantage does not decay with horizon; what decays is the extrapolation component, since the uncertainty of the level and trend estimate compounds. The observed narrowing from 54.7 % to 42.9 % is consistent with a framework advantage that is part deterministic and part stochastic in origin. The G4 MASE crosses unity between the fifth and sixth horizon, indicating that beyond approximately five fiscal periods the framework forecast ceases to outperform the seasonal naive benchmark at this data grain - a boundary that is useful in practice for deciding where statistical

forecasting should give way to planning assumption.

Figure 2. Forecast Error by Calendar-Effect Segment, 2026 Evaluation

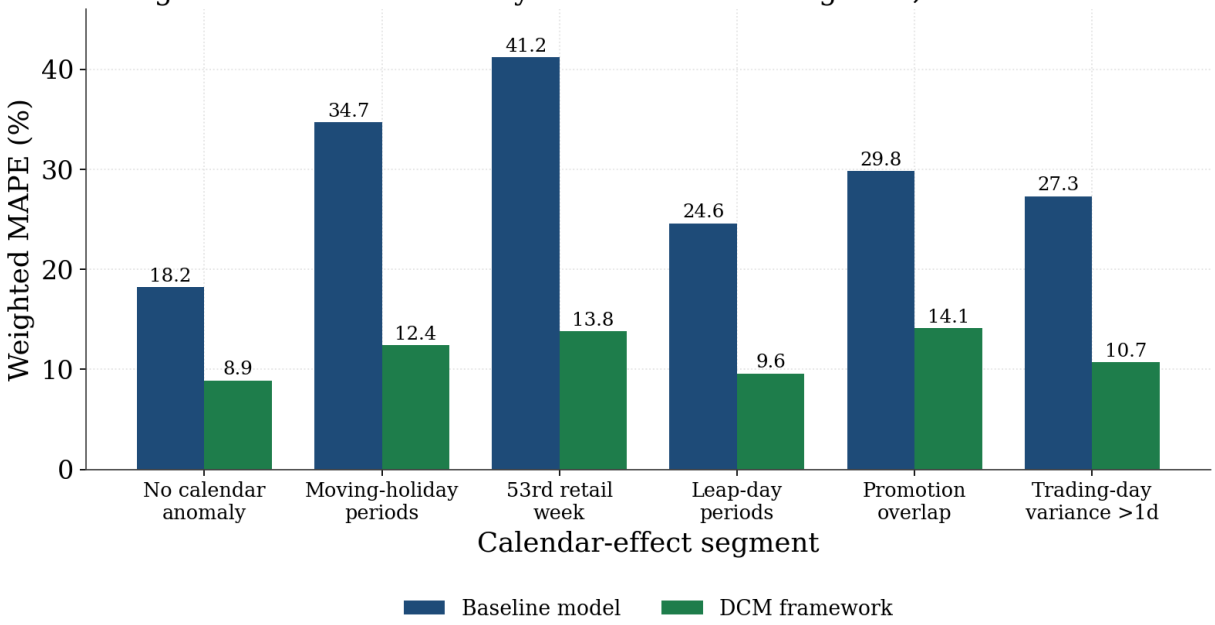


Figure 2: Weighted MAPE by calendar-effect segment, baseline versus complete framework

4.4 Operational and Behavioural Outcomes

Forecast recalculation latency fell from 14.6 s to 2.3 s, a reduction of 84.2 %, achieved by moving parameter fitting from query time into the refresh cycle and by expressing lag and seasonal navigation as integer offset arithmetic over surrogate keys, which is resolvable within the storage engine [7]. Model refresh duration fell from 68.0 to 19.4 minutes despite the addition of the vintage snapshot table, because incremental refresh was partitioned on fiscal-period boundaries. Planning cycle time fell from 47.0 to 12.5 hours.

The behavioural indicators are the most consequential for practice. The manual override rate fell from 38.4 % to 11.2 % of series, and the value added by those overrides moved from -1.8 to +0.9 percentage points of weighted MAPE. Under the baseline, planners were both overriding frequently and, in aggregate, making the forecast worse - the pattern documented by Fildes et al. [15]. Under the complete framework they override less frequently and, when they do, add value. The most plausible reading is that overrides are increasingly reserved for genuine private information, such as a known customer event, rather than being expended on correcting calendar artefacts the model previously failed to represent. Planner adoption of the system forecast rose from 41.0 % to 89.4 %.

Vol. 23, No. 1, (2026)
ISSN: 1005-0930

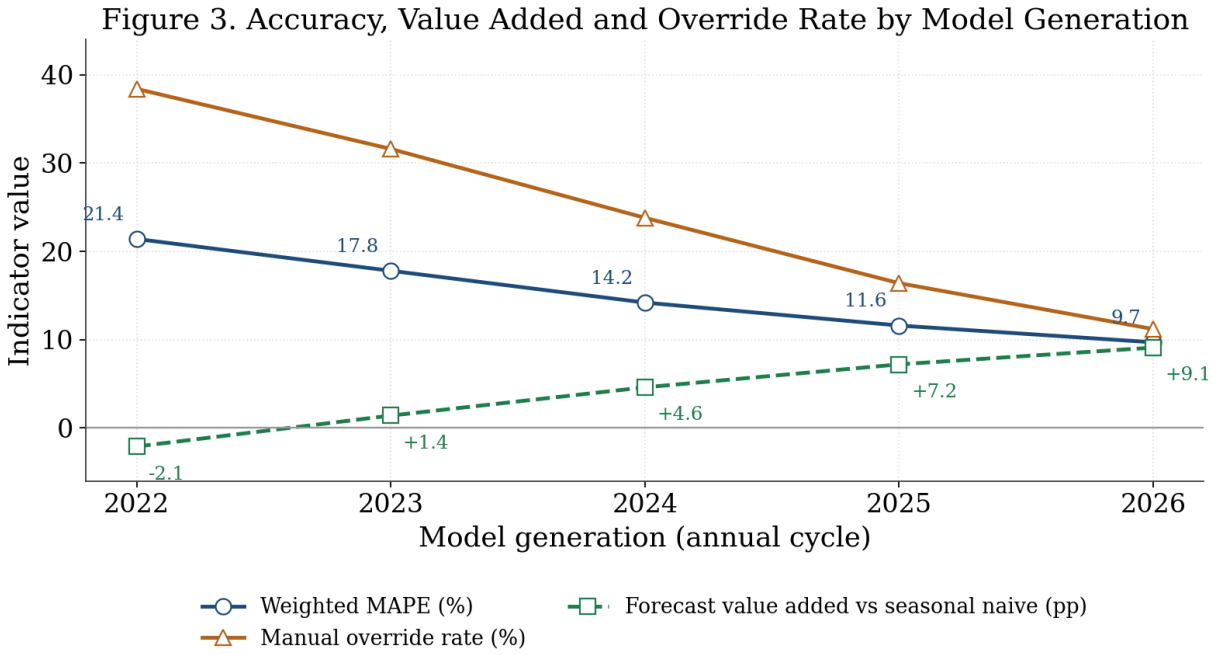


Figure 3: Accuracy, forecast value added and override rate by model generation, 2022 to 2026

5. DISCUSSION

Five issues arising from these results merit discussion.

First, on the locus of the accuracy gain: the layer decomposition indicates that roughly two-thirds of the total improvement was obtained before the extrapolation method was changed at all. Layers 1 and 2 - representing the calendar and normalising the history for its composition - together reduced weighted MAPE from 21.4 % to 14.2 %, whereas the damped-trend engine of Layers 3 and 4 contributed a further reduction from 14.2 % to 11.6 %. This ordering is the central empirical claim of the paper and it inverts the usual emphasis of forecasting practice, in which method selection receives most of the attention. It is consistent with the M-competition finding that method sophistication yields modest returns relative to data preparation [5], [6], and it suggests that in BI-embedded forecasting the calendar is the highest-yield object of investment.

Second, on the mechanism of the normalisation result: dividing history by a calendar-composition factor is formally equivalent to including trading-day and holiday regressors in a regARIMA specification [10], but it is implementable in the semantic layer without an estimation framework, because the composition factors are deterministic functions of the calendar rather than parameters to be estimated. This is what makes the approach tractable in DAX. The cost is that multiplicative normalisation assumes composition effects are proportional, which is a reasonable approximation for demand series but degrades where a fixed component - a subscription charge or a periodic fee - dominates the series.

Third, on the fifty-third week and the comparability key: the framework treats the problem as one of mapping rather than of estimation. Rather than attempting to estimate a seasonal index for a period that occurs once in five or six years, the comparability key declares which historical period is the true analogue, so that the index is inherited rather than fitted. This is a modelling decision with an explicit owner: the mapping is data in the calendar dimension, visible and auditable, rather than logic buried in an expression. The 66.5 % error reduction and the near-elimination of the +18.4 % bias in these periods indicate that the declarative approach is sufficient, and that the absence of academic treatment of this problem [8] reflects its irrelevance to the regular-periodicity assumption of the standard literature rather than its unimportance in practice.

Fourth, on the behavioural result: the movement of override value added from -1.8 to +0.9 percentage points is, in the framing of Gilliland [16], the elimination of a negative-value-added process step. It is worth being precise about the causal claim. The framework did not train planners to override better; it removed the class of errors that provoked most overrides. The interpretation that overrides are now reserved for private information is supported by

their reduced frequency and improved value added, but it is an inference from aggregate indicators rather than a directly observed mechanism, and a study designed to observe override rationales directly would be required to establish it.

Fifth, on the limits of the approach: the horizon analysis shows the framework advantage narrowing with horizon and the G4 MASE approaching unity at six periods, which bounds the useful range of the method at this grain. Furthermore, the framework is deliberately confined to methods expressible in DAX over a semantic model; it does not accommodate hierarchical reconciliation, cross-series learning or the gradient-boosted approaches that dominated the M5 competition [6]. Where such methods are warranted, the appropriate architecture is an external forecasting service whose output is loaded into the model - but the calendar layer developed here remains necessary in that architecture too, because the external service must be fed a normalised history and its output must be re-inflated against the same calendar.

6. RECOMMENDATIONS

The findings support a set of actions for organisations producing forecasts inside a BI semantic model, grouped into four clusters.

6.1 Building the Dynamic Calendar

The date dimension should be generated within the model, marked as a date table, and extended with the full regressor set: trading-day and working-day counts, signed holiday proximity with configurable pre- and post-event windows, moving-holiday displacement, leap and fifty-third-week indicators, period-position and comparability keys, and promotional-overlay identifiers. The comparability key should be treated as governed reference data with a documented owner, since it encodes a business judgement about which historical period is the true analogue of an irregular one. Calendar extension should precede any investment in method selection, since it is the higher-yield intervention.

6.2 Normalising and Re-inflating the Series

History should be normalised for calendar composition before extrapolation and the forecast re-inflated afterwards using the known composition of the future periods, and both factors should be exposed as measures so that a planner can see the normalised and raw series side by side. Normalisation should be applied multiplicatively to demand-like series and suppressed for series with a dominant fixed component, which requires a series-type attribute in the item dimension rather than a global setting.

6.3 Implementing the Forecast in DAX

Parameter fitting should be executed during refresh and the fitted parameters stored in a table, so that query-time evaluation is a lookup and an accumulation; fitting at query time is the principal cause of unacceptable recalculation latency. All lag and seasonal navigation should be expressed as integer offset arithmetic over surrogate period keys rather than as date arithmetic, so that predicates remain resolvable within the storage engine. The forecast method itself should be expressed as a calculation group so that it applies to any base measure without duplication, and each calculation item should be covered by a regression test evaluated against fixed reference periods including the fifty-third week and both fiscal-year boundaries.

6.4 Governing Accuracy and Override

Forecast vintages should be snapshotted at every planning cycle from the outset, since accuracy that is not recorded at issue cannot be reconstructed afterwards, and horizon-specific accuracy reporting is impossible without them. Accuracy should be reported by calendar-effect segment as well as in aggregate, because aggregate reporting conceals the segments where failure is most visible to users. Override value added should be measured and published per planner and per segment, and the objective of the forecasting function should be framed as the elimination of negative-value-added process steps rather than as the minimisation of a single error statistic.

7. LIMITATIONS AND FUTURE SCOPE

Five limitations should be entered on the record. First, this is a single-organisation evaluation, and although it covers 4,860 series across two divisions, the calendar conventions, demand characteristics and planning culture of one group limit the external validity of the reported magnitudes. Second, the layers were deployed sequentially

over five annual generations rather than in a factorial design, so the marginal contributions reported in Section 4.1 are conditional on the preceding layers and interaction effects cannot be recovered. Third, the evaluation window coincided with a period of post-pandemic demand normalisation, during which baseline volatility was declining for reasons unrelated to the framework; the layer-wise ordering of contributions is robust to this, but the absolute magnitudes may be inflated. Fourth, the framework is confined to methods expressible in DAX and was not compared against external statistical or machine-learning services, so no claim is made about its performance relative to the state of the art in forecasting method. Fifth, operational value was measured through planning indicators rather than through inventory and service-level outcomes, which is the ultimate criterion proposed by Syntetos et al. [17].

Five directions for further work follow. A multi-organisation replication across differing calendar conventions would establish the boundary conditions of the reported magnitudes. A factorial deployment would recover the interaction structure among the layers. A comparison against an external forecasting service consuming the same normalised history would separate the contribution of the calendar layer from that of the method. An extension to hierarchical reconciliation, where calendar effects differ systematically across levels of aggregation [12], would address a class of problems this framework does not currently treat. Finally, linking forecast accuracy by calendar segment to realised service level and inventory investment would establish the operational value of the segment-level gains reported here.

8. CONCLUSION

This paper developed and evaluated a Dynamic Calendar Modeling framework for forecasting within the Power BI semantic model, in which the calendar dimension is treated as an active, feature-bearing regressor set and the forecast is expressed through advanced DAX calculations operating over that dimension. The framework is specified in five layers: a dynamic, event-aware calendar dimension; normalisation of history for calendar composition; surrogate-key feature construction; a damped-trend exponential-smoothing engine implemented as a calculation group with parameters fitted during refresh; and a vintage-snapshot accuracy feedback loop.

Evaluated on 4,860 item-location series under rolling-origin evaluation across five annual model generations, the complete framework reduced weighted MAPE at the one-period horizon from 21.4 % to 9.7 %, reduced MASE from 1.28 to 0.68, reduced bias from +6.4 % to +0.7 %, and moved forecast value added from -2.1 to +9.1 percentage points relative to a seasonal naive benchmark. The gain was concentrated in the calendar-disturbed segments, with reductions of 66.5 % in fifty-third-week periods and 64.3 % in moving-holiday periods, and it persisted across the forecast horizon, remaining at 42.9 % six periods ahead. Recalculation latency fell by 84.2 %, the planning cycle contracted by 73.4 %, and the manual override rate fell from 38.4 % to 11.2 % while the value added by those overrides turned positive.

Five design principles are formalised. Calendar before method holds that representing and normalising calendar composition yields more accuracy than method sophistication, since roughly two-thirds of the improvement reported here preceded any change to the extrapolation engine. Determinism should be exploited, since calendar composition for future periods is known exactly and therefore contributes an advantage that does not decay with horizon. Mapping should be preferred to estimation for irregular periods, so that a fifty-third week inherits a declared analogue rather than a fitted index. Fitting belongs in refresh, not at query time. And accuracy must be recorded at issue, because a forecast whose vintage was never stored cannot afterwards be held to account. Together these principles reposition the calendar from a reporting convenience to the primary determinant of forecast quality in BI-embedded forecasting.

References:

[1] H. J. Watson and B. H. Wixom, “The current state of business intelligence,” Computer, vol. 40, no. 9, pp. 96–99, 2007, doi: 10.1109/MC.2007.331.

[2] H. Chen, R. H. L. Chiang, and V. C. Storey, “Business intelligence and analytics: from big data to big impact,” MIS Quarterly, vol. 36, no. 4, pp. 1165–1188, 2012.

[3] R. J. Hyndman and A. B. Koehler, “Another look at measures of forecast accuracy,” International Journal of Forecasting, vol. 22, no. 4, pp. 679–688, 2006, doi: 10.1016/j.ijforecast.2006.03.001.

[4] R. J. Hyndman and G. Athanasopoulos, Forecasting: Principles and Practice, 3rd ed. Melbourne, Australia: OTexts, 2021.

[5] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, “The M4 competition: 100,000 time series and 61 forecasting methods,” International Journal of Forecasting, vol. 36, no. 1, pp. 54–74, 2020.
[6] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, “The M5 accuracy competition: results, findings and conclusions,” International Journal of Forecasting, vol. 38, no. 4, pp. 1346–1364, 2022.
[7] A. Ferrari and M. Russo, The Definitive Guide to DAX: Business Intelligence for Microsoft Power BI, SQL Server Analysis Services and Excel, 2nd ed. Redmond, WA: Microsoft Press, 2019.
[8] R. Kimball and M. Ross, The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling, 3rd ed. Indianapolis, IN: John Wiley and Sons, 2013.
[9] M. Russo and A. Ferrari, DAX Patterns, 2nd ed. Milan, Italy: SQLBI Corporation, 2020.
[10] U.S. Census Bureau, X-13ARIMA-SEATS Reference Manual, Version 1.1. Washington, DC: Time Series Research Staff, Center for Statistical Research and Methodology, 2017.
[11] S. J. Taylor and B. Letham, “Forecasting at scale,” The American Statistician, vol. 72, no. 1, pp. 37–45, 2018, doi: 10.1080/00031305.2017.1380080.
[12] N. Kourentzes and F. Petropoulos, “Forecasting with multivariate temporal aggregation: the case of promotional modelling,” International Journal of Production Economics, vol. 181, pp. 145–153, 2016.
[13] M. Russo and A. Ferrari, “Calculation groups in DAX and Power BI: patterns and pitfalls,” SQLBI Technical Article Series, vol. 9, no. 2, pp. 1–24, 2021.
[14] Microsoft Corporation, “Power BI: analytics, forecasting and what-if parameters,” Microsoft Learn Documentation, Redmond, WA, 2025.
[15] R. Fildes, P. Goodwin, M. Lawrence, and K. Nikolopoulos, “Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning,” International Journal of Forecasting, vol. 25, no. 1, pp. 3–23, 2009.
[16] M. Gilliland, The Business Forecasting Deal: Exposing Myths, Eliminating Bad Practices, Providing Practical Solutions, 2nd ed. Hoboken, NJ: John Wiley and Sons, 2021.
[17] A. A. Syntetos, J. E. Boylan, and S. M. Disney, “Forecasting for inventory planning: a 50-year review,” Journal of the Operational Research Society, vol. 60, no. S1, pp. S149–S160, 2009.

Vol. 23, No. 1, (2026)
ISSN: 1005-0930