Auditing Real-Time Official Statistics: A Release-Event Ledger for U.S. GDP Nowcasting

Hyun Hak Kim*

August 2026

Abstract

Official economic statistics arrive in public release packages rather than immutable monthly tables. Reproducible real-time analysis must preserve when information became public, which observations and revisions were released jointly, and evidence for their timing. We develop a release-event ledger and audit framework that reconstructs origin-specific information sets. For four U.S. monthly labor-market and industrial-production series used in GDP nowcasting (2004–2025), the ledger records 8,475 events in 545 packages and distinguishes verified release times from conservative end-of-day assignments. At common month-end cutoffs, 138,732 of 138,864 candidate cells match archived FRED-MD vintages exactly, but monthly files do not recover intramonth release order. Across 387 sealed origins, assigning the previous labelled vintage delays public information at every origin and changes the unchanged forecasting-rule output by 0.953 percentage points on average in absolute value; assigning the same-month vintage introduces post-origin information at 202 origins and yields a 0.238-point discrepancy. Frozen package replay separates headline, revision, and interaction responses without causal-news claims. The contribution is an auditable dissemination-layer data contract and reproducibility protocol, not a claim of generally superior forecasts.

Keywords: official statistics; real-time data; data revisions; vintage data; information sets; nowcasting; reproducibility.

---

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

1 Introduction

Official economic time series reach users through a sequence of releases rather than as immutable tables. Each released value refers to an economic period, becomes public at a later time, and may subsequently be revised together with other rows in the same statistical product. A data representation adequate for latest-value analysis is therefore not necessarily adequate for reconstructing what a user could have known at a historical decision time. Official-statistics quality frameworks treat fitness for use as multidimensional, emphasizing such properties as reliability, timeliness, coherence, accessibility, clarity, and documented metadata (Biemer et al., 2014; European Statistical System, 2017; United Nations Statistics Division, 2019). Statistical-production frameworks likewise place dissemination and evaluation inside a process supported by cross-cutting metadata management (Signore et al., 2015; United Nations Economic Commission for Europe, 2025). These principles imply that a reproducible real-time analysis requires more than the latest value and its reference period: it also requires an auditable public availability time, release-package identity, revision lineage, and provenance.

A conventional monthly vintage records a valuable snapshot of the data state and is indispensable for avoiding ex-post substitution (Croushore and Stark, 2001; Koenig et al., 2003). By construction, however, a month-end snapshot need not identify the order of releases within that month, the rows released atomically in one official package, or whether a row entered as a new observation or as a revision. Those omissions are consequential whenever a monthly file is assigned to an earlier intramonth decision time, and they also limit the ability to audit how a downstream statistical output changed from one official release to the next.

This paper asks how a release-event representation of official macroeconomic statistics can be constructed, validated, and used to audit a real-time information set. The empirical case links monthly labor-market statistics released by the U.S. Bureau of Labor Statistics and industrial-production statistics released by the Federal Reserve Board to a U.S. GDP nowcasting application. Nowcasting is used as a demanding downstream test because the analytical output is updated repeatedly within a quarter. It is not used to claim a new or generally superior forecasting model.

The first contribution is a statistical-information contract and its empirical implementation. The contract links a series and reference period to its released value, public availability, release package, revision stage, archived official evidence, and source-file hash. The audited Core4 ledger covers payroll employment (PAYEMS), the unemployment rate (UNRATE), average weekly hours in manufacturing (AWHMAN), and industrial production (INDPRO) over 2004–2025. It contains 8,475 event rows, 545 packages, and 1,051 revision chains. The precision contract separates 363 packages supported by an exact official release time from 182 packages for which the official archive supports only a date and for which the ledger assigns a conservative end-of-day availability. This distinction prevents date-only evidence from supporting an intraday claim. A direct reconciliation against 264 FRED-MD vintages then assesses month-end value coherence: 138,732 of 138,864 candidate cells match exactly, while 122 INDPRO differences are confined to one March 2014 archive-ledger episode. The reconciliation validates values at common month-end cutoffs; it does not turn FRED-MD into an intraday timing source.

The direct representation audit then evaluates what common month-end reconciliation cannot establish. At each of 387 sealed intramonth origins, the event-time state is compared with the preceding monthly vintage and, as an explicitly post-dated diagnostic, with the vintage labelled by the origin's own month. The preceding vintage differs from the event-time state at every origin; it omits information already public by the origin and produces a mean absolute frozen-rule output difference of 0.953 percentage points (median 0.256). The same-month post-dated representation differs at 202 origins and produces a mean absolute difference of 0.238 (median 0.013), principally because it introduces values released after the origin. Pre-COVID means remain 0.254 and 0.111, respectively, so the distinction is not only a COVID-era artifact. These are representation discrepancies under one unchanged rule, not forecast-accuracy gains.

The second contribution is an audit methodology for this contract. We define the public-release filtration and require every input and fitted parameter used at an origin to be release eligible. Rows sharing an availability time and official package are treated atomically whenever their within-package order is not verified. For a frozen analytical rule, finite differences separate the response to headline observations, historical backfills, revisions, and their interactions. A stability bound connects the output change to the rule's sensitivity, while an information-coarsening result distinguishes omission at a common origin from a post-dated snapshot that can introduce future information. The mathematical ingredients are standard. Their role is to make the data contract, audit conditions, and replay quantities explicit and machine checkable.

The third contribution is a downstream application of the audited data states. Two bridge rules are specified before release-package impacts are examined, and only the factor-plus-residual (FR) rule passes the declared qualification gate. Holding its fitted parameters fixed, we replay 387 out-of-sample packages across 62 target quarters. The mean absolute complete-package update is 0.561 percentage points, with a 95% target-quarter block-bootstrap interval of [0.232, 1.137]. The mean absolute revision-only update is 0.078 percentage points, with interval [0.058, 0.099]. The ratio of absolute revision-only updates to absolute complete-package updates is 0.139, with interval [0.072, 0.287]. This revision absolute-update ratio is a sensitivity summary, not an additive structural share. Because the qualified outer panel contains no historical first-seen backfill, a two-component headline–revision Shapley allocation is identified from the already replayed states. It assigns a mean absolute 0.046 percentage points to revisions, or 0.083 relative to the sum of absolute complete-package updates. This is an order-invariant allocation of the interaction, not a causal revision effect.

The output magnitudes are concentrated in the COVID/recovery period, and the ledger has no post-2025 holdout. The intervals are descriptive uncertainty summaries conditional on the frozen design, not estimates of time-invariant population moments. We do not infer that event-time storage mechanically improves forecast accuracy, that FR is generally forecast superior, or that the estimated magnitudes transfer unchanged to other official statistical systems. The study also does not assess punctuality: the ledger verifies actual public availability, but a punctuality assessment would additionally require a complete record of scheduled release targets.

The paper does not claim priority for multidimensional quality management, statistical metadata systems, production architecture, real-time vintage data, event calendars, forecast-news accounting, or revision decomposition. Its narrow joint contribution is operational: it connects observation-level value histories to verified release events and provenance, validates the resulting bitemporal state against a conventional archive, and replays the semantic contents of official packages through an unchanged downstream rule. This makes the question of what was known when a falsifiable and reproducible statistical audit rather than an informal data-preparation assertion.

The remainder of the paper is organized as follows. Section 2 defines the audit framework and formal guarantees. Section 3 describes ledger construction, provenance, and month-end reconciliation. Section 4 explains the qualification of the downstream measurement rule and the replay protocol. Section 5 presents the audit findings, output consequences, and scope checks. Section 6 concludes. Proofs and detailed audit protocols are in the appendices.

1.1 Relation to the Official-Statistics and Nowcasting Literature

The official-statistics literature supplies the quality concepts that motivate the audit. Product-quality systems typically distinguish accuracy and reliability from relevance, timeliness and punctuality, comparability and coherence, and accessibility and clarity (Biemer et al., 2014). The European Statistics Code of Practice and the United Nations National Quality Assurance Framework similarly treat quality as a system-wide responsibility supported by documented procedures and metadata (European Statistical System, 2017; United Nations Statistics Division, 2019). Kenett and Shmueli (2016) broaden the perspective from the intrinsic properties of a statistical product to information quality: the usefulness of data depends also on the analytical goal and method. The present paper operationalizes a selected subset of these dimensions for a specific downstream use. Release lag is a timeliness measure, revision lineage concerns reliability and transparency, and archive reconciliation concerns value coherence. These diagnostics do not constitute an overall quality ranking of the producing agencies, do not identify the true value of an economic concept, and, without complete scheduled-release records, do not measure punctuality.

A second strand concerns metadata and production architecture. Signore et al. (2015) show why metadata about data content and statistical processes should be managed jointly for both internal and external users. Enterprise architectures organize methods, information, and technology across statistical production (Struijs et al., 2013), while the Generic Statistical Business Process Model provides common terminology for production activities and cross-cutting metadata and quality management (United Nations Economic Commission for Europe, 2025). The event ledger is not proposed as an alternative enterprise architecture or a complete metadata standard. It is a narrow audit layer for the dissemination boundary: it records the released state, its availability, its package context, and the evidence needed to reproduce a later use of that state.

A third strand studies uncertainty and revision in official economic statistics. Mazzi et al. (2021) emphasize that initial official estimates are uncertain and that revision information is part of communicating that uncertainty. Real-time macroeconomic research constructs vintage databases and shows that revisions can matter for empirical conclusions and forecasts (Croushore and Stark, 2001; Aruoba, 2008; Jacobs and van Norden, 2011). The present audit does not estimate a latent final value or classify agency revisions as statistical news or noise. It instead preserves the observed revision lineage and measures the response of one fixed downstream rule to alternative recorded package states.

Finally, official-statistics and macroeconomic nowcasting share an interest in timely analytical output. The recent JOS time-series literature highlights the pressure for more timely official series and the use of advance estimates and nowcasting (Bell et al., 2024). The European Statistics Awards explicitly evaluate both nowcast accuracy and the reproducibility and production potential of submitted methods (Zając et al., 2024). In macroeconomics, mixed-frequency regressions and dynamic-factor systems formalize ragged information sets and can attribute forecast changes to incoming releases (Ghysels et al., 2007; Andreou et al., 2010; Giannone et al., 2008; Bánbura et al., 2013; Bok et al., 2018). These studies establish the value of release-aware analysis; simultaneous releases and event-level updates are therefore not novelty claims here.

Existing work also decomposes nowcast changes into new-data, data-revision, and parameter-revision effects (Hayashi and Tachi, 2021; Almuzara et al., 2023), models the release path of GDP itself (Anesti et al., 2022), or attributes changes to GDP components (Higgins, 2014; O'Keeffe and Petrova, 2025). Our estimand differs in two ways. First, semantic subsets of the same verified package form counterfactual observed-data states. Second, fitted parameters are held fixed, so no parameter-revision effect or causal announcement effect is estimated. The defensible gap is thus not any individual metadata field or decomposition. It is the joint, reproducible mapping from archived official evidence to an event-eligible bitemporal state, atomic package semantics, revision lineage, and frozen downstream replay.

2 An Audit Framework for Real-Time Official Statistics

The framework is designed as an audit layer at the dissemination boundary. It does not replace an agency production system or prescribe a forecasting model. It links public release evidence to a reproducible analytical state, and it separates three questions that a conventional latest-value table cannot answer by itself: which values were public at an origin, which rows belonged to the same official package, and how revisions within that package changed a declared downstream output. Figure 1 shows the resulting chain of evidence and validation.

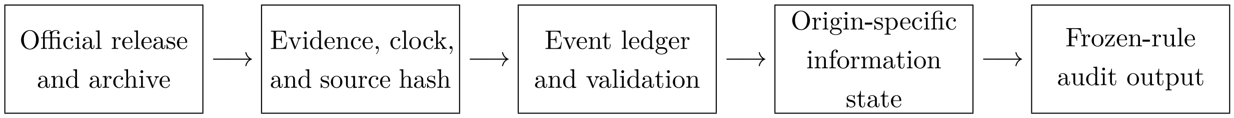

Figure 1: Release-event audit workflow. Agency releases and archived evidence identify values, public availability, and package membership. Deterministic validation then reconstructs the information state at a declared origin before any frozen downstream replay. Each arrow represents a machine-checkable mapping, not an inferred causal effect.

2.1 Public-release filtration, notation, and adaptedness

Work on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with public-release filtration $\{\mathcal{F}_t : t \in \mathcal{T}\}$. An event is

$$e_k = (i_k, r_k, v_k, a_k, x_k, p_k), \tag{1}$$

where i_k is the series, r_k its economic reference period, v_k the vintage or revision stage, a_k the verified availability time, x_k the released value, and p_k the release-package identifier. If $\mathcal{S}_t$ denotes release metadata publicly known by t, define

$$\mathcal{F}_t = \sigma(\mathcal{S}_t, \{e_k : a_k \le t\}) \,. \tag{2}$$

The event-ledger state is

$$X_t = C_t(\{e_k : a_k \le t\}) \,, \tag{3}$$

where C_t is a deterministic measurable map containing the declared transformations, lag state, ragged-edge mask, age, and coverage information. Let Θ_t denote the fitted parameters and tuning choices available at t. The real-time forecast is

$$\widehat{Y}_t = f_{\Theta_t}(X_t). \tag{4}$$

Both X_t and Θ_t must be $\mathcal{F}_t$-measurable; an exact predictor clock cannot repair target leakage or ex-post tuning.

Locked notation. The following notation is fixed for the remainder of the manuscript.

Symbol	Meaning
t	forecast origin or verified release time
a_k	verified public-availability time of event k
$\mathcal{F}_t$	public information available by t
X_t	state reconstructed from events available by t
Θ_t	release-eligible fitted parameters and tuning choices
p, $\mathcal{B}_t$	release-package identifier and atomic event batch
$K = \{H, B, R\}$	headline, backfill, and revision components
$U_S(X)$	state after replaying components $S \subseteq K$
$v_p(S)$	frozen-rule forecast change after replaying S
$D_p = v_p(K)$	total frozen-rule package update
I_p	aggregate second- and third-order interaction

Adaptedness assumptions. The framework uses five assumptions. They concern observability and accounting, not a parametric law for GDP or predictor releases.

Assumption 1 (Verified availability)**.** An event enters X_t if and only if its declared public-availability time satisfies $a_k \leq t$. A date-only package supports conservative date ordering, not exact intraday ordering.

Assumption 2 (Release-eligible fitting)**.** Θ_t is deterministic or $\mathcal{F}_t$-measurable. Every training label and tuning input used to construct Θ_t was publicly available by t.

Assumption 3 (Measurable forecasting rule)**.** The state map C_t and $(\theta, x) \mapsto f_\theta(x)$ are Borel measurable. Square integrability is imposed when conditional mean-squared error is used.

Assumption 4 (Atomic release package)**.** Rows with a common package identifier and verified availability time form an unordered batch. The complete post-package state depends on the set of rows, not on their storage order.

Assumption 5 (Component replay)**.** Headline H, backfill B, and revision R replay operators are well defined. For any selected subset of components, the final state is invariant to their application order.

Proposition 1 (Ledger-adapted forecasts)**.** *Under Assumptions 1–3, the forecast*

$$\widehat{Y}_t = f_{\Theta_t}(X_t)$$

is $\mathcal{F}_t$-measurable.

Proposition 1 is a sufficient condition for a genuinely real-time forecast. It also shows why timestamp verification alone is insufficient: parameters selected using unavailable targets would violate Assumption 2 even if every predictor timestamp were exact.

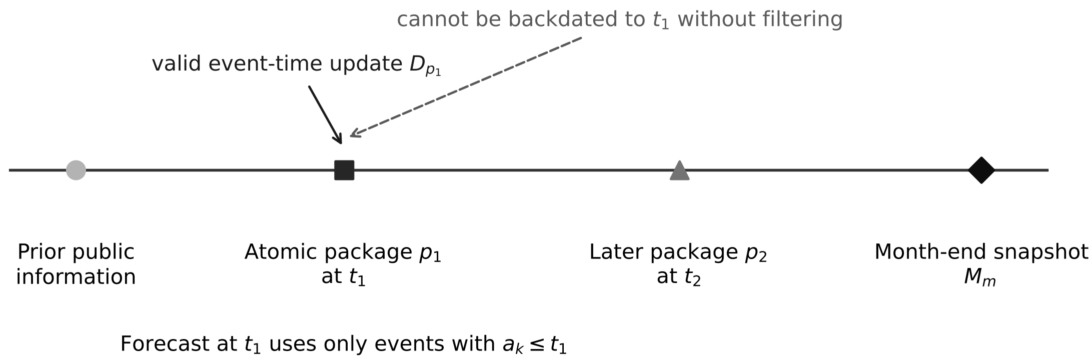

Figure 2: Event-time information and a later month-end snapshot. The diagram is conceptual. The solid arrow is a same-origin frozen-rule package update; the dashed arrow marks an invalid backdating operation unless the later snapshot is filtered back to the event-time information set.

For a snapshot M_m observed at $t_m > t$, a direct leakage diagnostic is whether the square-integrable output $Z_{t,m} = f_{\Theta_t}(M_m)$ satisfies

$$\mathbb{P}\{\mathrm{Var}(Z_{t,m} \mid \mathcal{F}_t) > 0\} > 0, \tag{5}$$

When it does, $Z_{t,m}$ is not $\mathcal{F}_t$-measurable and cannot be the output of a genuinely time-t procedure; Appendix Proposition 5 gives the formal statement and proof.

This diagnostic does not label every monthly snapshot as contaminated. A month-end file is appropriate at month-end, and an earlier-origin forecast can remain adapted if it discards every post-t component relevant to its output. The narrower point is that the month-end file alone cannot certify the earlier intramonth information set.

Figure 2 summarizes the timing distinction. An atomic package at t_1 yields a well-defined same-origin update conditional on the recorded ledger and frozen rule. A month-end snapshot observed after a later package at t_2 may be used for a forecast at t_1 only after every post-t_1 event relevant to the stored state has been removed. A monthly vintage label by itself does not perform that filtering.

2.2 Package updates and component accounting

Let $\mathcal{B}_t(p) = \{e_k : a_k = t,\ p_k = p\}$ denote one package, let X_{t^-} be the pre-package state, and let $U_A(X_{t^-})$ be the state after applying rows $A \subseteq \mathcal{B}_t(p)$. Conditional on the recorded package and the frozen rule, the package-total update is

$$D_p = f_{\Theta_t}(U_{\mathcal{B}_t(p)}(X_{t^-})) - f_{\Theta_t}(X_{t^-}). \tag{6}$$

Proposition 2 (Package invariance and row-attribution indeterminacy)**.** *Under Assumption 4, D_p*

is invariant to the storage order of package rows. For a permutation π of its q rows, define the sequential contribution

$$c_{\pi,j} = f_{\Theta_t}(U_{\{\pi(1),\ldots,\pi(j)\}}(X_{t^-})) - f_{\Theta_t}(U_{\{\pi(1),\ldots,\pi(j-1)\}}(X_{t^-})). \tag{7}$$

Then $\sum_{j=1}^{q} c_{\pi,j} = D_p$ for every π, but individual row contributions need not be invariant to π. Without verified within-package timing or an explicit allocation rule, sequential row-level attribution is not uniquely determined.

This result makes the package, rather than an arbitrary row in a stored file, the empirical event unit.

Headline, backfill, revision, and interaction. Let $K = \{H, B, R\}$. For $S \subseteq K$, let $U_S(X_{t^-})$ be the state after applying precisely the components in S, and define

$$v_p(S) = f_{\Theta_t}(U_S(X_{t^-})) - f_{\Theta_t}(X_{t^-}), \qquad v_p(\varnothing) = 0. \tag{8}$$

Proposition 3 (Exact finite-difference decomposition)**.** *Under Assumption 5, define, for every nonempty $S \subseteq K$,*

$$\mu_p(S) = \sum_{T \subseteq S} (-1)^{|S|-|T|} v_p(T). \tag{9}$$

Then

$$D_p = v_p(K) = \sum_{\varnothing \neq S \subseteq K} \mu_p(S). \tag{10}$$

In particular, the empirically reported aggregate interaction

$$I_p = v_p(K) - v_p(\{H\}) - v_p(\{B\}) - v_p(\{R\}) \tag{11}$$

is the sum of all pairwise and third-order interactions.

Thus *revision-only update* means the singleton finite difference $v_p(\{R\})$. It is not a structural revision contribution unless the higher-order terms involving R vanish.

This object is also distinct from a conventional nowcast-news decomposition. There, "news" is a released value minus its model expectation, and forecast changes may be separated into new-observation, data-revision, and parameter-revision effects (Giannone et al., 2008; Hayashi and Tachi, 2021). Here every value in $v_p(S)$ is observed in the same official package and Θ_t is held fixed. The exercise measures sensitivity to counterfactual package states; it does not recover a surprise or a parameter-revision effect.

If the feature state after replaying S is additive, $z_S = z_0 + \sum_{j \in S} \delta_j$, and the frozen rule is affine, $f(z) = \alpha + \beta' z$, then

$$D_p = \sum_{j \in K} \beta' \delta_j, \tag{12}$$

and $\mu_p(S) = 0$ for every $|S| \geq 2$; Appendix Corollary 1 records the formal result and proof.

An affine terminal regression does not alone imply zero interaction. The event-to-feature map, missingness rule, or state replacement can be nonadditive.

Stability of a frozen forecasting rule. Let $Z(X)$ be the complete feature representation passed to the forecasting rule, including any transformed lags and residual features.

If f is L-Lipschitz on the relevant feature domain, the frozen update obeys

$$|D_p| \leq L \, \| Z(U_{\mathcal{B}_t(p)}(X_{t^-})) - Z(X_{t^-}) \|. \tag{13}$$

If Z is additionally K_Z-Lipschitz under a raw-state metric ρ, then

$$|D_p| \leq L K_Z \, \rho(U_{\mathcal{B}_t(p)}(X_{t^-}), X_{t^-}). \tag{14}$$

For an affine rule $f(z) = \alpha + \beta' z$, $D_p = \beta' \Delta Z_p$ exactly and $|D_p| \leq \|\beta\|_* \|\Delta Z_p\|$. Appendix Proposition 6 gives the formal assumptions and proof.

A large forecast update can therefore reflect a large released-data change, strong rule sensitivity in that direction, a missingness-pattern change, or a combination of these mechanisms. It is not by itself a large causal news effect.

2.3 Coarsening, snapshot leakage, and the empirical audit

Proposition 4 (Oracle MSE under nested information sets)**.** *Let $Y \in L^2$ and let $\mathcal{G}_t \subseteq \mathcal{F}_t$ be information sets available at the same forecast origin. If $m_G = \mathbb{E}[Y \mid \mathcal{G}_t]$ and $m_F = \mathbb{E}[Y \mid \mathcal{F}_t]$, then*

$$\mathbb{E}[(Y - m_G)^2] - \mathbb{E}[(Y - m_F)^2] = \mathbb{E}[(m_F - m_G)^2] \geq 0. \tag{15}$$

Equality holds if and only if $m_F = m_G$ almost surely.

Proposition 4 is an oracle, same-origin result. It does not imply that an estimated finite-sample model must improve when given a richer representation. Nor does it compare a real-time forecast with a later snapshot: a post-dated information set is generally not a sub-σ-field of $\mathcal{F}_t$ and can achieve smaller retrospective error solely by containing future information.

Connection to the empirical audit. The theoretical conditions map directly to the empirical pipeline. Assumption 1 is examined by the S0 source and timestamp contract. Assumption 2 is implemented by S3 release-eligible training and target-release purging. Assumption 3 holds for the deterministic state construction and fixed ridge prediction maps. Assumption 4 is checked by the one-family, one-clock package validation in S0. Assumption 5 is enforced by S3's component and complete-package state reconstruction. The propositions therefore discipline the interpretation of the audit; they do not turn the qualified FR rule into a true conditional-mean model.

3 Construction and Validation of the U.S. Release-Event Ledger

3.1 Ledger construction, provenance, and component labels

The headline ledger is fixed before GDP outcomes or package-level forecast impacts are examined. It contains PAYEMS, UNRATE, and AWHMAN from the BLS Employment Situation release, and INDPRO from the Federal Reserve's G.17 Industrial Production and Capacity Utilization release. These series give a continuous, auditable Core4 scope over 2004–2025. Capacity utilization (CUMFNS) is not admitted to the headline scope because first-vintage coverage in the assembled ledger begins only in 2014. Personal consumption expenditures (PCE) belongs to a third release family and is retained only in the documented Core6 coverage sensitivity. Neither exclusion is selected using GDP forecast performance.

Table 1 reports the frozen scope. A series-level package count is the number of packages containing that series. These counts are not additive because a single Employment Situation package can release PAYEMS, UNRATE, and AWHMAN together.

Table 1: Official series and frozen Core4 event-ledger scope

Series	Producer and release	Reference frequency	Recorded lineage	Clock evidence	Event rows	Packages
AWHMAN	BLS, Employment Situation	Monthly	First release, backfill, revisions	Official time or date	884	262
INDPRO	Federal Reserve Board, G.17	Monthly	First release, backfill, revisions	Official time	4,112	281
PAYEMS	BLS, Employment Situation	Monthly	First release, backfill, revisions	Official time or date	2,991	263
UNRATE	BLS, Employment Situation	Monthly	First release and revisions	Official time or date	488	263
Core4 total	Two producers and release families	Monthly	1,051 reference-period chains	Mixed precision	8,475	545

Notes: The total package count is the number of unique package identifiers, not the sum of the series rows. "Backfill" denotes a historical reference period first observed after the package headline, not a separate agency policy classification. Clock evidence records the strongest archive support available; date-only packages receive the conservative end-of-day convention. Scope is fixed without reading GDP outcomes.

Two time axes and source provenance. The ledger is bitemporal in the limited but essential sense that it separates the economic reference period r_k from public availability a_k in equation (1). For example, a January employment observation can enter the information set only when the corresponding February release becomes public. Revisions preserve the same reference period but receive their own availability event. Forecast-state reconstruction uses a_k; lag and mixed-frequency transformations use r_k.

Released values are obtained from locally archived ALFRED source files. Their availability mapping is obtained independently from an official release registry constructed from BLS and Federal Reserve archive evidence. A strict event row is retained only when the value/vintage observation and the official release record join under the declared series and release-family contract. Each row stores paths and SHA-256 hashes for both the official evidence and ALFRED source. The S0 audit re-hashes the referenced files rather than accepting a manifest entry at face value. All 8,475 Core4 rows pass both evidence-file and ALFRED-file checks, and every package maps to the official registry.

The package key is `release_package_id`. Rows sharing this key must also share one release family, actual release time, availability time, time zone, and timestamp-precision classification. This restriction prevents values from different public releases from being combined merely because they have the same vintage date.

Timestamp-precision contract. The empirical clock distinguishes two cases. The 363 *official-time* packages, containing 5,678 event rows, have a verified publication time and may support event-time ordering. For 182 historical BLS packages, containing 2,797 event rows, the archive establishes the official release date but not a release time. These *official-date-only* packages are assigned 23:59:59 `America/New_York` on the verified date. The assignment is deliberately conservative: it ensures that the package cannot be used earlier on that date, but it is not evidence that the release occurred at the end of the day. Exact intraday claims therefore use the *official-time* subset as a separate robustness sample.

The availability time used by the ledger equals the verified public release time when one is observed and the conservative end-of-day time otherwise. It is not the download time, file-modification time, or ALFRED query date. All events fall inside the declared 2004–2025 window, package clocks are internally constant, and revision numbers are monotone within each series–reference-period chain.

Headline, backfill, and revision labels. The S1 census partitions every Core4 event into one of three ledger-native components. A row is a *headline first observation* when its revision number is zero and its reference period equals the package headline period. It is a *historical first-seen backfill* when its revision number is zero but its reference period predates the package headline. A row with a positive revision number is a *revision*. This rule yields 1,048 headline rows, three backfills, and 7,424 revision rows across 1,051 reference-period chains. The partition is exhaustive and mutually exclusive.

Revision magnitudes remain in official source units. Payroll revisions, for example, cannot be ranked against unemployment-rate or industrial-production revisions from their raw numerical magnitudes. The paper uses these labels to replay forecast states; it does not infer a common structural revision shock from cross-series source units. Detailed component and chain statistics appear in Appendix B.

3.2 Direct reconciliation with FRED-MD vintages

To distinguish the event ledger from a self-generated benchmark, we compare it with the archived monthly vintages of FRED-MD (McCracken and Ng, 2016). The archive contains 264 labelled vintages spanning the 2004–2025 study window. For each vintage month, we reconstruct the ledger state at 23:59:59 `America/New_York` on the final calendar day and compare raw Core4 levels by series and reference period. Thus the exercise asks whether the event ledger and the conventional archive agree at month-end. It does not infer an intramonth availability time from FRED-MD.

Every one of 138,864 candidate cells is classified as an exact match, a value mismatch, present only in one source, or missing in both. As Table 2 shows, 138,732 cells match exactly and ten

Table 2: Month-end reconciliation with archived FRED-MD vintages

Series	Candidate cells	Exact matches	Both missing	Value mismatches
AWHMAN	34,716	34,714	2	0
INDPRO	34,716	34,591	3	122
PAYEMS	34,716	34,714	2	0
UNRATE	34,716	34,713	3	0
Total	138,864	138,732	10	122

Notes: Comparisons use raw levels and the reconstructed ledger state at month-end. No cell is present in only one source. All value mismatches occur for INDPRO in the March 2014 FRED-MD vintage. FRED-MD supplies a monthly vintage label, not the event ledger's intramonth release clock.

are missing in both sources. The remaining 122 cells are all INDPRO values in the March 2014 FRED-MD vintage. Adjacent vintages agree with the ledger, so the audit labels them an isolated archive–ledger value difference; it does not attribute the episode to an agency revision, an archive error, or a ledger error. The mean ledger-minus-archive difference within the episode is -0.0464 in the raw INDPRO index and the maximum absolute difference is 0.5266.

The reconciliation sharpens the paper's contribution in an important way. The event ledger is not valuable because conventional monthly-vintage values are generally wrong; almost all comparable month-end values coincide. Its additional content is the sequence and composition of the releases that lead from one month-end state to the next.

4 Evaluation Design: Information Coarsening and Frozen-Rule Replay

4.1 Evaluation sample and downstream measurement instrument

The empirical object is the change in the output of an unchanged forecasting rule when a verified package enters its information set. The rule must use the monthly predictors, because an intercept-only rule would mechanically assign a zero update to every package. It must also clear a minimal forecasting credibility gate before its package responses are interpreted. This gate is a screen against an empirically unresponsive or grossly misspecified measuring device; it is not a model-selection tournament and does not make the surviving rule a proposed forecasting model.

The target is the advance-vintage compound annualized quarter-on-quarter growth rate of real GDP, in percentage points,

$$y_q = 100\left[\left(\frac{Y_q}{Y_{q-1}}\right)^4 - 1\right], \tag{16}$$

where both levels are taken from the target quarter's first available GDP vintage. We retain nowcast origins for target years 2004–2025. The scored outer sample begins in 2010 because the earlier observations supply initial estimation and validation histories. The inclusive 2010Q1–2025Q3 calendar contains 63 quarters, but the sealed panel contains 62. The 2018Q3 advance GDP event, dated October 26, 2018, has no verified match in the official target-release registry and is therefore excluded; a later GDP vintage cannot substitute without changing the advance-vintage estimand. All other quarters in the range contain six or seven qualified package origins.

Two transparent ridge bridge candidates were declared before package impacts were examined. This limited qualification exercise protects the audit from using an unresponsive or grossly inadequate downstream measuring instrument; it is not a search for a new forecasting model. Neither package updates nor their semantic components enter the candidates, tuning, or eligibility decision. Every feature transformation, coefficient, penalty choice, and GDP training label is subject to release eligibility and target-release purging.

The factor candidate, F, summarizes the Core4 snapshot with a train-only factor; the factor-plus-residual candidate, FR, retains source-specific residuals. They are screened against recursive Student-t location and AR-$t(1)$ controls under eight non-overlapping outer blocks, with each target quarter receiving total weight one. The pre-declared gate requires finite outputs, an overall RMSE ratio no larger than 1.05 relative to the better control, and the same threshold in at least one declared subperiod. F fails. FR's ratios are 0.928 overall, 1.044 pre-COVID, and 1.131 after 2021, so it qualifies only as the frozen measurement rule used below. Appendix C, including Table 10, reports the feature maps, folds, penalty choices, controls, and complete decision.

Frozen package replay. For each qualified outer origin, we fit FR once on the release-eligible history and construct five same-origin states: the state immediately before the package, three states that add only headline, backfill, or revision rows, and the complete post-package state. The fitted standardization, factor loading, residual map, ridge coefficients, and target lags are identical across these states. Their forecast differences therefore estimate $v_p(\{H\})$, $v_p(\{B\})$, $v_p(\{R\})$, and D_p as defined in Section 2. They are counterfactual outputs of a frozen rule, not causal effects of releases on GDP.

The complete post-package state is required to reproduce both the strict data adapter and the corresponding qualification forecast. The raw transformed matrix tolerance is 10^{-5} and the forecast tolerance is 10^{-8}; the largest observed reproduction discrepancy is 1.78×10^{-15}. We also refit the identical FR rule immediately after a predictor package on exactly the same historical snapshot rows. Because no new GDP label arrives and past event-time rows are not rewritten, this deterministic counterfactual is zero up to numerical precision. It is an implementation check, not observed institutional refitting behavior.

Monthly-representation counterfactuals. At each sealed origin, we also replace the event-time Core4 state with each of two deterministic monthly-vintage representations. The *previous-month-available* representation uses the FRED-MD file labelled by the preceding calendar month. It avoids assigning a later-labelled file to the origin but may omit releases that occur after its cutoff and before the origin. The *same-month-post-dated* representation assigns the file labelled by the origin's own month back to that intramonth origin. It is intentionally diagnostic because it may contain values released later in the month.

For both comparisons, the fold, selected ridge penalty, release-eligible estimation history, target lags, and FR fitting algorithm are unchanged. We compare 56 raw cells (four series over 14 monthly reference periods), 48 transformed cells (four series over 12 feature periods), and the resulting FR output. No candidate is added or retuned, and realized GDP errors are not compared across the representations. The estimand is a deterministic representation discrepancy, not an accuracy gain.

4.2 Descriptive uncertainty

Uncertainty intervals condition on the qualified, frozen FR rule. We resample target-quarter clusters using a circular moving-block bootstrap with blocks of four consecutive quarters, 50,000 replications, and seed 2499. Each draw retains all package observations attached to a sampled quarter. Reported 95% intervals are percentile intervals. They quantify sampling variation in the descriptive package-accounting moments under this resampling design; they neither account for model-selection uncertainty nor test a causal GDP-news effect. With 62 target-quarter clusters, they should not be read as tests of release-family or period differences, or as confidence intervals for a time-invariant process.

5 Audit Findings and Downstream Consequences

5.1 Information-representation audit

Table 3 gives the direct comparison. The previous-month representation differs from the event-time raw and transformed states at all 387 origins. It has 10.297 different raw cells per origin on average and changes the frozen FR output at every origin; the mean absolute output discrepancy is 0.953 percentage points, with median 0.256 and 90th percentile 0.975. The maximum is 49.487 at the June 16, 2020 G.17 origin. This extreme is a COVID-shock observation: the corresponding pre-COVID mean, median, and 90th percentile are 0.254, 0.195, and 0.546.

The same-month post-dated representation differs at 202 origins. Its overall mean absolute output discrepancy is 0.238 percentage points, with median 0.013 and 90th percentile 0.562. The zero or near-zero median reflects the 185 origins whose states are identical or whose output differs only at numerical precision; it does not validate post-dating at the other origins. The pre-COVID mean remains 0.111. Thus the empirical footprint is not confined to the COVID/recovery window, although the largest output discrepancies are concentrated there.

The cell-level classification identifies the mechanism. Of 3,985 raw differences under the previous-month assignment, 3,975 are values already public by the origin but released after the snapshot cutoff; ten are archive–ledger value differences. Of 1,402 same-month differences, 1,376 are values released after the origin, 25 are archive–ledger value differences, and one is present only in the archive. The comparison therefore measures delayed inclusion in the first case and predominantly premature inclusion in the second, rather than a broad failure of month-end value coherence.

Figure 3 reports the full output-discrepancy distribution. Panel (a) shows that the previous-month assignment produces a broader right tail, while the post-dated assignment has a large

Table 3: Event-time versus monthly-vintage representation audit

Monthly assignment	Period	Origins	State differs	Mean raw cells	$\overline{\lvert\Delta f\rvert}$	Median $\lvert\Delta f\rvert$	P90 $\lvert\Delta f\rvert$
Previous month	All	387	387	10.297	0.953	0.256	0.975
Previous month	Pre-COVID	243	243	10.342	0.254	0.195	0.546
Previous month	COVID/recovery	50	50	10.140	5.324	0.936	8.775
Previous month	Post-2021	94	94	10.266	0.433	0.308	1.049
Same month, post-dated	All	387	202	3.623	0.238	0.013	0.562
Same month, post-dated	Pre-COVID	243	127	3.683	0.111	0.006	0.378
Same month, post-dated	COVID/recovery	50	26	3.420	0.786	0.063	2.524
Same month, post-dated	Post-2021	94	49	3.574	0.277	0.025	0.900

Notes: Each origin compares 56 raw cells and 48 transformed cells. State differs counts origins with at least one raw-cell difference; the transformed state differs at the same number of origins in every row. Δf is the monthly-representation output minus the event-time output under the same FR algorithm, fold-specific penalty, and release-eligible estimation history. Output differences are annualized GDP-growth percentage points. No accuracy comparison, model search, retuning, or J2 bootstrap is performed.

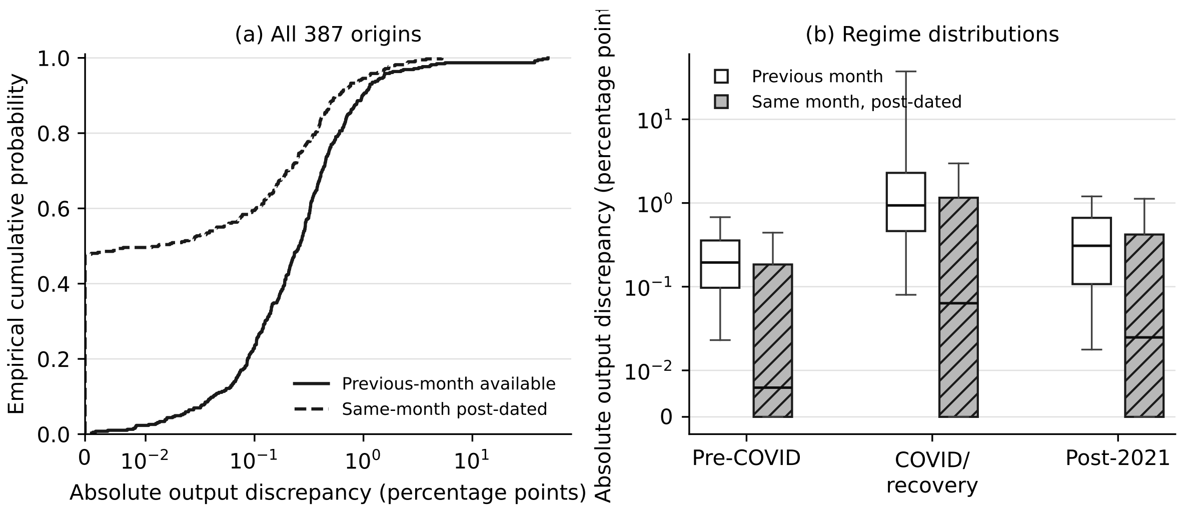

Figure 3: Distribution of absolute frozen-output discrepancies between the event-time and monthly-vintage representations. Panel (a) uses a symmetric-log horizontal scale. In panel (b), boxes show interquartile ranges, center lines are medians, and whiskers are the 5th and 95th percentiles; the vertical scale is symmetric-log. White boxes denote the previous-month assignment and hatched gray boxes the same-month post-dated assignment.

mass at or near zero. Panel (b) separates the three declared periods. Dispersion rises sharply during COVID/recovery for both assignments, but the boxes remain above a purely zero numerical discrepancy before and after the shock window.

Timing evidence is reported separately from output sensitivity. The 387-origin panel contains 256 packages supported by an official time and 131 supported by an official date only. Among packages with an observed headline lag, the median lag from reference-month end to release is five days for the Employment Situation and 16 days for G.17; the corresponding 90th percentiles are eight and 17 days. The previous-month mean absolute output discrepancy is 1.183 for exact-time origins and 0.502 for date-only origins; same-month figures are 0.147 and 0.416. Because precision groups differ in family and temporal composition, these partitions are not estimates of a timestamp-precision

Table 4: Frozen-rule package-update inference by release family

Release family	Packages	$\overline{\lvert D_p \rvert}$	$\overline{\lvert v_p(\{R\}) \rvert}$	Q_R
All packages	387	0.561 [0.232, 1.137]	0.078 [0.058, 0.099]	0.139 [0.072, 0.287]
Employment Situation	188	0.681 [0.178, 1.639]	0.039 [0.031, 0.049]	0.057 [0.028, 0.198]
Federal Reserve G.17	199	0.448 [0.274, 0.670]	0.114 [0.077, 0.156]	0.255 [0.189, 0.349]

Notes: Brackets contain 95% target-quarter circular moving-block bootstrap percentile intervals. Updates are in annualized GDP-growth percentage points; Q_R is unitless. In the stored output the backward-compatible column name contains 'share', but the reported estimand is the revision absolute-update ratio in equation (28).

effect.

5.2 Package, component, and measurement accounting

The qualified outer panel contains 387 package updates spanning 62 target quarters: 188 Employment Situation packages and 199 G.17 packages. Across all packages, the mean absolute nowcast update is 0.561 percentage points and its 95% block-bootstrap interval is [0.232, 1.137]. The signed mean is only 0.011 percentage points, so the absolute moment describes movement rather than a systematic upward or downward shift.

Table 4 separates the two release families. The mean absolute update is 0.681 percentage points for Employment Situation packages and 0.448 for G.17 packages. The Employment Situation interval is much wider, which foreshadows the period concentration documented below. These magnitudes describe the response of the frozen FR rule; they are not estimates of a release's effect on realized GDP.

Headline, revision, and interaction terms. The all-package mean absolute headline-only finite difference is 0.580 percentage points. The corresponding revision-only value is 0.078, with a 95% interval of [0.058, 0.099]. No historical first-seen backfill occurs among the 387 qualified outer packages, so the empirical backfill-only update is zero in this panel; this is a sample fact, not a theoretical restriction on $v_p(\{B\})$.

The mean absolute aggregate interaction is 0.155 percentage points. This term is not expected to vanish merely because the terminal regression is ridge-linear: raw vintage replacements pass through log differences, first differences, the ragged-edge state, and train-only factor-residual features. Consistent with Proposition 3, the absolute headline, revision, and interaction moments do not add to the absolute total.

Because no backfill occurs in the qualified outer panel, the implemented aggregate interaction is the nonadditivity between headline and revision states. The empty, headline-only, revision-only, and complete states consequently also identify the two-component Shapley allocation

$$\phi_{p,R} = \frac{1}{2} v_p(\{R\}) + \frac{1}{2}\{D_p - v_p(\{H\})\}, \qquad \phi_{p,H} = D_p - \phi_{p,R}. \tag{17}$$

The maximum efficiency residual is 4.44×10^{-16}. The mean absolute revision allocation is 0.046

percentage points overall, 0.032 for Employment Situation packages, and 0.060 for G.17 packages. Its median absolute value is 0.021 overall. Relative to the sum of absolute complete-package updates, $\sum_p |\phi_{p,R}|/\sum_p |D_p|$ is 0.083 overall, compared with 0.047 and 0.134 by family. This is an order-invariant efficient allocation of the observed headline–revision interaction, not a causal attribution. The three-component formula remains available for an extension with observed backfills, but the current empirical result contains no backfill allocation.

The revision absolute-update ratio is $Q_R = \sum_p |v_p(\{R\})|/\sum_p |D_p|$. It equals 0.139 overall, with a 95% interval of [0.072, 0.287]. The ratio is higher for G.17 (0.255) than for the Employment Situation (0.057), but it is not a structural decomposition: signs and component interactions can cause the numerator to exceed the denominator. Lemma 1 formalizes this limitation. The lower Shapley ratio of 0.083 shows why the singleton and allocated quantities should be kept distinct: the latter shares the headline–revision interaction according to the explicit convention in equation (17).

Implication for real-time macroeconomic forecasting. The application establishes a measurement implication rather than a new forecasting-performance result. Near-complete agreement between the ledger and month-end FRED-MD values does not certify that a monthly snapshot represents every earlier intramonth forecast origin. Direct comparison shows that using the preceding labelled vintage delays already-public information at every origin, whereas assigning the same-month vintage prematurely includes later information at 202 origins. The respective mean absolute frozen-output discrepancies are 0.953 and 0.238 percentage points. These differences, together with the package sequence and composition, are what the release-event audit preserves.

The results also caution against treating the released headline as the only input change in a package. The revision-only state moves the frozen rule by 0.078 percentage points on average in absolute value, while the order-invariant Shapley allocation is 0.046 because headline and revision states are not additive. These figures are not a prescription for changing a production nowcast, nor a statement about the economic importance of a release; they show which aspects of the recorded real-time data state a specified GDP nowcasting rule is sensitive to.

5.3 Temporal concentration, robustness, and scope

Table 5 partitions the sealed output without reselecting or refitting a period-specific model. Mean absolute package updates rise from 0.202 percentage points before COVID to 2.676 during 2020–2021, then fall to 0.364 after 2021. The COVID/recovery interval contains only eight target quarters but dominates the overall magnitude. Revision-only changes also rise across these periods, whereas their ratio to total absolute movement falls to 0.056 in 2020–2021 because complete package updates increase much more sharply.

The partition shows why the overall qualification and update means cannot be read as temporally invariant performance facts. It also explains why FR can clear the overall gate while remaining weaker than the better control after 2021. The defensible conclusion is descriptive heterogeneity in a frozen-rule audit, not a separately estimated regime effect.

Table 5: Descriptive period partition of package updates

Period	Years	Packages	Target quarters	$\overline{\lvert D_p \rvert}$	$\overline{\lvert v_p(\{R\}) \rvert}$	Q_R
Pre-COVID	2010–2019	243	39	0.202	0.048	0.236
COVID/recovery	2020–2021	50	8	2.676	0.151	0.056
Post-2021	2022–2025	94	15	0.364	0.117	0.320

Notes: Entries are deterministic partitions of the 387-package frozen-rule panel. Although the declared pre-COVID partition starts in 2004, the scored outer panel begins in 2010. The post-2021 panel ends at target quarter 2025Q3. No model is selected within a period and no separate period-wise inference is conducted. Updates are in percentage points except for Q_R.

Timestamp-precision sensitivity. The headline analysis uses every validated package under the timestamp contract in Section 3. As a stricter descriptive check, we remove every package whose archive verifies only the release date. The official-time-only sample retains 256 of 387 outer packages. Its mean absolute complete update is 0.738 percentage points, the mean absolute revision-only update is 0.100, and the revision absolute-update ratio is 0.135, compared with 0.561, 0.078, and 0.139 in the full sample. Appendix Table 11 reports the family-level results.

This comparison is not a controlled estimate of the effect of timestamp precision. Every G.17 package has an official time, so its 199-package result is unchanged. By contrast, only 57 of 188 Employment Situation packages survive, and they cover only 19 rather than 62 target quarters. The BLS exact-time-only mean absolute update is 1.748 percentage points largely because the restricted panel has a different temporal composition. Accordingly, the exercise confirms which estimates can support exact intraday ordering; it does not show that exact timestamps cause larger forecast updates.

Core6 coverage without model expansion. The Core4 ledger is the only scope for which a forecasting rule passed the frozen qualification protocol. A ledger-only sensitivity appends capacity utilization (CUMFNS) and personal consumption expenditures (PCE). This Core6 archive contains 13,808 event rows and 806 distinct packages, compared with 8,475 rows and 545 packages in Core4. CUMFNS contributes 2,779 rows but begins only in June 2014; PCE contributes 2,554 rows and begins in March 2004. Detailed coverage appears in Appendix Table 14.

No Core6 forecasting rule is fitted, tuned, or scored. The coverage exercise therefore shows that the bitemporal infrastructure can hold the additional series, but it provides no evidence that Core6 improves GDP nowcasts. In particular, the incomplete early CUMFNS history prevents it from silently replacing the continuous 2004–2025 Core4 headline design.

6 Discussion and Conclusion

Real-time nowcasting requires an auditable answer to a basic question: what was publicly known at the stated forecast origin? This paper represents releases as a bitemporal event ledger, defines forecasts on the resulting public-release filtration, and treats a common-clock release package as the audited event unit. The framework distinguishes legitimate same-origin information coarsening from post-dated snapshot substitution and gives an exact finite-difference accounting of headline, backfill, revision, and interaction updates.

For U.S. GDP nowcasting, the Core4 ledger records 8,475 event rows and 545 packages over 2004–2025. Its values reconcile almost completely with 264 FRED-MD month-end vintages, while retaining the intramonth order and package composition that the monthly labels do not supply. Direct comparison at 387 origins shows that the previous-month representation delays public information at every origin and yields a mean absolute frozen-output discrepancy of 0.953 percentage points. Assigning the same-month vintage back to the origin introduces future information at 202 origins and yields a mean absolute discrepancy of 0.238. The pre-COVID counterparts, 0.254 and 0.111, confirm that the result is not only a shock-period artifact.

Conditioning on the credibility-qualified FR measuring rule, the 387 outer-sample packages produce a mean absolute nowcast update of 0.561 percentage points. The mean absolute revision-only update is 0.078 and its absolute-update ratio is 0.139; after allocating the headline–revision interaction by the two-component Shapley rule, the corresponding figures are 0.046 and 0.083. The much larger 2020–2021 updates show that these sample averages are not temporally homogeneous.

For producers and disseminators, the operational implication is not that monthly vintage archives should be replaced. They remain the independent benchmark for month-end value coherence. The implication is that uses requiring intramonth reconstruction need a complementary dissemination-layer contract: reference period, released value, public-availability evidence and precision, package identity, revision lineage, and immutable provenance. Preserving these fields at release time allows later users to validate an origin-specific state without inferring a release clock from a monthly label. The same contract can audit analytical outputs other than GDP nowcasts, provided that their inputs and fitted objects are subject to the same eligibility rule.

The evidence has important limits. Core4 covers four U.S. series from two release families and does not establish portability to other statistical systems. The 131 date-only outer packages support conservative daily ordering, not exact intraday attribution. Scheduled-release histories are incomplete, so the paper measures actual release lag rather than punctuality. The downstream magnitudes are conditional on one qualified FR rule and are neither causal announcement effects nor evidence that FR is a generally superior GDP model. Bootstrap intervals condition on the selected rule, the 62-quarter panel, and the declared four-quarter block design; they do not test differences across agencies or regimes. COVID/recovery observations are unusually influential, although the representation discrepancy remains present before COVID.

The validated ledger ends on December 31, 2025; the last qualified origin is September 16, 2025 and targets 2025Q3. A prospective extension should append new official evidence and GDP vintages under the same contract, freeze the current qualification decision before observing new package impacts, and publish the validation failures as well as the successful replays. Such an extension would test the audit protocol outside the reconstructed window without silently retuning the measurement instrument.

In summary, the empirical value of the release-event ledger is not that monthly archive values are generally wrong. It is that value history, availability, package semantics, and provenance remain jointly auditable between monthly cutoffs. The U.S. application shows that this added information is observable, machine-checkable, and consequential for a declared downstream output. That is the paper's official-statistics contribution; forecasting is the stress test, not the proposed innovation.

A Proofs

Proposition 1.

Proof. Every event used by X_t belongs to $\mathcal{F}_t$ by Assumption 1. Because C_t is deterministic and measurable, X_t is $\mathcal{F}_t$-measurable. By Assumption 2, Θ_t is also $\mathcal{F}_t$-measurable. Hence the pair (Θ_t, X_t) is $\mathcal{F}_t$-measurable. Assumption 3 and closure of measurability under composition imply that $f_{\Theta_t}(X_t)$ is $\mathcal{F}_t$-measurable. □

Proposition 5 (Post-dated snapshot leakage diagnostic)**.** *Let M_m be a snapshot observed at $t_m > t$, and suppose $Z_{t,m} = f_{\Theta_t}(M_m)$ is square integrable. If condition* (5) *holds, then $Z_{t,m}$ is not $\mathcal{F}_t$-measurable and cannot be the output of a genuinely time-t forecasting procedure.*

Proof. If a square-integrable random variable Z is $\mathcal{F}_t$-measurable, then $\mathbb{E}[Z \mid \mathcal{F}_t] = Z$ almost surely. Consequently,

$$\mathrm{Var}(Z \mid \mathcal{F}_t) = \mathbb{E}[(Z - \mathbb{E}[Z \mid \mathcal{F}_t])^2 \mid \mathcal{F}_t] = 0 \tag{18}$$

almost surely. Applying this necessary property to $Z_{t,m}$ contradicts condition (5). □

Proposition 2.

Proof. Assumption 4 makes the complete post-package state $U_{\mathcal{B}_t(p)}(X_{t^-})$ a function of the set of package rows, so D_p is storage-order invariant. For every permutation, summing (7) telescopes:

$$\sum_{j=1}^{q} c_{\pi,j} = f_{\Theta_t}(U_{\mathcal{B}_t(p)}(X_{t^-})) - f_{\Theta_t}(X_{t^-}) = D_p. \tag{19}$$

Individual terms need not be order invariant. Let $f(z_1, z_2) = z_1 z_2$, let the pre-package state be $(0, 0)$, and let two named rows change the state by $\delta_1 = (1, 0)$ and $\delta_2 = (0, 1)$. Applying row 1 then row 2 assigns contributions $(0, 1)$ to the named rows. Reversing the order assigns $(1, 0)$ to those same named rows. The package total is one in both orders, but the row attribution changes. Thus row-level sequential attribution requires verified order or an additional allocation convention. □

Proposition 3.

Proof. Sum (9) over all $S \subseteq K$ and interchange the finite sums:

$$\sum_{S\subseteq K} \mu_p(S) = \sum_{T\subseteq K} v_p(T) \sum_{S:T\subseteq S\subseteq K} (-1)^{|S|-|T|} \tag{20}$$

$$= \sum_{T\subseteq K} v_p(T)(1-1)^{|K|-|T|} \tag{21}$$

$$= v_p(K). \tag{22}$$

The inner sum is zero unless $T = K$, in which case it is one. Because $\mu_p(\varnothing) = v_p(\varnothing) = 0$ and $\mu_p(\{j\}) = v_p(\{j\})$, separating singleton from higher-order terms gives (11) and

$$I_p = \mu_p(\{H, B\}) + \mu_p(\{H, R\}) + \mu_p(\{B, R\}) + \mu_p(\{H, B, R\}). \tag{23}$$

$\square$

Corollary 1 (Affine and additive case)**.** *Suppose $z_S = z_0 + \sum_{j\in S} \delta_j$ and $f(z) = \alpha + \beta' z$. Then $D_p = \sum_{j\in K} \beta'\delta_j$ and $\mu_p(S) = 0$ for every $|S| \geq 2$.*

Proof. For every $S \subseteq K$,

$$v_p(S) = \beta' \sum_{j\in S} \delta_j. \tag{24}$$

Substitution in (9) leaves $\beta'\delta_j$ for a singleton and cancels every feature increment in each finite difference of order two or higher. Summing the singleton terms yields the result. $\square$

Proposition 6 (Forecast-update stability)**.** *If f is L-Lipschitz on the relevant feature domain, then equation* (13) *holds. If Z is additionally K_Z-Lipschitz under a raw-state metric ρ, then equation* (14) *holds. For an affine rule $f(z) = \alpha + \beta' z$, $D_p = \beta'\Delta Z_p$ and $|D_p| \leq \|\beta\|_* \|\Delta Z_p\|$.*

Proof. Apply Lipschitz continuity of f to the pre- and post-package feature vectors:

$$|D_p| \leq L\|Z(U_{\mathcal{B}_t(p)}(X_{t^-})) - Z(X_{t^-})\|. \tag{25}$$

Applying the K_Z-Lipschitz inequality for Z to the right-hand side gives (14). For $f(z) = \alpha + \beta' z$, direct subtraction gives $D_p = \beta'\Delta Z_p$, and Hölder's inequality gives $|D_p| \leq \|\beta\|_* \|\Delta Z_p\|$. $\square$

If a feature map jumps when a missingness indicator changes, the second bound requires a raw-state metric that includes this discrete indicator. It should not be asserted under a continuous-value norm alone.

Proposition 4.

Proof. Write

$$Y - m_G = (Y - m_F) + (m_F - m_G). \tag{26}$$

Because $\mathcal{G}_t \subseteq \mathcal{F}_t$, $m_F - m_G$ is $\mathcal{F}_t$-measurable. Moreover, $\mathbb{E}[Y - m_F \mid \mathcal{F}_t] = 0$, so the cross term satisfies

$$\mathbb{E}[(Y - m_F)(m_F - m_G)] = 0. \tag{27}$$

Expanding the square gives (15). Nonnegativity follows from the squared right-hand side, and equality holds exactly when $m_F = m_G$ almost surely. □

Remark 1 (Post-dated leakage counterexample). Let Y have mean zero and positive variance, let $\mathcal{F}_t$ be the trivial σ-field, and suppose a later release reveals Y, so that $\mathcal{H}_{t_m} = \sigma(Y)$ for $t_m > t$. The real-time oracle is zero and has MSE $\mathrm{Var}(Y) > 0$. The post-dated forecast $\mathbb{E}[Y \mid \mathcal{H}_{t_m}] = Y$ has zero retrospective MSE. The apparent gain is entirely due to information unavailable at t and is not a real-time forecast improvement.

Revision absolute-update ratio. For the empirical packages, define

$$Q_R = \frac{\sum_p |v_p(\{R\})|}{\sum_p |v_p(K)|}. \tag{28}$$

Lemma 1 (The ratio is not an additive share)**.** *Without additivity and sign restrictions, Q_R need not lie in* $[0, 1]$ *and is not a structural decomposition share.*

Proof. For one package, let the revision-only update be 2 and let the other component and interaction terms sum to -1.5. The total update is 0.5, so the absolute ratio is $2/0.5 = 4 > 1$. If the terms cancel exactly, the denominator is zero and the ratio is undefined. The observed value 0.139 is therefore a relative magnitude, not an additive variance or structural share. □

Order-invariant component allocation. If component attributions that sum exactly to D_p are later required, define the Shapley allocation (Shapley, 1953)

$$\psi_{p,j} = \sum_{S \subseteq K \setminus \{j\}} \frac{|S|!(|K| - |S| - 1)!}{|K|!} \left[v_p(S \cup \{j\}) - v_p(S) \right]. \tag{29}$$

Proposition 7 (Order-invariant efficient allocation)**.** $\sum_{j \in K} \psi_{p,j} = D_p$. *This is the unique allocation satisfying efficiency, symmetry, the dummy property, and additivity.*

Proof. Equation (29) averages component j's marginal contribution over all $|K|!$ component orderings. Within each ordering, the marginal contributions of all components telescope to $v_p(K) - v_p(\varnothing) = D_p$; averaging proves efficiency. Symmetry, the dummy property, and additivity follow directly from the marginal-contribution formula.

For uniqueness, the unanimity games $u_T(S) = \mathbf{1}\{T \subseteq S\}$, for nonempty $T \subseteq K$, form a basis for all set functions satisfying $v_p(\varnothing) = 0$. In u_T, the dummy property assigns zero to components outside T, while symmetry and efficiency assign $1/|T|$ to each component in T. Additivity therefore fixes the allocation on every basis element and hence on every admissible set function. The allocation in (29) has these values and is unique. □

The current outer panel contains no backfill component. Its empty, headline-only, revision-only, and full states therefore provide every coalition needed for the two-component restriction of (29). The empirical allocation is reported in equation (17). A future three-component application with observed backfills would require the additional pairwise states before applying the full formula.

B Provenance and Timestamp Audit

Strict-row inclusion and audit conditions. The Core4 audit requires all of the following conditions before any GDP outcome is loaded:

1. the set of headline series equals PAYEMS, UNRATE, AWHMAN, and INDPRO;
2. every event identifier is unique and every package identifier is present;
3. every package joins the official release registry;
4. a package has one release family, actual time, availability time, and timestamp-precision class;
5. verification status and precision labels belong to the pre-declared accepted sets;
6. PAYEMS, UNRATE, and AWHMAN map only to the Employment Situation family and INDPRO maps only to G.17;
7. every official evidence path and ALFRED source path exists and matches its stored SHA-256 hash;
8. every event clock follows the declared exact-time or conservative date-only rule;
9. all events lie inside the study window; and
10. revision numbers are monotone within series–reference-period chains.

All ten conditions pass. The audit records no evidence-file or ALFRED-file hash error. Because the source paths point to a local archive, the check verifies the bytes used by the study; it does not assert that an external website will remain unchanged or continuously accessible.

Core event fields. Table 6 groups the minimum fields needed to reconstruct a released value, place it in an origin-specific information set, identify its atomic package, and verify its provenance. The public disclosure version keeps portable artifact identifiers and hashes while removing workstation paths.

Table 6: Core fields in the strict event ledger

Field group	Interpretation
‘series_id’, ‘reference_period’, ‘value’	Released series, economic reference period, and published value.
‘revision_number’, ‘previous_value’, ‘revision_delta’	Position and change within a series–reference-period vintage chain.
‘release_family’, ‘release_package_id’	Official product family and atomic release bundle.
‘headline_reference_period’	Reference period designated as the package headline for the component classification.
‘actual_release_at’, ‘available_at’, ‘timezone’	Verified release clock, forecast-availability clock, and `America/New_York` time-zone contract.
‘verification_status’, ‘verified_by’	Evidence status and construction rule.
‘evidence_file’, ‘evidence_sha256’	Archived official release evidence and its byte-level hash.
‘alfred_source_file’, ‘alfred_source_sha256’	Archived released-value source and its byte-level hash.

Notes: Backticks are used here only to reproduce machine-readable field names. The complete data dictionary accompanies the replication package.

Component census. The deterministic component rule is exhaustive in the frozen Core4 ledger. Table 7 reports how its 8,475 rows divide into headline first observations, historical first-seen backfills, and revisions by series.

Table 7: Ledger-native component counts by series

Series	Headline first	Historical backfill	Revision	Total events
AWHMAN	262	1	621	884
INDPRO	262	1	3,849	4,112
PAYEMS	262	1	2,728	2,991
UNRATE	262	0	226	488
Total	1,048	3	7,424	8,475

Notes: Headline and backfill rows have revision number zero. Their classification depends on whether the row’s reference period equals the package headline reference period. Every positive-revision-number row is a revision.

Revision-chain depth and source units. Revision depth differs substantially by producer and series. Because the released values retain their official source units, Table 8 reports within-series diagnostics and does not impose a cross-series ranking of raw revision magnitudes.

Table 8: Revision-chain diagnostics

Series	Revision rows	Revised periods	Median depth	Maximum depth	Median $\lvert\Delta\rvert$
AWHMAN	621	223	2	8	0.1000
INDPRO	3,849	261	15	24	0.1073
PAYEMS	2,728	261	10	19	19.0000
UNRATE	226	147	1	4	0.1000

Notes: Depth counts positive revision stages in the ledger chain. $|\Delta|$ is in each source's raw unit: hours for AWHMAN, an index for INDPRO, thousands of employees for PAYEMS, and percentage points for UNRATE. Magnitudes must not be compared across rows without a declared normalization.

Reconciliation classification. For vintage m, series i, and reference period r, let $L_{i,r,m}$ be the raw level in the reconstructed month-end ledger and $F_{i,r,m}$ the corresponding FRED-MD entry. The cell classifier is

$$\mathcal{C}_{i,r,m} = \begin{cases} \text{both missing}, & L_{i,r,m} \text{ and } F_{i,r,m} \text{ are missing}, \\ \text{ledger only}, & L_{i,r,m} \text{ is observed and } F_{i,r,m} \text{ is missing}, \\ \text{FRED-MD only}, & L_{i,r,m} \text{ is missing and } F_{i,r,m} \text{ is observed}, \\ \text{exact match}, & L_{i,r,m} = F_{i,r,m}, \\ \text{value mismatch}, & \text{otherwise}. \end{cases} \tag{30}$$

Equality is evaluated on the stored numerical levels after parsing, with no GDP target and no forecast model. The 122 value mismatches share one series and one vintage and agree again in adjacent archive vintages. The label 'isolated archive–ledger difference' is descriptive and intentionally leaves the source mechanism unresolved.

C Forecast-Rule Qualification

Candidate feature maps. At origin t, each Core4 series is represented on a 12-month grid anchored to the target-quarter end. PAYEMS and INDPRO use log differences, UNRATE uses first differences, and AWHMAN enters in levels. Future or unreleased grid cells remain missing. For each series, let $\ell_{t,i}$ be its latest observed transformed cell on this grid. Means and standard deviations used below are estimated only on the training indices $\mathcal{I}_t$; missing standardized entries are set to zero and accompanied by indicators.

Stack the standardized latest observations as $\widetilde{\ell}_t$. The first right singular vector a_t, estimated on $\mathcal{I}_t$, is signed so that its INDPRO loading is nonnegative. Define the snapshot factor and source residuals as

$$g_t = \widetilde{\ell}_t' a_t, \qquad r_t = \widetilde{\ell}_t - g_t a_t. \tag{31}$$

Let h_t contain the latest four GDP growth releases strictly available at the origin and let m_t^h and m_t^ℓ denote their missingness indicators and the Core4 latest-value indicators. The two candidate designs are

$$z_t^F = (g_t, h_t, m_t^h), \tag{32}$$
$$z_t^{FR} = (g_t, h_t, m_t^h, r_t, m_t^\ell). \tag{33}$$

Each forecast is an intercept-plus-ridge regression in its declared feature vector. The model contains no release-age or coverage interaction; those were not part of the S3 candidates. Quarter weights sum to one within each target quarter in both ridge fitting and scoring. For target-quarter set $\mathcal{Q}$ and n_q origins attached to quarter q, the qualification loss is

$$\text{RMSE} = \left\{ \frac{1}{|\mathcal{Q}|} \sum_{q\in\mathcal{Q}} \frac{1}{n_q} \sum_{t:q(t)=q} (y_q - \widehat{y}_t)^2 \right\}^{1/2}. \tag{34}$$

Fold construction and tuning record. For an outer block beginning in year b, origins in $[b-2, b)$ form the validation set and earlier rows form the inner estimation set, subject in both cases to the target-release purge. The selected penalty minimizes equation (34), with ties broken in favor of the smaller penalty. Table 9 reports the complete frozen selection record. FR selects 100 in every block; F selects 10 for 2020–2021 and 0.1 for 2024–2025, and 100 otherwise.

Table 9: Nested-block qualification and selected ridge penalties

Outer years	Validation years	Inner rows	Validation rows	Outer rows	α_F	α_{FR}
2010–2011	2008–2009	87	38	50	100	100
2012–2013	2010–2011	131	44	50	100	100
2014–2015	2012–2013	181	44	49	100	100
2016–2017	2014–2015	231	43	50	100	100
2018–2019	2016–2017	280	44	44	100	100
2020–2021	2018–2019	330	38	50	10	100
2022–2023	2020–2021	374	44	50	100	100
2024–2025	2022–2023	424	44	44	0.1	100

Notes: Candidate grid is $\{0.1, 1, 10, 100\}$. Inner and validation counts are package-origin rows, not independent GDP quarters. At every outer origin, features and coefficients are refitted on the expanding release-eligible history while the block-selected penalty is fixed.

Controls, gate, and replay checks. The Student-t control estimates a location and scale with degrees of freedom fixed at five from the unique GDP releases available at an origin. The AR-t control recursively projects the same release history using one lag; if a valid autoregressive fit is unavailable in a short early history, it uses the Student-t location. No such replacement changes the common outer scoring panel.

The better-control RMSEs are 6.203 overall, 1.033 pre-COVID, and 1.765 in the normalization range. F's corresponding ratios are 1.108, 1.054, and 1.216; FR's are 0.928, 1.044, and 1.131.

Applying the declared conjunction of finite output, the overall threshold, and at least one period threshold yields F = ineligible and FR = eligible. Had neither candidate qualified, the protocol would have prohibited interpreting forecast-update or refit-sensitivity results.

Table 10: Pre-declared forecast-rule qualification

	Target-quarter-weighted RMSE			Ratio to better control			
Rule	Overall	Pre-COVID	Normalization	Overall	Pre-COVID	Normalization	Eligible
F	6.873	1.089	2.146	1.108	1.054	1.216	No
FR	5.757	1.078	1.996	0.928	1.044	1.131	Yes
Student-t location	6.203	1.033	1.765	—	—	—	Control
AR-t(1)	6.790	1.070	1.772	—	—	—	Control

Notes: The overall sample contains 387 package origins and 62 target quarters; pre-COVID contains 243 origins and 39 quarters through 2019Q4; and normalization contains 94 origins and 15 quarters from 2022Q1. Ratios use the smaller control RMSE in each column. Eligibility requires an overall ratio no larger than 1.05 and a ratio no larger than 1.05 in at least one of the two period columns, together with finite forecasts. Qualification permits package- update accounting only.

For all 387 eligible replay origins, the reconstructed complete package state matches the strict adapter and the post-package forecast matches its qualification forecast. The maximum recorded difference is 1.7764×10^{-15}, below the declared forecast tolerance of 10^{-8}. The identical-history deterministic refit sensitivity has mean absolute value 3.10×10^{-17} and maximum absolute value 1.7764×10^{-15}. These are numerical reproducibility checks. They are not estimates of how a forecaster would revise a production model after observing a release.

D Additional Empirical Results

Timestamp-precision results. Table 11 separates the full validated panel from the subset with official publication times. Date-only packages remain valid for the paper's conservative daily ordering but do not support an exact intraday claim.

Table 11: Timestamp-precision sensitivity of frozen-rule updates

Precision scope	Release family	Packages	Target quarters	$\overline{\lvert D_p \rvert}$	$\overline{\lvert v_p(\{R\}) \rvert}$	Q_R
All validated	All	387	62	0.561	0.078	0.139
All validated	Employment Situation	188	62	0.681	0.039	0.057
All validated	Federal Reserve G.17	199	62	0.448	0.114	0.255
Official time only	All	256	62	0.738	0.100	0.135
Official time only	Employment Situation	57	19	1.748	0.049	0.028
Official time only	Federal Reserve G.17	199	62	0.448	0.114	0.255

Notes: Updates are in percentage points; Q_R is unitless. The official-time-only restriction removes all conservative date-only packages. G.17 is unchanged because every G.17 package has an official time. The BLS restriction changes both the number and temporal composition of target-quarter clusters, so cross-scope differences are not timestamp-precision effects.

Full period-by-family results. Table 12 expands the main-text period partition by release family. The COVID/recovery concentration in complete package updates is most pronounced for the Employment Situation, while G.17 has the larger revision-only magnitude in every period. These are comparisons of descriptive moments; the protocol does not test equality across families or periods.

Table 12: Descriptive period-by-family package updates

Period	Release family	Packages	Target quarters	$\overline{\lvert D_p\rvert}$	$\overline{\lvert v_p(\{R\})\rvert}$	Q_R
Pre-COVID	All	243	39	0.202	0.048	0.236
Pre-COVID	Employment Situation	117	39	0.191	0.038	0.200
Pre-COVID	Federal Reserve G.17	126	39	0.212	0.057	0.267
COVID/recovery	All	50	8	2.676	0.151	0.056
COVID/recovery	Employment Situation	25	8	3.887	0.071	0.018
COVID/recovery	Federal Reserve G.17	25	8	1.465	0.231	0.158
Post-2021	All	94	15	0.364	0.117	0.320
Post-2021	Employment Situation	46	15	0.183	0.024	0.131
Post-2021	Federal Reserve G.17	48	15	0.538	0.205	0.382

Notes: Pre-COVID is scored over 2010–2019 despite the declared partition beginning in 2004; post-2021 ends at 2025Q3. Entries are deterministic partitions of sealed package outputs, without period-specific selection or inference. Updates are in percentage points except for Q_R.

Mechanically selected package examples. To avoid choosing visually dramatic releases after inspecting the updates, one package is selected mechanically within each period. Let m_j be the median of $|D_p|$ in period j. The reported package minimizes $\big||D_p| - m_j\big|$, with ties resolved first by origin and then by package identifier. The period medians are 0.163, 0.620, and 0.254 percentage points for pre-COVID, COVID/recovery, and post-2021. Table 13 reports the resulting same-origin histories.

Table 13: Mechanically selected representative package histories

Period	Package	Family	Target	Pre	Post	D_p	$v_p(\{H\})$	$v_p(\{R\})$	I_p	Precision
Pre-COVID	2018-04-17	G.17	2018Q2	2.248	2.084	−0.163	−0.125	0.020	−0.058	Official time
COVID/recovery	2020-09-04	Employment Situation	2020Q3	−3.736	−3.134	0.603	0.591	0.042	−0.030	Official time
Post-2021	2023-04-07	Employment Situation	2023Q2	1.654	1.916	0.262	0.256	0.006	0.001	Date only

Notes: Forecasts and updates are in annualized GDP-growth percentage points. Backfill-only updates are zero for all three packages. Component columns are singleton finite differences, and I_p is the aggregate interaction. The date-only package is assigned the conservative end-of-day clock and is not an exact intraday example.

The examples illustrate the accounting identity without claiming that a median package is representative of an entire release process. In the 2018 G.17 case, the positive revision-only term is more than offset by the negative headline and interaction terms. In the two Employment Situation cases, the headline term supplies most of the signed package update, while revision and interaction terms remain separately visible.

Core6 ledger coverage. Table 14 records the exact additional ledger coverage from CUMFNS and PCE. It is included to distinguish availability of auditable event history from eligibility for the qualified forecasting exercise; only the former is established for these two series.

Table 14: Core6 ledger-coverage sensitivity

Scope	Series	Event rows	Packages	First availability	Last availability	Headline eligible
Core4 headline	All	8,475	545	2004-02-06	2025-12-23	Yes
Core6 ledger only	All	13,808	806	2004-02-06	2025-12-23	No
Core6 ledger only	CUMFNS	2,779	147	2014-06-16	2025-12-23	No
Core6 ledger only	PCE	2,554	261	2004-03-01	2025-12-23	No

Notes: Distinct package totals do not equal the sum of series rows because one package can contain several series. Core6 is a coverage diagnostic only. No Core6 feature map, rule qualification, forecast score, or package update is reported.

References

Almuzara, M., Baker, K., O'Keeffe, H., and Sbordone, A. (2023). The new york fed staff nowcast 2.0. Technical report, Federal Reserve Bank of New York. Technical report.

Andreou, E., Ghysels, E., and Kourtellos, A. (2010). Regression models with mixed sampling frequencies. *Journal of Econometrics*, 158(2):246–261.

Anesti, N., Galvão, A. B., and Miranda-Agrippino, S. (2022). Uncertain kingdom: Nowcasting gross domestic product and its revisions. *Journal of Applied Econometrics*, 37(1):42–62.

Aruoba, S. B. (2008). Data revisions are not well behaved. *Journal of Money, Credit and Banking*, 40(2–3):319–340.

Bánbura, 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, pages 195–237. Elsevier, Amsterdam.

Bell, W. R., Di Fonzo, T., Ladiray, D., McElroy, T. S., and Smith, P. A. (2024). Preface to time series special issue of the journal of official statistics. *Journal of Official Statistics*, 40(4):499–507.

Biemer, P. P., Trewin, D., Bergdahl, H., and Japec, L. (2014). A system for managing the quality of official statistics. *Journal of Official Statistics*, 30(3):381–415.

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.

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

European Statistical System (2017). *European Statistics Code of Practice*. Eurostat, Luxembourg, revised edition edition.

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.

Hayashi, F. and Tachi, Y. (2021). The nowcast revision analysis extended. *Economics Letters*, 209:110112.

Higgins, P. (2014). GDPNow: A model for GDP "nowcasting". Working Paper 2014-7, Federal Reserve Bank of Atlanta.

Jacobs, J. P. A. M. and van Norden, S. (2011). Modeling data revisions: Measurement error and dynamics of "true" values. *Journal of Econometrics*, 161(2):101–109.

Kenett, R. S. and Shmueli, G. (2016). From quality to information quality in official statistics. *Journal of Official Statistics*, 32(4):867–885.

Koenig, E. F., Dolmas, S., and Piger, J. (2003). The use and abuse of real-time data in economic forecasting. *Review of Economics and Statistics*, 85(3):618–628.

Mazzi, G. L., Mitchell, J., and Carausu, F. (2021). Measuring and communicating the uncertainty in official economic statistics. *Journal of Official Statistics*, 37(2):289–316.

McCracken, M. W. and Ng, S. (2016). FRED-MD: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, 34(4):574–589.

O'Keeffe, H. and Petrova, K. (2025). Component-based dynamic factor nowcast model. Staff Report 1152, Federal Reserve Bank of New York.

Shapley, L. S. (1953). A value for n-person games. In Kuhn, H. W. and Tucker, A. W., editors, *Contributions to the Theory of Games II*, volume 28 of *Annals of Mathematics Studies*, pages 307–317. Princeton University Press, Princeton, NJ.

Signore, M., Scanu, M., and Brancato, G. (2015). Statistical metadata: A unified approach to management and dissemination. *Journal of Official Statistics*, 31(2):325–347.

Struijs, P., Camstra, A., Renssen, R., and Braaksma, B. (2013). Redesign of statistics production within an architectural framework: The dutch experience. *Journal of Official Statistics*, 29(1):49–71.

United Nations Economic Commission for Europe (2025). *Generic Statistical Business Process Model: Version 5.2*. United Nations Economic Commission for Europe, Geneva.

United Nations Statistics Division (2019). *United Nations National Quality Assurance Frameworks Manual for Official Statistics: Including Recommendations, the Framework and Implementation Guidance*. United Nations, New York.

Zając, A., Karlberg, M., and Museux, J.-M. (2024). The european statistics awards for nowcasting: A new approach to engage with the scientific community and to foster improved timeliness of official statistics. *Journal of Official Statistics*, 40(4):633–658.