When Does Monthly Information Improve Investment Nowcasts? Benchmark Synchronisation and Evidence from Korea

Hyun Hak Kim*

September 2026

Abstract

Forecast comparisons can misattribute the value of a newly released quarterly target to monthly information when a challenger and benchmark use different information dates. We distinguish the held-AR replacement comparison (H), the same-date operational comparison (O), and a fixed-operator monthly-input counterfactual (M). In simulations with an informative target update but no monthly signal, the test based on H rejects in 45.1% of replications, versus 0.9% and 1.1% for O and M. We evaluate five mixed-frequency procedures and eight benchmarks at twelve monthly issue dates for 42 quarters of Korean aggregate gross fixed capital formation. Across 48 evaluation cells, only the O comparison for seasonally adjusted annualised quarter-on-quarter growth at $h = 0$ survives Holm adjustment (raw $p = 0.0003$; Holm–48 $p = 0.0144$). The equipment-only restricted MIDAS bridge has the lowest RMSE (5.623), outperforming every higher-dimensional MIDAS specification. The calendar-based pseudo-real-time design uses latest-vintage histories, observed evaluation-period target-release dates, a rule-based earlier calendar, and assumed predictor lags rather than complete historical vintages. The evidence supports synchronised benchmarks and economically grounded bridge comparators, not general MIDAS dominance or a uniform marginal value of monthly updates.

Keywords: investment nowcasting; mixed-frequency data; benchmark synchronisation; forecast evaluation; information sets.

JEL classification: C22; C53; E22; E27.

* Professor, Department of Economics, Kookmin University, Seoul, Korea. hyunhak.kim@kookmin.ac.kr.

1 Introduction

A benchmark is not a passive yardstick when forecasts are updated repeatedly. Macroeconomic nowcasts are produced along a release calendar: monthly indicators arrive asynchronously, a new quarterly target observation eventually becomes available, and the forecasting procedure may be re-estimated at each issue date. The nowcasting literature has developed sophisticated ways to process this ragged edge and combine high- and low-frequency information (Giannone et al., 2008; Bańbura et al., 2013; Bok et al., 2018). Yet an evaluation that compares an updated mixed-frequency forecast with a benchmark left at an earlier information date changes more than the model. It changes the target history, the estimation state, and the information available to the two forecasts. Consequently, a measured gain need not identify the value of the newly released monthly data that motivated the update.

This distinction matters whenever a sequence of forecasts is produced for the same target period. Suppose that a forecaster updates an investment nowcast after an industrial-production release, while the reported autoregressive benchmark remains the forecast issued a month earlier. The resulting loss differential answers a legitimate operational question—whether to replace the held forecast with the current forecasting system—but it does not isolate the incremental content of the industrial-production release. Recent work shows a related identification problem for the directional accuracy of temporal aggregates: an aggregate can appear predictable by construction when the no-change benchmark is defined at the wrong sampling point (McCarthy and Snudden, 2025). The problem studied here is different. It concerns squared-error point forecasts, quarterly target updates, re-estimation, and a sequence of monthly information sets. The common lesson is that benchmark construction determines what a forecast comparison measures.

The paper separates three comparisons. A held-benchmark comparison, denoted H, compares the current mixed-frequency forecast with an AR forecast retained from an earlier issue date. The first estimand of interest, operational replacement value O, instead compares the current mixed-frequency procedure with the live AR available at the same issue date. It is an end-to-end comparison of two executable forecasting systems under a synchronised information set. The second estimand, fixed-operator monthly-input value M, applies the same fitted mixed-frequency operator to updated monthly inputs and to a same-horizon counterfactual that admits no new monthly observations. It is therefore an input ablation, not a model horse race. An exact add-and-subtract decomposition shows that H equals O plus the gain from updating the AR itself. A second decomposition relates O and M but imposes no sign ordering between them. A monthly input may improve a particular fitted procedure even when that procedure loses to a live AR, while an operational procedure may beat the AR for reasons other than the newly admitted monthly inputs.

A sample-matched simulation demonstrates that this distinction can be empirically important. In a design with no monthly predictive content but an informative quarterly target update, the test based on the held-AR comparison rejects in 45.1% of replications. Tests based on the synchronised O and fixed-operator M comparisons reject in 0.9% and 1.1%, respectively. These figures describe

one calibration rather than a universal distortion rate. Their role is to show that an apparently successful monthly-data procedure can inherit the value of a target update when the benchmark is held fixed. Synchronising the operational benchmark and isolating the input counterfactual largely remove that attribution error in this design. This treatment of expanding information sets is related to the distinction between ideal and estimated forecasts (Holzmann and Eulert, 2014) and to multi-horizon procedures that ask which models can be distinguished at each nowcast horizon (Fosten and Gutknecht, 2021). The present analysis asks a logically prior question: what does the reported forecast gain identify at a given horizon?

The empirical application concerns real gross fixed capital formation (GFCF) in Korea. Investment is an informative stress test because it is a volatile national-accounts component that is central to assessments of aggregate demand and productive capacity and has closely related monthly indicators. Existing work documents that mixed-frequency performance varies across GDP components (Foroni and Marcellino, 2014), that financial information can help nowcast real investment (Degiannakis, 2022), and that mixed-frequency methods can be useful for Korean GDP (Kim and Swanson, 2018). The target here is narrower and the update path is denser. For each quarter from 2016Q1 through 2026Q2, the exercise constructs twelve issue-date information sets, indexed from $h = 11$ to $h = 0$. The last three, $h = 2, 1, 0$, fall inside the target quarter and are nowcasts; earlier dates are forecasts. The focal target is seasonally adjusted annualised quarter-on-quarter log growth. Non-seasonally adjusted QoQ and year-on-year growth are retained as secondary transformations.

The model menu is designed to separate information representation from evaluation design. The five challengers—a cross-validated ridge bridge, targeted MIDAS, factor U-MIDAS, sparse-group MIDAS, and factor-augmented sparse-group MIDAS—represent dense shrinkage, economic targeting, factor compression, grouped sparsity, and combined sparse and dense information, respectively (Marcellino and Schumacher, 2010; Siliverstovs, 2017; Babii et al., 2022; Beyhum and Striaukas, 2026). The comparison set also contains eight benchmarks: a live direct AR, a recursively selected BIC-AR, a recursive mean, a no-change forecast, equipment-only and construction-only restricted MIDAS bridges, and two fixed-weight forecast combinations. Simple bridge models are important competitors rather than merely ancillary baselines (Baffigi et al., 2004). Model rankings are allowed to vary with the forecast horizon, as they often do in mixed-frequency applications (Kuzin et al., 2011). Estimation, preprocessing, and tuning are performed recursively, and all reported comparisons use common target quarters and issue dates.

The operational evidence is strong but narrow. Across the 48-cell evaluation family, the only result that survives the global adjustment is the seasonally adjusted QoQ O comparison at $h = 0$. The raw omnibus p-value is 0.0003 and the Holm-adjusted value is 0.0144; the conclusion is unchanged under block lengths of three, four, and six quarters. This is evidence that at least one of the five mixed-frequency challengers improves on the synchronised live AR at the end of the quarter. It is not a test that selects the best model among all thirteen procedures, and it does not extend across the full horizon path or across target transformations.

The stronger benchmark horse race changes the economic interpretation. At $h = 0$, the

equipment-only restricted MIDAS bridge has the lowest RMSE, 5.623. Targeted MIDAS is the best of the five challengers, with an RMSE of 5.866, followed by factor-augmented sparse-group MIDAS at 6.018 and sparse-group MIDAS at 6.050. The targeted procedure has an RMSE roughly 23% lower than that of the live AR, whose RMSE is 7.616, but it does not improve on the equipment bridge. Sparse-group MIDAS improves on the BIC-AR after both the Holm-9 and Holm-27 adjustments, whereas factor-augmented sparse-group MIDAS does so only after the Holm-9 adjustment; neither improves on the equipment bridge. The result is therefore not that a high-dimensional MIDAS specification dominates simpler forecasting rules. It is that at least one mixed-frequency procedure has operational value at the end of the quarter, while a direct, economically matched indicator performs at least as well as the more complex procedures.

The fixed-operator evidence answers a different question. For seasonally adjusted QoQ growth, moving from the pre-quarter monthly input set $X^{(0)}$ to the quarter-end set $X^{(3)}$ reduces RMSE for each of the five challengers; the cumulative gains range from 6.66% to 29.81% relative to each model's own no-new-month counterfactual. These percentages cannot rank models because the counterfactual RMSE differs across fitted procedures. More importantly, the monthly loss contributions are episodic. Depending on the model and update month, only 38.1%–66.7% of quarter-specific updates yield a positive loss contribution. Although the raw omnibus p-value for the M comparison in the QoQ, $h = 0$ cell is 0.0052, its Holm-adjusted value across 48 cells is 0.2444. The appropriate conclusion is therefore descriptive and model-specific. Monthly forecast updates can have material cumulative value, but the evidence does not support uniform or globally significant input value.

The vintage evidence further limits the claim. The main predictor histories are August 2026 latest-vintage observations. Actual quarterly target-release dates are used from 2009Q4 onward, including throughout the evaluation period; earlier training quarters use a 28-day rule. Monthly availability relies on deterministic series-specific lag assumptions rather than a complete archive of historical publication timestamps. The exercise is thus latest-vintage, calendar-based pseudo-real-time, not a historical real-time backtest (Croushore and Stark, 2001; Croushore, 2006). An archive of 44 preliminary aggregate GFCF releases is too short to re-estimate the complete recursive procedure under the prespecified requirements for estimation and inference windows. A separate 41-quarter outcome-only rescoring leaves the latest-vintage training histories and forecasts frozen. The favourable quarter-end ranking of targeted, sparse-group, and factor-augmented sparse-group MIDAS against the live AR is not reversed, but this partial diagnostic cannot establish full real-time performance.

These findings offer three contributions. First, the paper provides an evaluation discipline for forecasting under evolving information sets. It distinguishes a held-forecast replacement comparison, same-origin operational replacement value, and fixed-operator monthly-input value, and it uses simulation to show why their separation can matter in finite samples. This is an evaluation framework, not a new estimator or a claim of new forecasting theory. Second, the application evaluates a demanding expenditure component over a dense monthly path while including strong

quarterly, direct-indicator, and combination benchmarks. The equipment bridge's leading rank is a substantive result, consistent with timeliness and economically grounded predictor choice being at least as important as model complexity. Third, the analysis treats horizon, transformation, benchmark, and estimand as dimensions of the claim rather than robustness labels. Global multiplicity adjustment, alternative transformations with less favourable results, incomplete vintage coverage, and unstable monthly contributions are reported alongside the favourable quarter-end result.

The remainder of the paper is organised as follows. Section 2 develops the forecast-value framework for evolving information sets. Section 3 describes Korean investment, the predictor panel, and the monthly availability assumptions. Section 4 presents the forecasting models and recursive evaluation design. Section 5 reports the operational results, fixed-operator update evidence, and robustness exercises. Section 6 draws the practical implications and concludes. The appendices document data construction, implementation details, update accounting, and additional empirical results.

2 Forecast Value under Evolving Information Sets

Repeated forecasting creates a measurement problem before it creates a model-selection problem. For a target quarter t, let $\tau_{t,h}$ denote the issue date h months before the quarter end and let $\mathcal{I}_{t,h}$ collect the information admissible on that date. The empirical exercise uses $h = 11, \ldots, 0$, with smaller h indicating a later information set. The quarterly target history, the monthly ragged edge, and the fitted state of a forecasting procedure can all change between two adjacent issue dates. A loss comparison is interpretable only after specifying which of those objects are allowed to change.

2.1 Issue dates and forecast objects

Write y_t for the realised transformed target and $L(y_t, f)$ for the loss of forecast f. The empirical analysis uses squared loss, but the accounting identities below require only that the displayed losses be well defined. Fix a model m, target t, and issue date h. Five forecast objects are needed. The held AR forecast $a^H_{t,h}$ was issued at an earlier origin and then retained. The live AR $a^O_{t,h}$ is recomputed at $\tau_{t,h}$. The operational mixed-frequency forecast $g^O_{m,t,h}$ is also produced at $\tau_{t,h}$, using the recursive preprocessing, tuning, and coefficient estimates that its procedure would actually employ. Synchronisation means that $a^O_{t,h}$ and $g^O_{m,t,h}$ have the same issue date, release-calendar-admissible quarterly target history, latest target lag, and training quarters. It does not mean that their regressors are identical: using monthly predictors is precisely what distinguishes the challenger from the direct AR.

The remaining two forecasts define a controlled input comparison. Let $X^0_{t,h}$ denote the no-new-month predictor input constructed under the lag-based availability scheme for horizon h, $X^1_{t,h}$ the corresponding updated input, and $T^0_{t,h}$ the quarterly-target state held common to both. Let $S^M_{m,t,h}$ contain the feature construction, scaling, factor extraction, tuning choices, and coefficients,

all obtained without admitting $X^1_{t,h}$. The fixed-operator forecasts are

$$g^0_{m,t,h} = f_m(S^M_{m,t,h}, X^0_{t,h}, T^0_{t,h}),$$
$$g^1_{m,t,h} = f_m(S^M_{m,t,h}, X^1_{t,h}, T^0_{t,h}). \tag{1}$$

One state is shared within this pair, while a separate state is fitted for each horizon. In particular, $g^0_{m,t,h}$ is a same-horizon counterfactual; it is not the operational forecast from an earlier issue date. This distinction prevents changes in the target history, preprocessing, or coefficients from being relabelled as the contribution of monthly inputs. At the pre-quarter origin $h = 3$, the baseline and updated input coincide, so M is structurally zero.

2.2 Three comparisons and two exact decompositions

Define realised gains so that a positive value favours the current challenger or updated input:

$$d^H_{m,t,h} = L(y_t, a^H_{t,h}) - L(y_t, g^O_{m,t,h}), \tag{2}$$
$$d^O_{m,t,h} = L(y_t, a^O_{t,h}) - L(y_t, g^O_{m,t,h}), \tag{3}$$
$$d^M_{m,t,h} = L(y_t, g^0_{m,t,h}) - L(y_t, g^1_{m,t,h}). \tag{4}$$

Their population counterparts are $\theta^j_{m,h} = E(d^j_{m,t,h})$ for $j \in \{H, O, M\}$; their empirical counterparts are the averages of the same loss differentials over common target quarters. The three objects answer different questions, summarised in Panel A of Table 1. H asks whether a user should replace a previously issued AR forecast with the current mixed-frequency forecast. O asks whether the current mixed-frequency system improves on a live direct AR. M asks whether updated monthly inputs improve a particular fitted procedure when its operator and quarterly-target state are fixed. Thus O is an end-to-end system comparison, whereas M is an input ablation rather than a model horse race.

The first identity isolates what a held benchmark adds. Adding and subtracting the live AR loss in Equation (2) gives, path by path,

$$d^H_{m,t,h} = d^O_{m,t,h} + u^{AR}_{t,h},$$
$$u^{AR}_{t,h} = L(y_t, a^H_{t,h}) - L(y_t, a^O_{t,h}). \tag{5}$$

When the terms are integrable, $\theta^H_{m,h} = \theta^O_{m,h} + E(u^{AR}_{t,h})$. The synchronisation term can reflect a newly available quarterly target, a different horizon equation, and AR re-estimation. A comparison with the held forecast can therefore answer a valid replacement-policy question, but its gain cannot be attributed to current monthly inputs alone. Nor is a large rejection rate for H necessarily a size distortion: if retaining the held forecast has genuinely become inferior, the rejection answers the composite replacement question. The attribution error arises only when that result is described as monthly-input evidence.

Table 1: Forecast-value comparisons and admissible interpretations

	First forecast	Second forecast	What a positive value supports
Panel A: comparison map			
H	Earlier, held AR	Current operational procedure	Replacement of the held forecast; the gain may include benchmark updating
O	Same-issue live AR	Current operational procedure	End-to-end replacement of the live AR by the mixed-frequency system
M	Fixed-operator no-new-month counterfactual	Fixed-operator updated input	Monthly-input value within the fitted procedure and target state
Panel B: joint signs for a fixed model and horizon			
O sign	M sign	Joint pattern	Admissible interpretation
$\theta^O > 0$	$\theta^M > 0$	Both comparisons favour the second forecast	The system beats the live AR and its controlled monthly-input update has positive value
$\theta^O > 0$	$\theta^M \leq 0$	Only O is positive	The system has replacement value, but this ablation does not support monthly-input value
$\theta^O \leq 0$	$\theta^M > 0$	Only M is positive	The input helps within the fitted system, but the system does not beat the live AR
$\theta^O \leq 0$	$\theta^M \leq 0$	Neither comparison is positive	Neither positive-value claim is supported

Notes: Every row refers to the same target, transformation, model, and horizon. Panel B classifies population quantities; sample-based claims require the corresponding model-level and familywise inference.

A second add-and-subtract identity links the operational and fixed-operator comparisons:

$$d^O_{m,t,h} = d^M_{m,t,h} + \underbrace{\{L(y_t, a^O_{t,h}) - L(y_t, g^0_{m,t,h})\}}_{\text{baseline-alignment gap}} + \underbrace{\{L(y_t, g^1_{m,t,h}) - L(y_t, g^O_{m,t,h})\}}_{\text{operator-state gap}}. \tag{6}$$

The first gap compares the live AR with the no-new-month counterfactual. The second compares the fixed-operator updated-input forecast with the fully operational forecast, whose state may have been recursively re-estimated. Neither gap has a general sign. Even if the operational and fixed-operator updated-input forecasts happen to coincide, the baseline-alignment gap remains. Consequently, θ^O and θ^M have no general ordering by sign or magnitude. The four cases in Panel B are logical interpretations, not additional tests. In particular, separate omnibus rejections for O and M need not identify the same winning model and cannot by themselves establish a joint positive claim for any named procedure.

For an oracle squared-error forecast, nested information has non-negative population value. Experiment M, however, applies an estimated and potentially misspecified operator to two input vectors, so its population or sample gain need not be positive. Appendix C gives the oracle result, the exact telescoping decomposition of within-quarter updates, and the squared-loss interpretation

used in the empirical analysis.

3 Korean Investment and the Monthly Information Set

Korean gross fixed capital formation is both economically important and difficult to assess in real time. In the January 2026 World Bank archive snapshot, GFCF averaged 30.7 per cent of Korea's GDP over 2015–2024 (World Bank, 2026). The quarterly national accounts disaggregate GFCF into construction, equipment, and intellectual property products (IPP). These components display different short-run dynamics and have different monthly proxies. On average over 2016Q1–2026Q1, construction accounted for 47.9 per cent of nominal GFCF, equipment for 31.0 per cent, and IPP for 21.1 per cent. Equipment is also the most volatile component. Over the common evaluation window, the standard deviation of seasonally adjusted annualised QoQ growth is 14.8 percentage points for equipment, compared with 9.1 for construction, 4.5 for IPP, and 7.4 for aggregate GFCF. These differences motivate both the aggregate forecasting exercise and the component diagnostics.

3.1 Targets and transformations

The national-accounts targets are obtained from the Bank of Korea Economic Statistics System (ECOS) (Bank of Korea, 2026b). The focal series is real aggregate GFCF, seasonally adjusted and measured as a chain-volume series with reference year 2020. Let $Y_{j,t}^{SA}$ denote the published quarterly level of aggregate GFCF or of component j. The headline target is its annualised one-quarter log change,

$$y_{j,t}^{q} = 400\{\log(Y_{j,t}^{SA}) - \log(Y_{j,t-1}^{SA})\}. \tag{7}$$

This target closely matches the short-term question posed during a quarter: how investment activity is changing relative to the preceding quarter. The focal analysis therefore concerns aggregate seasonally adjusted QoQ growth. Construction, equipment, and IPP results provide secondary evidence on the possible economic sources of aggregate predictability.

Table 2 also records two non-seasonally adjusted robustness targets: annualised QoQ growth with quarter-of-year controls and YoY growth. Their exact definitions and interpretation are given in Appendix A.

The balanced evaluation sample contains the same 42 target quarters, 2016Q1–2026Q2, at every issue date. Holding the target quarters fixed prevents changes in sample composition from being mistaken for horizon effects. The quarterly and monthly estimation histories are drawn from a common late-August 2026 ECOS snapshot. Section 4 explains how expanding-window estimation and all preprocessing steps are applied recursively at each forecast origin.

3.2 Monthly predictors and lag-based availability masking

The monthly information set contains 24 predefined predictor groups constructed from 25 ECOS source series. The difference arises because the yield on three-year AA-minus-rated corporate

Table 2: Korean investment targets and descriptive characteristics

Panel A: target definitions

Role	Seasonal basis	Transformation	Evaluation window
Focal	SA real GFCF	$400\{\log(Y_t) - \log(Y_{t-1})\}$	2016Q1–2026Q2
Robustness	NSA real GFCF	$400\{\log(Y_t) - \log(Y_{t-1})\}$	2016Q1–2026Q2
Robustness	NSA real GFCF	$100\{\log(Y_t) - \log(Y_{t-4})\}$	2016Q1–2026Q2

Panel B: composition and focal-target volatility

Series	Mean nominal share (%)	SA annualised QoQ standard deviation
Aggregate GFCF	100.0	7.4
Construction	47.9	9.1
Equipment	31.0	14.8
IPP	21.1	4.5

Notes: The targets are Bank of Korea ECOS chain-volume measures with reference year 2020. The NSA-QoQ forecasting equations additionally include unpenalised quarter-of-year indicators. Nominal shares use seasonally adjusted current-price observations from 2016Q1 through 2026Q1, the last nominal quarter in the August 2026 snapshot. Standard deviations use the 42 common evaluation quarters and are expressed in annualised log-growth percentage points.

bonds and the yield on three-year Korean Treasury bonds are combined into a single credit-spread predictor. The panel deliberately combines broad macroeconomic signals with indicators directly related to the three investment components. Six common-macro groups consist of all-industry production, manufacturing production, manufacturing utilisation, the leading composite index, the won–US dollar exchange rate, and the credit spread. Four construction groups cover completed construction, orders, housing approvals, and a business survey. Eight equipment groups cover investment and machinery indices, trade, surveys, and credit. Six IPP groups contain service-production series, business surveys, and the KOSDAQ index. Table 3 summarises these blocks.

Series transformations and source-history coverage are documented in Appendix A.

Table 3: Monthly predictor blocks and timing assumptions

Block	Groups	ECOS series	Representative indicators	Lag classes (days)
Common macro	6	7	Production, utilisation, leading index, exchange rate, credit spread	1, 30
Construction	4	4	Completed construction, orders, approvals, business outlook	0, 30, 45
Equipment	8	8	Equipment and machinery indices, trade, surveys, credit	0, 15, 30, 60
IPP	6	6	Service production, business outlook surveys, KOSDAQ	0, 1, 30
Total	24	25		0, 1, 15, 30, 45, 60

Notes: All source series are from Bank of Korea ECOS. Across the 25 source series, the assumed lag distribution is four series at zero days, four at one day, two at 15 days, 13 at 30 days, one at 45 days, and one at 60 days. These lags are deterministic availability assumptions rather than observations from a historical release-timestamp archive.

At every issue date, lag-based availability masking precedes imputation, standardisation, factor

extraction, and regression construction. The detailed masking rule and its timing caveats appear in Appendix A.

3.3 Twelve issue dates and the vintage boundary

For every target quarter, forecasts are produced at the end of twelve consecutive months. The earliest origin is indexed by $h = 11$ and the quarter-end origin by $h = 0$. For example, the sequence for 2026Q1 begins at the end of April 2025 and continues monthly through March 2026. Origins $h = 11, \ldots, 3$ precede the target quarter and are forecasts; $h = 2, 1, 0$ occur in January, February, and March and are nowcasts. Figure 1 shows the timing and the distinction between updating the complete operational system and updating only the fixed operator's monthly inputs.

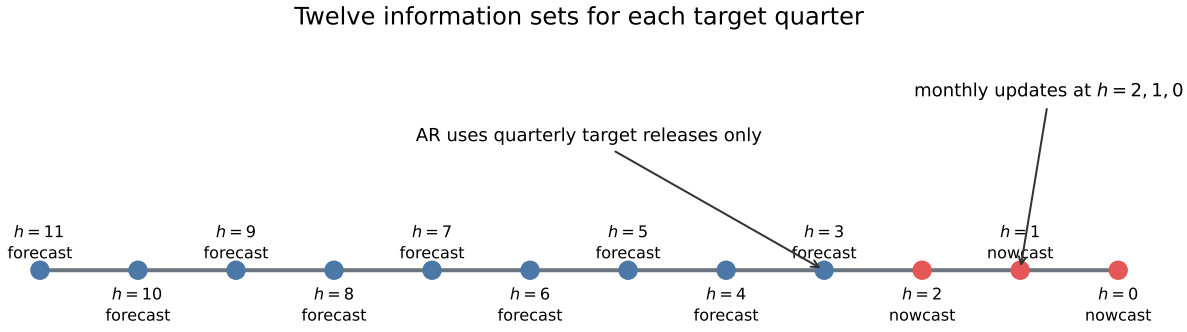

Figure 1: Information sets and forecast comparisons along the issue-date path

Notes: The operational comparison uses the live AR and each challenger at the same issue date. The fixed-operator comparison holds the forecasting state and quarterly target information fixed while changing the lag-masked monthly input. The specific contents of the ragged edge vary across source series.

A release calendar and a historical data vintage are distinct objects (Croushore and Stark, 2001; Croushore, 2006). ECOS provides the currently revised history rather than the sequence of values observed by a forecaster at every historical issue date. The main analysis therefore combines actual quarterly release dates from 2009Q4 onward, rule-based dates for earlier training quarters, and deterministic monthly lag masks with an August 2026 latest-vintage panel. It reconstructs which reference periods are deemed admissible under those timing rules, but not the first published value of every observation. We therefore refer to the twelve origins as issue dates and to their associated data as information sets, not as data vintages.

The quarterly release-calendar construction and the preliminary-outcome archive are documented in Appendix A.

4 Forecasting Models and Recursive Evaluation

The forecasting exercise holds the set of target quarters fixed while varying both the information available at twelve monthly issue dates and the way in which that information is processed. Every

model is estimated as a direct forecast separately for each target transformation and horizon. A historical training row for lead h is reconstructed at the same relative issue date. Thus, a longer-lead training observation cannot use the more complete monthly panel that subsequently became available for that historical quarter. The direct AR and each challenger use the same quarterly release calendar and the latest admissible target observation. Monthly observations enter only after passing the lag-based availability mask summarised in Section 3 and detailed in Appendix A.

4.1 Recursive estimation and tuning

The estimation window begins in 2006Q1 and expands through time. The common evaluation sample contains 42 consecutive quarters from 2016Q1 through 2026Q2. Hyperparameters are selected using eight sequential one-step-ahead validation folds within each outer estimation sample. The protocol requires at least 24 observations in the first inner estimation window, although this eligibility floor does not bind. Across the twelve-lead operational design, the realised smallest outer sample is 37 quarters and the smallest initial inner training sample is 29; at $h = 0$, the corresponding sample sizes are 40 and 32. Within every validation fold, the availability mask is applied and the scaler and principal components are re-estimated using only information admissible at the fold cutoff.

Additional imputation, eligibility, and principal-component rules are reported in Appendix B.

4.2 Core procedures and stronger benchmarks

The common quarterly benchmark is a direct AR based on the most recently released quarterly target,

$$y_t = c_h + q_t'\delta_h + \phi_h y_{\ell(t,h)} + u_{t,h}, \tag{8}$$

where $y_{\ell(t,h)}$ need not equal y_{t-1} if that observation has not yet been published, and q_t contains the NSA-QoQ quarter indicators when applicable. The live AR is re-estimated at each issue date. It can react to a newly released quarterly target but never observes the monthly predictor panel.

The five mixed-frequency procedures span dense shrinkage, economic targeting, factor compression, grouped sparsity, and combined sparse and dense information. Table 4 summarises the complete model menu; Appendix B gives the lag dictionaries, penalty grids, factor construction, and convergence criteria.

The stronger benchmarks ensure that the live AR is not the sole standard of comparison. At every origin, BIC-AR selects $p \in \{0, \ldots, 4\}$ on the common sample for which four lags can be constructed; ties favour the lower order. The equipment-only bridge uses the seasonally adjusted equipment-investment index, while the construction-only bridge uses real seasonally adjusted completed construction. Both use twelve lags and the same three-term Legendre basis as targeted MIDAS, but estimate all coefficients by OLS without tuning. The two forecast combinations have fixed weights and therefore cannot benefit from ex post weight selection. The operational ranking contains 13 forecasts in total: five core challengers and eight benchmarks, including the live

Table 4: Forecasting procedures and strengthened benchmarks

Model	Monthly or quarterly representation	Estimation	Evaluation role
Panel A: operational benchmark and core challengers			
Live AR(1)	Most recent admissible quarterly target	Recursive OLS	Synchronised quarterly benchmark
Ridge bridge	24 three-month summaries	Nested partial ridge	Dense monthly bridge
Targeted MIDAS	Five groups, 12 lags compressed to 15 terms	Nested partial ridge	Economic targeting
Factor U-MIDAS	Two raw-panel factors at three positions	Nested partial ridge	Dense common component
SG-MIDAS	24 groups and 72 dictionary terms	Sparse-group penalty	Sparse monthly component
FASG-MIDAS	72 dictionary terms plus two factors	Sparse-group penalty; factors unpenalised	Dense plus sparse component
Panel B: strengthened operational benchmarks			
BIC-AR(p)	Zero to four released quarterly lags	Recursive BIC and OLS	Flexible quarterly benchmark
Recursive mean	Released target history	Expanding mean	Low-complexity benchmark
No-change growth	Latest released quarterly growth	No estimation	Persistence benchmark
Equipment-only bridge	One equipment index; 12 lags compressed to three terms	OLS, no tuning	Direct indicator benchmark
Construction-only bridge	One construction index; 12 lags compressed to three terms	OLS, no tuning	Direct indicator benchmark
Five-challenger average	Equal weights on the five challengers	Fixed weights of 1/5	Diversification benchmark
AR-plus-five average	Equal weights on the live AR and five challengers	Fixed weights of 1/6	Conservative combination

Notes: All regressions include an intercept, the most recent admissible quarterly target when available, and unpenalised quarter-of-year indicators for NSA-QoQ. Preprocessing and tuning are recursive. The two single-indicator bridges use the same lag-based availability masks and Legendre basis as the broader models.

AR. This menu is particularly important for investment because a timely direct indicator may outperform a much broader predictor system (Baffigi et al., 2004).

4.3 Implementing the operational and fixed-operator comparisons

Experiment O is run at all twelve issue dates. At every h, each challenger forecast is produced from its current recursively fitted state and compared with the live AR computed from the same quarterly information set. This implements $d^{O}_{m,t,h}$ in Equation (3).

Experiment M is run at $h = 3, 2, 1, 0$. For each transformation, target quarter, model, and horizon, one state $S^{M}_{m,t,h}$ is fitted and shared by the no-new-month and updated-input forecasts. The quarterly target history and preprocessing cutoff are held at the pre-quarter information state. Within this frozen-state design, historical monthly feature rows are constructed at the same relative lead h as the prediction row. Thus, M does not train a representation at one issue-date lead and apply it to an input from another. The realised minimum effective sample contains 39 quarters, and the smallest initial inner training sample contains 31. At $h = 3$, the two input vertices coincide, so every model differential is exactly zero; this structural cell is retained.

Appendix B distinguishes these horizon-specific pairs from the fixed-state monthly path used to decompose the quarter-end endpoint.

4.4 Loss, inference, and multiplicity

Squared error is the primary loss function. For a candidate m and the appropriate benchmark b,

$$\mathrm{RMSE}_{m,h} = \left\{ \frac{1}{42} \sum_{t} (y_t - \widehat{y}_{m,t,h})^2 \right\}^{1/2}, \qquad \mathrm{Skill}_{m,h} = 1 - \frac{\mathrm{RMSE}_{m,h}}{\mathrm{RMSE}_{b,h}}. \tag{9}$$

MAE is a secondary descriptive measure. In O, b is the common live AR. In M, b is each model's own no-new-month counterfactual. Hence, M skill values describe input updating within a model and cannot rank models with different counterfactual denominators.

The aggregate omnibus family contains 48 cells: three transformations by twelve O horizons plus three transformations by four M horizons. Each cell contains the same 42 quarters and five challengers. For O, the cell statistic is Hansen's one-sided consistent superior predictive ability (SPA) statistic, maximised over the five challengers (Hansen, 2005). For M, it is a one-sided studentised maximum-t statistic. The reference computation uses 9,999 common circular moving-block bootstrap draws over the 42 target quarters, seed 20260830, and a main block length of four quarters. Block lengths of three and six quarters are reported as sensitivity checks. Resampling is applied to the realised loss panel and does not entail model re-estimation. Raw p-values for all 48 cells are reported, and Holm adjustment over the complete family provides the familywise screen (Holm, 1979). The three structural M cells at $h = 3$ remain in the family with statistic zero and $p = 1$. An omnibus rejection identifies a transformation–horizon–estimand cell; it is not a model-specific, post-selection p-value for the challenger with the largest sample-average gain.

The strengthened-benchmark exercise is a separate operational family. At $h = 0$, targeted MIDAS, SG-MIDAS, and FASG-MIDAS are each compared with the BIC-AR, equipment-only bridge, and construction-only bridge. The nine one-sided comparisons per transformation use 9,999 null-centred circular moving-block bootstrap draws and the same main and sensitivity block lengths. Holm-9-adjusted p-values within each transformation and Holm-27-adjusted p-values across the three transformations are reported. A substantive superiority statement additionally requires at least 5 per cent RMSE skill. These benchmark families are not pooled with the O/M family because they answer a different comparison question.

Appendix B documents the design chronology, calibration caveats, and preliminary-outcome inference.

5 Empirical Results and Robustness

5.1 Operational performance across issue dates

The first empirical result is localisation rather than broad model dominance. Figure 2 reports each core challenger's skill relative to the same-issue live AR over the twelve monthly issue dates for the focal SA-QoQ target and the NSA-YoY robustness target. Considering the complete 13-procedure menu for seasonally adjusted QoQ growth, the model with the lowest RMSE changes substantially along the information path. The equipment-only bridge is best at $h = 0$ and $h = 4$; targeted MIDAS is best at $h = 1$, $h = 2$, and $h = 5$; the construction-only bridge is best at $h = 3$; factor U-MIDAS is best at $h = 6, \ldots, 9$; and BIC-AR is best at $h = 10$ and $h = 11$. Direct monthly indicators are most competitive near the target quarter, factor compression records the lowest RMSE at several intermediate leads, and the quarterly benchmark remains difficult to improve on at the longest leads. No specification is best along the entire path.

The formal operational comparison reinforces this conclusion. The omnibus test compares the five mixed-frequency challengers with the synchronised live AR at each transformation and issue date. Among the 36 operational cells, only the seasonally adjusted QoQ cell at $h = 0$ has a raw p-value below 0.05. Its raw value is 0.0003. The next smallest operational values are 0.0579 for NSA-QoQ at $h = 0$, 0.0728 for seasonally adjusted QoQ at $h = 8$, and 0.1232 for the same target at $h = 6$. The evidence therefore supports a quarter-end operational nowcast claim, not general superiority across forecasting and nowcasting horizons.

Neither alternative target transformation yields an additional globally significant result; detailed rankings are reported in Appendix D.

5.2 The quarter-end horse race and global evidence

Table 5 broadens the $h = 0$ comparison beyond the live AR. Among all thirteen procedures, the equipment-only bridge has the lowest seasonally adjusted QoQ RMSE, 5.623. Targeted MIDAS ranks second at 5.866, followed by FASG-MIDAS at 6.018 and SG-MIDAS at 6.050. The five-challenger combination has RMSE 6.159 and the combination of the live AR with the five challengers

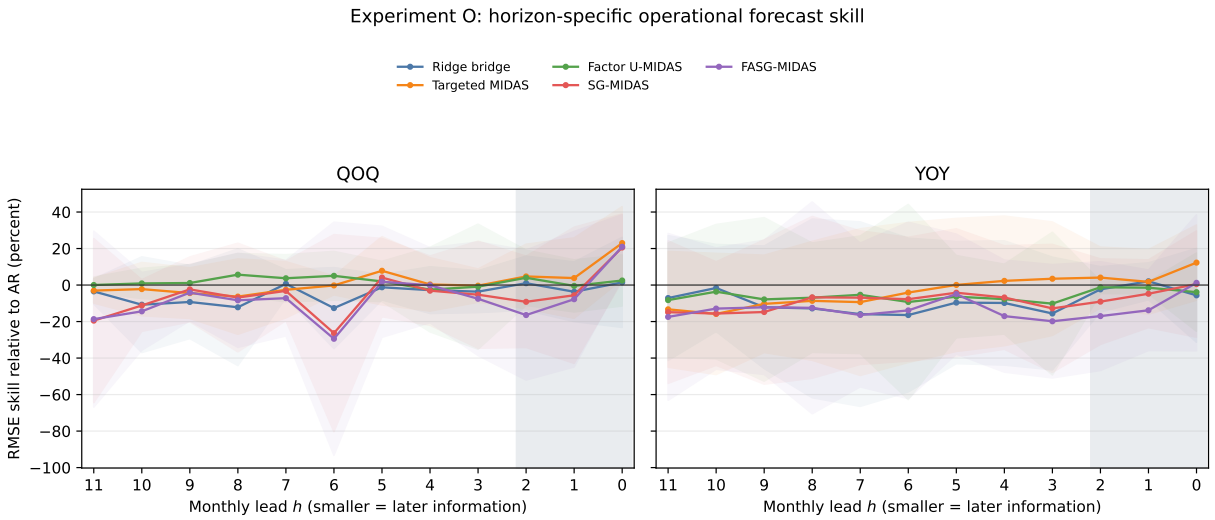

Figure 2: Operational forecast performance across twelve issue dates

Notes: The left panel uses seasonally adjusted annualised QoQ aggregate GFCF growth; the right panel uses NSA-YoY growth. Skill is one minus the ratio of a procedure's RMSE to the same-issue live-AR RMSE. Each point uses the same 42 target quarters. The vertical division distinguishes forecasts from the three within-quarter nowcasts. NSA-QoQ results are summarised in the global evidence matrix and detailed in Appendix D.

has RMSE 6.303. The BIC-AR and recursive mean both have RMSE 7.448, while the live AR has RMSE 7.616. Their equality reflects the selection of zero lags for this weakly autocorrelated target. The no-change forecast is last, with RMSE 9.935.

Table 5: Quarter-end seasonally adjusted QoQ forecast ranking

Rank	Procedure	RMSE	MAE
1	Equipment-only bridge	5.623	4.629
2	Targeted MIDAS	5.866	4.922
3	FASG-MIDAS	6.018	4.885
4	SG-MIDAS	6.050	4.594
5	Five-challenger average	6.159	4.769
6	AR-plus-five average	6.303	4.882
7	Construction-only bridge	7.256	5.671
8	Factor U-MIDAS	7.427	5.792
9	BIC-AR	7.448	5.955
9	Recursive mean	7.448	5.955
11	Ridge bridge	7.505	5.594
12	Live AR	7.616	6.114
13	No-change growth	9.935	8.167

Notes: All procedures use 42 quarters, 2016Q1–2026Q2, and the $h = 0$ information set. RMSE and MAE are percentage points of annualised log growth. The table ranks realised accuracy; formal comparisons appear in Table 6 and the global omnibus test.

The ranking has two implications. First, mixed-frequency procedures can deliver sizeable quarter-end gains relative to the live AR. Targeted MIDAS, SG-MIDAS, and FASG-MIDAS reduce RMSE by 22.98, 20.56, and 20.98 per cent. Second, complexity is not required to obtain a

large quarter-end gain. The equipment bridge improves on targeted MIDAS by 4.14 per cent and on both sparse procedures by about 7 per cent. Its ranking is consistent with timeliness and the close economic link of a direct indicator being at least as important as a broader predictor set or a more elaborate penalty.

The strong-benchmark tests in Table 6 make this qualification explicit. Targeted MIDAS improves on BIC-AR by 21.24 per cent, but its raw and Holm-9 p-values are 0.0133 and 0.0931. SG-MIDAS improves on BIC-AR by 18.76 per cent, with $p_{\text{raw}} = 0.0001$ and $p_{\text{Holm-9}} = 0.0009$, but it is 7.60 per cent worse than the equipment bridge. FASG-MIDAS improves on BIC-AR by 19.19 per cent, with $p_{\text{raw}} = 0.0024$ and $p_{\text{Holm-9}} = 0.0192$, but is 7.04 per cent worse than the equipment bridge. Only the SG-MIDAS–BIC-AR comparison remains significant when the three transformations are pooled into the 27-comparison strong-benchmark family. No focal challenger meets the prespecified pairwise economic and statistical gates against all three strong benchmarks.

Table 6: Strong-benchmark comparisons at the quarter end

Challenger	Benchmark	Skill (%)	Raw p	Holm-9	Holm-27
Targeted MIDAS	BIC-AR	21.24	0.0133	0.0931	0.3325
Targeted MIDAS	Equipment bridge	−4.32	0.7020	1.0000	1.0000
Targeted MIDAS	Construction bridge	19.16	0.0313	0.1878	0.7199
SG-MIDAS	BIC-AR	18.76	0.0001	0.0009	0.0027
SG-MIDAS	Equipment bridge	−7.60	0.7639	1.0000	1.0000
SG-MIDAS	Construction bridge	16.62	0.0383	0.1915	0.8426
FASG-MIDAS	BIC-AR	19.19	0.0024	0.0192	0.0624
FASG-MIDAS	Equipment bridge	−7.04	0.8012	1.0000	1.0000
FASG-MIDAS	Construction bridge	17.06	0.0628	0.2512	1.0000

Notes: Positive skill favours the challenger. One-sided squared-loss tests use 9,999 circular moving-block draws and the main block length of four. Holm-9 adjusts the nine displayed comparisons; Holm-27 pools the corresponding comparisons across all three target transformations. An operational omnibus rejection against the live AR does not identify a winner in this table.

Multiplicity across the full search is more restrictive still. The 48-cell family contains all 36 operational cells and the twelve fixed-operator input cells. At the main four-quarter block length, three cells have raw p-values below 0.05: O/SA-QoQ/$h = 0$ (0.0003), M/SA-QoQ/$h = 0$ (0.0052), and M/NSA-YoY/$h = 0$ (0.0255). After Holm adjustment over all 48 cells, only O/SA-QoQ/$h = 0$ remains significant, with $p_{\text{Holm-48}} = 0.0144$. Its Bonferroni value is also 0.0144, and its Holm-48 values remain below 0.05 at block lengths three and six, at 0.0192 and 0.0048. This is the strongest formal result: the omnibus evidence implies that at least one mixed-frequency challenger improves on the synchronised live AR at the end of the quarter. It neither identifies a named challenger nor implies that any challenger beats the equipment bridge. The result does not extend to other horizons or transformations.

5.3 Fixed-operator value of within-quarter updates

The operational result does not reveal whether its gain is generated by newly admitted monthly inputs or by the complete forecasting system. The fixed-operator comparison addresses the nar-

rower question. Table 7 applies each model's quarter-end fitted state to the four input vertices $X^{(0)}, \ldots, X^{(3)}$, where $X^{(0)}$ is the pre-quarter input and $X^{(3)}$ is the quarter-end input. Comparing $X^{(3)}$ with $X^{(0)}$, all five procedures have positive cumulative seasonally adjusted QoQ skill relative to their own no-new-month counterfactual. The gains range from 6.66 per cent for factor U-MIDAS to 29.81 per cent for FASG-MIDAS.

Table 7: Fixed-operator within-quarter update path

	$X^{(0)}$	$X^{(3)}$	Endpoint	Mean loss contribution		
Procedure	RMSE	RMSE	skill (%)	Month 1	Month 2	Month 3
Ridge bridge	8.732	7.564	13.38	15.079	4.313	−0.355
Targeted MIDAS	7.757	5.876	24.25	12.349	1.073	12.219
Factor U-MIDAS	7.822	7.301	6.66	9.935	−1.917	−0.145
SG-MIDAS	8.618	6.699	22.27	3.611	2.496	23.282
FASG-MIDAS	9.440	6.626	29.81	12.043	9.655	23.513

Notes: The target is seasonally adjusted annualised QoQ growth and there are 42 quarters. $X^{(0)}$ is the pre-quarter input and $X^{(3)}$ the quarter-end input, evaluated with the same $h = 0$ state. Endpoint skill uses each model's own $X^{(0)}$ RMSE. The final three columns are mean reductions in squared loss, not RMSE skills; their sum equals the endpoint mean loss gain.

These percentages are not a model ranking. FASG-MIDAS has the largest percentage gain partly because its no-new-month RMSE is 9.440; its updated RMSE is 6.626. Targeted MIDAS has a smaller percentage gain but a lower updated RMSE, moving from 7.757 to 5.876. The different denominators explain why fixed-operator skill must be interpreted within each procedure.

Monthly contributions are episodic and may be negative; their model-by-month patterns and interpretation are reported in Appendix C.

The familywise evidence is weaker than the economic magnitudes. For the focal endpoint, the fixed-operator omnibus raw p-value is 0.0052, but its Holm-48 value is 0.2444. Adjusted values at block lengths three and six are 0.3854 and 0.0987. The NSA-YoY endpoint has a raw value of 0.0255 and a Holm-48 value of 1.0000. For NSA-QoQ, endpoint skills range from −0.58 to 11.50 per cent and the omnibus raw value is 0.1183. The data therefore support sizeable descriptive fixed-operator gains at the focal endpoint, but not a familywise claim that monthly-input value is positive after accounting for the full transformation, horizon, and estimand search.

5.4 Benchmark-synchronisation simulation

The benchmark-synchronisation simulation shows why O and M must remain separate. It uses 1,000 replications in each signal cell, 42 recursive out-of-sample quarters, and a one-candidate design. When both the monthly signal and target update are absent, tests based on the held, synchronised operational, and fixed-operator comparisons reject in 1.5, 1.1, and 1.4 per cent of replications. When the target update is informative but monthly predictive content is zero, the test based on the held comparison rejects in 45.1 per cent, while tests based on O and M reject in only 0.9 and 1.1 per cent. The stricter event in which the held-comparison test rejects while both controlled tests do not occurs in 44.1 per cent. Figure 3 reports the full set of rejection rates across

the four signal designs.

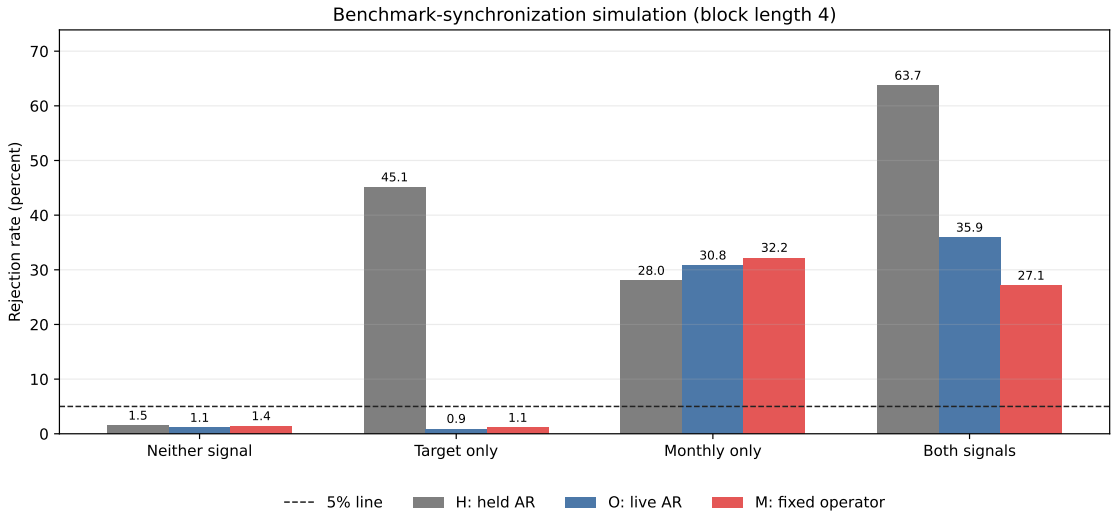

Figure 3: Benchmark-synchronisation simulation

Notes: Rejection rates are based on 1,000 replications, 42 recursive evaluation quarters, 999 bootstrap draws, and the main block length of four. The target-only cell has an informative quarterly-target update but no monthly predictive signal. The held comparison's rejection in that cell is a valid composite replacement result; it is an attribution error only if described as monthly-input evidence.

Appendix D reports the component evidence, the remaining simulation cells, preliminary-outcome rescoring, and additional robustness diagnostics. Table 8 consolidates the quarter-end evidence across estimands and target transformations.

Table 8: Quarter-end evidence across estimands and target transformations

Cell	Largest descriptive skill	Raw p	Holm-48	Interpretation
O, SA-QoQ	22.98% versus live AR	0.0003	0.0144	Headline evidence
M, SA-QoQ	29.81% versus own baseline	0.0052	0.2444	Secondary
O, NSA-YoY	12.28% versus live AR	0.1369	1.0000	No familywise support
M, NSA-YoY	20.63% versus own baseline	0.0255	1.0000	Raw only
O, NSA-QoQ	12.77% versus live AR	0.0579	1.0000	No familywise support
M, NSA-QoQ	11.50% versus own baseline	0.1183	1.0000	Descriptive

Notes: All entries use $h = 0$ and 42 quarters. The descriptive skill is the largest among the five core challengers and does not itself constitute post-selection inference. The p-values are model-set omnibus values; Holm-48 covers all transformation, horizon, and estimand cells.

The combined evidence supports one bounded conclusion. At least one mixed-frequency procedure improves on the synchronised live AR for the quarter-end forecast of seasonally adjusted investment growth. The gain is horizon-specific, does not establish an improvement over the best direct-indicator bridge, and cannot be promoted to a general result about every monthly update, target transformation, model, or data vintage.

6 Discussion and Conclusion

The central empirical result favours a mixed-frequency operational procedure but is narrower than a conventional comparison with a held quarterly benchmark would suggest. Across the 48-cell family spanning estimands, target transformations, and issue dates, only the operational comparison for seasonally adjusted annualised QoQ GFCF growth at $h = 0$ survives the global familywise adjustment. Its raw omnibus p-value is 0.0003 and its Holm-48-adjusted value is 0.0144, with the conclusion unchanged under alternative block lengths. The evidence supports the claim that at least one mixed-frequency procedure can improve on the synchronised live AR at the end of the target quarter. It does not establish a general advantage at earlier horizons, for alternative transformations, or for the full class of MIDAS models.

The stronger benchmark comparison further qualifies the result. At $h = 0$, the equipment-only bridge records the lowest RMSE, 5.623, compared with 5.866 for targeted MIDAS, 6.018 for FASG-MIDAS, 6.050 for SG-MIDAS, and 7.616 for the live AR. The leading high-dimensional procedures record substantially lower RMSEs than the synchronised AR, and some also improve on the BIC-selected AR after pairwise multiplicity adjustment. None, however, meets the stated pairwise economic and statistical gates against the BIC-AR, equipment bridge, and construction bridge. Model complexity is therefore not necessary to obtain a quarter-end gain. A single-indicator, economically targeted bridge is at least as accurate in this sample as procedures that shrink, group, or compress a larger predictor set.

This ranking has a direct practical implication. A forecasting institution considering a high-dimensional nowcasting system should compare it not only with an autoregression but also with direct-indicator bridges that encode the relevant expenditure mechanism. For Korean investment, the equipment bridge is a substantive competitor rather than a weak baseline. It is transparent, parsimonious, and difficult for more elaborate models to beat in this sample. The appropriate conclusion is not to discard high-dimensional methods: targeted and sparse procedures remain competitive alternatives to the live AR. It is to retain the equipment bridge as a permanent benchmark and require added complexity to demonstrate value beyond it.

The fixed-operator results clarify why an absence of broad model dominance does not imply that monthly forecast updates are uninformative. Holding the quarter-end operator and target state fixed, moving from $X^{(0)}$ to $X^{(3)}$ lowers RMSE for all five challengers. The cumulative reductions range from 6.66 per cent for factor U-MIDAS to 29.81 per cent for FASG-MIDAS. These are within-model measures and cannot rank procedures. The improvement is also uneven. Across model–month cells, only 38.1 to 66.7 per cent of the individual forecast-update contributions are positive, depending on the model and update month. The raw M-test p-value at $h = 0$ is 0.0052, but its Holm-48 value is 0.2444. Monthly updates can have material cumulative value while remaining episodic and model-specific; a positive monthly-input-value claim is not supported by the global inferential screen.

Variation across horizons points in the same direction. The operational procedure with the lowest RMSE changes as the quarter approaches: factor-based methods record the lowest RMSE

at several intermediate leads, targeted specifications become more competitive closer to the target quarter, and the equipment bridge is strongest at quarter end. Neither NSA-QoQ nor NSA-YoY produces an additional globally significant result. The component exercise is also most favourable for equipment investment, while evidence for construction and IPP is weaker. These patterns are consistent with an economic interpretation centred on a timely equipment signal, but the component comparisons are descriptive and the chain-volume aggregate is not exactly additive. They are not a structural decomposition of aggregate forecast gains.

The simulation explains why the evaluation design matters beyond this application. When monthly predictors contain no predictive information but a new quarterly target observation is informative, the test based on the held comparison rejects in 45.1 per cent of replications, whereas tests based on O and M reject in 0.9 and 1.1 per cent. The 45.1 per cent rate is not a size estimate for H: the informative target update makes the held-forecast replacement estimand positive. The attribution error arises only when this legitimate replacement rejection is described as evidence that monthly inputs help. Separating H, O, and M makes the claim attached to each comparison explicit.

Several limitations bound the empirical conclusions. The main analysis contains 42 evaluation quarters and therefore has limited power for comparing many models over twelve issue dates. It uses August 2026 latest-vintage predictor histories. Actual quarterly target-release dates are observed from 2009Q4 onward; earlier training quarters use a 28-day rule. The monthly ragged edge is constructed from deterministic series-specific lag assumptions rather than archived historical publication timestamps. The exercise is calendar-based pseudo-real-time, not a complete real-time backtest. The archive of 44 preliminary-stage aggregate outcomes is too short to re-estimate the recursive model menu while preserving an adequate evaluation window. In the feasible 41-quarter outcome-only diagnostic, latest-vintage training histories and forecasts remain frozen and only the evaluation outcomes are replaced. The favourable ranking of targeted MIDAS, SG-MIDAS, and FASG-MIDAS against the live AR is not reversed, but this exercise cannot establish the effects of real-time predictor vintages or recursive re-estimation.

The highest-value extension is therefore a versioned archive containing the first release and subsequent revisions of quarterly investment and monthly predictors together with actual publication timestamps. Such an archive would permit a genuinely real-time replication, a distinction between news and revisions, and cleaner attribution of each update. A longer sample or harmonised cross-country application would provide more power to study state dependence, structural change, and horizon-specific model selection. Forecast combinations that condition on issue date and give persistent weight to economically matched bridge indicators are another promising direction, provided that their tuning remains fully recursive.

The broader conclusion is methodological as well as applied. For seasonally adjusted annualised QoQ GFCF growth at the end of the quarter, at least one mixed-frequency procedure can improve on the synchronised live AR, while the equipment bridge records the lowest RMSE in the thirteen-procedure horse race. Fixed-operator updates produce economically meaningful cumulative gains,

but the associated omnibus cell does not survive Holm-48 adjustment and their monthly contributions are not uniformly positive. A credible claim therefore requires a live same-origin benchmark, a fixed-operator input counterfactual, strong direct-indicator competitors, and multiplicity control over the reported estimand, horizon, and transformation family. Once those disciplines are imposed, the evidence supports selective operational evaluation of mixed-frequency procedures rather than general MIDAS dominance.

Data availability and declarations

Data availability. The national-accounts targets and monthly indicator series are publicly available from the Bank of Korea ECOS database. Dated quarterly target-release announcements from 2009Q4 onward are available from the Bank of Korea website; earlier training-quarter release dates follow a 28-day rule. Historical monthly availability is represented by documented series-specific lag assumptions, rather than by a complete archive of publication timestamps. The GFCF-to-GDP share is obtained from the World Bank's World Development Indicators Database Archives. The main analysis uses the August 2026 latest-vintage predictor panel and is therefore a calendar-based pseudo-real-time exercise, not a complete historical-vintage backtest.

Funding. An earlier stage of this research received financial support from the Bank of Korea. No grant number was assigned.

Competing interests. The author declares no competing interests.

Disclaimer. The author has no current affiliation with the Bank of Korea. The views expressed are solely those of the author.

References

Babii, A., Ghysels, E., and Striaukas, J. (2022). Machine learning time series regressions with an application to nowcasting. *Journal of Business & Economic Statistics*, 40(3):1094–1106.

Baffigi, A., Golinelli, R., and Parigi, G. (2004). Bridge models to forecast the euro area GDP. *International Journal of Forecasting*, 20(3):447–460.

Bańbura, M., Giannone, D., Modugno, M., and Reichlin, L. (2013). Now-casting and the real-time data flow. In Elliott, G. and Timmermann, A., editors, *Handbook of Economic Forecasting*, volume 2A, chapter 4, pages 195–237. Elsevier.

Bank of Korea (2026a). Economic statistics publication calendar and quarterly national accounts releases. `https://www.bok.or.kr/portal/stats/statsPublictSchdul/listCldr.do?menuNo=200775`. Historical calendar pages accessed in August 2026.

Bank of Korea (2026b). Economic statistics system (ECOS). `https://ecos.bok.or.kr/`. Data retrieved 27 August 2026.

Beyhum, J. and Striaukas, J. (2026). Factor-augmented sparse MIDAS regressions with an application to nowcasting. *Journal of Business & Economic Statistics*. Published online.

Bok, B., Caratelli, D., Giannone, D., Sbordone, A. M., and Tambalotti, A. (2018). Macroeconomic nowcasting and forecasting with big data. *Annual Review of Economics*, 10:615–643.

Clark, T. E. and West, K. D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics*, 138(1):291–311.

Croushore, D. (2006). Forecasting with real-time macroeconomic data. In Elliott, G., Granger, C. W. J., and Timmermann, A., editors, *Handbook of Economic Forecasting*, volume 1, chapter 17, pages 961–982. Elsevier.

Croushore, D. and Stark, T. (2001). A real-time data set for macroeconomists. *Journal of Econometrics*, 105(1):111–130.

Degiannakis, S. (2022). Stock market as a nowcasting indicator for real investment. *Journal of Forecasting*, 41(5):911–919.

Foroni, C. and Marcellino, M. (2014). A comparison of mixed frequency approaches for nowcasting euro area macroeconomic aggregates. *International Journal of Forecasting*, 30(3):554–568.

Foroni, C., Marcellino, M., and Schumacher, C. (2015). Unrestricted mixed data sampling (MIDAS): MIDAS regressions with unrestricted lag polynomials. *Journal of the Royal Statistical Society: Series A*, 178(1):57–82.

Fosten, J. and Gutknecht, D. (2021). Horizon confidence sets. *Empirical Economics*, 61(2):667–692.

Ghysels, E., Sinko, A., and Valkanov, R. (2007). MIDAS regressions: Further results and new directions. *Econometric Reviews*, 26(1):53–90.

Giannone, D., Reichlin, L., and Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, 55(4):665–676.

Hansen, P. R. (2005). A test for superior predictive ability. *Journal of Business & Economic Statistics*, 23(4):365–380.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2):65–70.

Holzmann, H. and Eulert, M. (2014). The role of the information set for forecasting—with applications to risk management. *The Annals of Applied Statistics*, 8(1):595–621.

Kim, H. H. and Swanson, N. R. (2018). Methods for backcasting, nowcasting and forecasting using factor-MIDAS: With an application to korean GDP. *Journal of Forecasting*, 37(3):281–302.

Kuzin, V., Marcellino, M., and Schumacher, C. (2011). MIDAS vs. mixed-frequency VAR: Nowcasting GDP in the euro area. *International Journal of Forecasting*, 27(2):529–542.

Marcellino, M. and Schumacher, C. (2010). Factor MIDAS for nowcasting and forecasting with ragged-edge data: A model comparison for german GDP. *Oxford Bulletin of Economics and Statistics*, 72(4):518–550.

McCarthy, M. and Snudden, S. (2025). Predictable by construction: Assessing forecast directional accuracy of temporal aggregates. *Applied Economics*, pages 1–16. Published online 16 October 2025.

Siliverstovs, B. (2017). Short-term forecasting with mixed-frequency data: A MIDASSO approach. *Applied Economics*, 49(13):1326–1343.

World Bank (2026). World development indicators database archives. `https://databank.worldbank.org/databases/archives`. Korea archive versions through January 2026.

Appendices

A Data Construction and Vintage Coverage

A.1 Alternative target transformations

Two alternative transformations of the corresponding non-seasonally adjusted levels assess whether the findings depend on published seasonal adjustment or on the growth convention. The first is the annualised one-quarter log change

$$y_{j,t}^{q,NSA} = 400\{\log(Y_{j,t}^{NSA}) - \log(Y_{j,t-1}^{NSA})\}, \tag{10}$$

for which the forecasting equations include unpenalised quarter-of-year indicators. The second is the four-quarter log change

$$y_{j,t}^{y} = 100\{\log(Y_{j,t}^{NSA}) - \log(Y_{j,t-4}^{NSA})\}. \tag{11}$$

The YoY transformation removes deterministic quarterly seasonality without using an adjusted series, but it produces a more persistent, overlapping target series. RMSE magnitudes are therefore compared only within a transformation, never across QoQ and YoY scales. Table 2 in the main text summarises the target definitions, component shares, and focal-target volatility.

A.2 Monthly transformations and availability masks

Activity and service series that are seasonally adjusted by their providers enter as monthly log changes. Non-seasonally adjusted monthly flows and stocks generally enter as 12-month log changes. The cumulative housing-approval series is first converted into a monthly flow and then transformed as a 12-month log change. Survey indices are expressed as deviations from 100. Daily financial prices and yields are aggregated to calendar-month means before growth rates or spreads are constructed. The leading indicator is standardised using only the relevant training window. The aligned panel spans January 1970 through August 2026, although individual series begin on different dates and the shortest histories start in September 2009.

At a given issue date, a monthly observation is admitted only if the end of its reference period plus its assigned lag is no later than the forecast origin. Unavailable observations are masked before training-window imputation, standardisation, factor extraction, or regression construction. A fixed predictor schema is retained across origins, so changing availability alters the ragged edge rather than the identity of the candidate model. The lag classes should not be interpreted as 25 empirically verified historical publication schedules. A lag that is too long may attenuate the measured quarter-end input contribution, whereas a lag that is too short may admit information prematurely. These opposing risks are one reason for classifying the exercise as pseudo-real-time.

A.3 Quarterly release timing and the preliminary archive

Quarterly target availability is determined from dated Bank of Korea advance-release calendars (Bank of Korea, 2026a). Actual dates are available from 2009Q4 onward and cover all relevant quarterly release events over the evaluation period. For earlier training quarters, the registry assigns an advance-release date 28 days after quarter-end. The previous quarter's advance estimate normally enters during the first month of the target quarter, but its admission is determined by the observed publication date rather than by a fixed quarterly lag. The AR and monthly models consequently share the same admissible quarterly target history at a given issue date, while only the monthly models process the evolving predictor edge.

A separate archive of Bank of Korea national-accounts publications contains 44 consecutive preliminary estimates of aggregate GFCF from 2015Q2 through 2026Q1. Each observation is a published real, seasonally adjusted simple QoQ growth rate, converted to the target scale as $400\log(1 + q_t/100)$. Relative to the August 2026 series, the median absolute revision in simple QoQ growth is 0.705 percentage points, the mean absolute revision is 0.885 points, and the largest absolute revision is 2.673 points in 2018Q3. The archive is too short for full recursive re-estimation of the five challengers under preliminary target histories. The feasible diagnostic therefore freezes the fitted models and point forecasts and replaces only the evaluation outcome for 41 quarters, 2016Q1–2026Q1. The resulting evidence is described as preliminary-outcome rescoring rather than as evidence from vintage-trained forecasts.

B Recursive Estimation and Implementation Details

B.1 Additional preprocessing conditions

After origin-specific standardisation, unavailable monthly values are set to zero, which corresponds to the training-window mean under that scaler. A fixed column schema is retained across origins. The monthly samples used for principal-component extraction begin in January 2000 and must span at least 120 calendar months. A series is eligible if it contains at least 60 observations at an outer cutoff or at least 24 observations at an inner-fold cutoff. All standardisation, factor extraction, tuning, and coefficient estimation are repeated recursively; no full-sample preprocessing state is carried into an out-of-sample origin.

B.2 Mixed-frequency specifications

Five mixed-frequency procedures span dense, economically targeted, factor, sparse, and factor-augmented specifications. They all include an intercept, the most recent admissible quarterly target, and any applicable seasonal controls as unpenalised regressors. The ridge bridge uses one availability-filtered summary for each of the 24 monthly predictor groups: the mean of its three most recent standardised calendar positions. Its ridge penalty is selected from $\{0.1, 1, 10, 100\}$.

The targeted MIDAS model retains five predictors prespecified on economic grounds for their

relationship with aggregate GFCF: all-industry production, the leading composite index, the seasonally adjusted equipment-investment index, real seasonally adjusted completed construction, and the credit spread. Twelve monthly lags of each predictor are projected onto three shifted-Legendre terms of degrees zero through two, normalised to unit root-mean-square. The resulting 15 lag-basis coefficients are estimated by partial ridge using the same inner penalty grid as the dense bridge. This procedure combines economic targeting and lag-polynomial compression with regularisation; it is not an unregularised restricted MIDAS regression (Ghysels et al., 2007).

Factor U-MIDAS recursively extracts two principal components from the eligible raw monthly panel and includes the three most recent monthly positions of each factor without imposing a lag polynomial. Its six factor-position coefficients are ridge regularised (Marcellino and Schumacher, 2010; Foroni et al., 2015). SG-MIDAS instead maps twelve lags of every predictor group into the same three-term Legendre dictionary, producing 72 candidate terms. It estimates them with the sparse-group penalty

$$\lambda \left[(1-\alpha) \sum_{g=1}^{24} \sqrt{p_g} \|\beta_g\|_2 + \alpha \|\beta\|_1 \right], \qquad p_g = 3, \quad \alpha = 0.5. \tag{12}$$

FASG-MIDAS adds two recursively estimated principal components of the complete 72-term dictionary to the unpenalised part of this regression, allowing dense common movement and sparse predictor-specific departures (Babii et al., 2022; Beyhum and Striaukas, 2026). For both sparse models, α is fixed and $\lambda/\lambda_{\max} \in \{0.05, 0.1, 0.2, 0.4, 0.8, 1\}$ is chosen by the eight inner folds, with ties favouring the larger penalty. The sparse optimisations use a relative objective tolerance of 10^{-8} and a maximum KKT violation of 10^{-7}; all outer and inner problems used in the reported forecasts converged. Table 4 in the main text summarises the complete set of forecasting procedures and benchmarks.

B.3 Horizon-specific fixed-operator implementation

The horizon-specific implementation differs from the fixed-state update path detailed in Appendix C. The reported M value at $h = 2$, for example, uses a separately fitted $S^M_{m,t,2}$. By contrast, the intermediate $X^{(1)}_t$ forecast in the $X^{(0)}$–$X^{(3)}$ path applies the quarter-end state $S^M_{m,t,0}$. The same $h = 0$ state is applied to all four input panels solely to allocate the endpoint update over three monthly steps. The path's $X^{(1)}$ and $X^{(2)}$ forecasts are not copied from the horizon-specific $h = 2$ and $h = 1$ forecasts.

B.4 Design disclosures and auxiliary inference

The study was not preregistered. The expanded model menu and the initial quarter-end reporting rules were fixed after earlier diagnostics but before the final expanded-model forecasts were produced. NSA-QoQ, the stronger benchmark menu, and the 48-cell global screen were revision-stage additions. The global adjustment disciplines the realised search path; it does not retrospectively

turn the study into a prospective confirmatory experiment. A separate simulation matched to the evaluation-sample length also did not establish exact finite-sample calibration of the moving-block procedures. These procedures are retained as the stated inferential reporting convention, while inference is interpreted jointly with effect sizes, block-length stability, results under alternative target transformations, and the stronger benchmarks.

For the preliminary-outcome diagnostic, frozen O forecasts are rescored against preliminary aggregate GFCF outcomes for the 41 common quarters. Forecast origins, estimation histories, preprocessing states, tuning choices, coefficients, and point forecasts are unchanged. At $h = 0$, raw squared-loss differentials are assessed using a Bartlett-kernel HAC variance estimator with autocovariances through lag four. A Clark–West loss differential is also reported as a nested-model diagnostic (Clark and West, 2007). Holm adjustment is applied across the five challengers separately for the raw and Clark–West statistics. Because the challengers are tuned or regularised rather than fixed-dimensional nested OLS models, the Clark–West statistic is treated as a diagnostic rather than as an exact test.

C Fixed-Operator Information and Update Accounting

C.1 Oracle information value and the monthly update path

Nested information has non-negative value for an oracle squared-error forecast. If $\mathcal{F}_0 \subseteq \mathcal{F}_1$, $Y \in L^2$, and $\mu_j = E(Y \mid \mathcal{F}_j)$, then the standard projection argument gives

$$E\{(Y - \mu_0)^2 - (Y - \mu_1)^2\} = E\{(\mu_1 - \mu_0)^2\} \geq 0. \tag{13}$$

This result concerns two ideal conditional means (Holzmann and Eulert, 2014). Experiment M instead applies one estimated, regularised, and potentially misspecified operator to two input vectors. Equation (13) therefore does not imply that $\theta^M_{m,h}$ is non-negative for an estimated or misspecified procedure. Nor must its sample counterpart be positive. This is why an input can be informative in principle while a particular implementation fails to use it reliably.

The within-quarter accounting used in Section 5 sharpens this distinction. Let $X_t^{(0)}$ denote the lag-masked monthly panel at the end of the month before the target quarter, and let $X_t^{(1)}, X_t^{(2)}, X_t^{(3)}$ denote the panels at the ends of the first, second, and third months of that quarter. Thus, for $h = 3, 2, 1, 0$, the current panel is indexed by $k = 3 - h$. Holding a single $h = 0$ (quarter-end) state fixed, define

$$\widetilde{g}^{(k)}_{m,t} = f_m(S^M_{m,t,0}, X_t^{(k)}, T^0_{t,0}), \qquad k = 0, 1, 2, 3. \tag{14}$$

The contribution of the kth forecast update is

$$\Delta_{m,t,k} = L(y_t, \widetilde{g}^{(k-1)}_{m,t}) - L(y_t, \widetilde{g}^{(k)}_{m,t}), \qquad k = 1, 2, 3, \tag{15}$$

so that the endpoint comparison telescopes exactly:

$$\sum_{k=1}^{3} \Delta_{m,t,k} = L(y_t, \widetilde{g}_{m,t}^{(0)}) - L(y_t, \widetilde{g}_{m,t}^{(3)}) = d_{m,t,0}^{M}. \tag{16}$$

The intermediate forecasts in this path keep the $h = 0$ state fixed. They are not the updated forecasts from the separately fitted $h = 2$ and $h = 1$ Experiment M pairs. Moreover, moving from $X_t^{(k-1)}$ to $X_t^{(k)}$ admits observations that pass their series-specific lag masks and advances the rolling lag window. Hence $\Delta_{m,t,k}$ is a one-month forecast-update contribution under a fixed operator, not the causal effect of an isolated data release, a release surprise, or a vintage revision.

For squared loss, the sign also depends on the direction of the forecast error, not only on the size of the revision. For $\Delta f = f_1 - f_0$,

$$(y - f_0)^2 - (y - f_1)^2 = 2(y - f_0)\Delta f - (\Delta f)^2 = \Delta f(2y - f_0 - f_1). \tag{17}$$

A large forecast revision is thus neither necessary nor sufficient for an accuracy improvement. Finally, the percentage skill reported for M is normalised by each model's own no-new-month RMSE. It is meaningful within a fitted procedure but cannot be used to rank competing models. These restrictions motivate the separate operational horse race, fixed-operator path, and benchmark-synchronisation simulation in the main text.

C.2 Monthly contribution patterns

The monthly decomposition also shows that the average endpoint gains are not delivered uniformly. The first within-quarter update has a positive mean loss contribution for every model. The second contribution is positive for four models but negative for factor U-MIDAS. The third is strongly positive for targeted MIDAS, SG-MIDAS, and FASG-MIDAS, but slightly negative for the ridge bridge and factor U-MIDAS. Across the fifteen model–month cells, the share of individual quarters with a positive contribution ranges from 38.1 to 66.7 per cent, depending on the model and update month. A mean contribution can be negative even when a majority of quarters are positive, as for the final ridge update, because a few adverse revisions have large squared-error costs. Moreover, a transition from $X^{(k-1)}$ to $X^{(k)}$ admits all observations passing their lag-based availability masks and advances the rolling lag window. It is a forecast-update contribution, not a causal attribution to a named release or surprise.

D Additional Empirical Results and Robustness

D.1 Alternative target transformations

The alternative target transformations lead to the same caution. For NSA-YoY growth, targeted MIDAS has the lowest quarter-end RMSE, 2.449, but the operational omnibus p-value is 0.1369. Its RMSE gains relative to BIC-AR, the equipment bridge, and the construction bridge are 14.14,

7.24, and 4.31 per cent, respectively; none survives transformation-specific Holm adjustment. For NSA-QoQ, targeted MIDAS is again best at $h = 0$, with RMSE 11.386, but BIC-AR and the equipment bridge are close, at 11.511 and 11.541. The omnibus p-value is 0.0579, and BIC-AR has the lowest RMSE at $h = 1, \ldots, 8$. These outcomes make seasonally adjusted QoQ the focal case rather than one favourable member of a uniformly successful set.

D.2 Investment components

The component results help interpret the aggregate pattern, although they form a secondary descriptive family. Equipment investment provides the clearest evidence. In O, targeted MIDAS reduces equipment RMSE by 24.31 per cent relative to the live AR; its pointwise 95 per cent interval is $[7.51, 37.41]$. In M, FASG-MIDAS has a 32.67 per cent gain relative to its own no-new-month forecast, with interval $[22.72, 41.75]$. The evidence is smaller for other components. For construction, the largest O skill is 9.59 per cent for FASG-MIDAS and the largest M skill is 9.83 per cent for SG-MIDAS; both intervals include zero. For IPP, factor U-MIDAS has the largest O skill, 5.61 per cent, while the largest M skill is 6.64 per cent for the ridge bridge. These entries are selected from 30 component–estimand–model cells and the intervals are pointwise rather than familywise.

A component-share approximation provides a scale check. Over 82 quarters from 2006Q1 through 2026Q2, the difference between component-share-implied aggregate growth and official aggregate growth has RMSE 0.1995 percentage points, or 2.59 per cent of the standard deviation of aggregate growth in the same sample. The small discrepancy permits component evidence to inform the economic interpretation, but does not turn chain-volume components into an additive national-accounts identity. Taken together, the results are consistent with equipment-related monthly information playing an important role in quarter-end performance, but they do not identify the structural source of the aggregate gain.

D.3 Additional simulation cells

In the target-update-only cell of Figure 3, rejection based on the held comparison is not intrinsically erroneous: the decision rule that replaces a stale AR forecast with the current procedure has genuinely benefited from new quarterly-target information. The error would be to attribute that rejection to monthly predictors. When only monthly information is predictive, rejection rates are 28.0, 30.8, and 32.2 per cent for the held, operational, and fixed-operator comparisons. With both signals, they are 63.7, 35.9, and 27.1 per cent. These values illustrate attribution in a stylised design. They are neither calibrated predictions for Korea nor a validation of the empirical bootstrap.

D.4 Preliminary-outcome rescoring

Outcome revisions are economically material. As documented in Appendix A, the available Korean archive contains 44 consecutive preliminary-stage aggregate outcomes, but the recursive design requires 32 prior target quarters. It would leave at most twelve fully re-estimated evaluation ob-

servations, below the sample gates for reporting and inference. A full real-time five-model backtest is therefore not feasible and is not claimed.

The narrower outcome-only diagnostic retains the frozen forecasts, issue dates, model states, and latest-vintage training histories, but scores them against preliminary aggregate outcomes for 41 common quarters. At $h = 0$, the live AR has RMSE 8.784. FASG-MIDAS, targeted MIDAS, and SG-MIDAS have RMSEs of 7.090, 7.239, and 7.431, corresponding to gains of 19.3, 17.6, and 15.4 per cent. Tests based on their raw squared-loss differentials remain significant after Holm adjustment over the five challengers, with adjusted p-values 0.0100, 0.0100, and 0.0006. Clark–West diagnostics yield the same three-model classification. The ridge bridge and factor U-MIDAS improve by only 3.3 and 1.7 per cent. This exercise shows that the favourable ranking against the live AR is not reversed when the evaluation outcome is changed from the latest estimate to the preliminary-stage value. Because the training targets and predictors remain latest-vintage, this is only a partial preliminary-outcome diagnostic. It has not been extended to the equipment and construction benchmarks.

D.5 Other robustness checks

Other checks qualify rather than overturn the result. Removing 2020Q1–2021Q4 leaves 34 quarters. Targeted MIDAS still improves on BIC-AR by 21.66 per cent for seasonally adjusted QoQ growth but remains 4.66 per cent worse than the equipment bridge. SG-MIDAS continues to improve on BIC-AR after the 27-comparison adjustment, whereas no focal procedure dominates all three strong benchmarks. A separate calibration audit did not establish reliable finite-sample size control for the moving-block procedures; their p-values are therefore interpreted jointly with effect sizes and block-length sensitivity. Predictor timing also remains calendar-based, and sparse tuning is too variable to support structural feature rankings: selected penalties change between adjacent origins in 30.1 per cent of aggregate sparse paths and the number of active groups changes in 48.0 per cent.