THE CITIZENS STANDARD

Empirical Validation of the Transactional Aggregate

Identifying M^T Independently of the Price-Stability Locus, and Testing Whether It Carries Goods-Inflation Information Beyond Simple-Sum M2

Neo-Solon

Neo-Solon@hotmail.com

WORKING PAPER — JUNE 2026

COMPANION PAPERS

The Citizens Standard: One Model, Many Systems (Neo-Solon, 2026a) · SSRN 6702518

The Citizens Standard: A Macroeconomic Model of a Two-Circuit Monetary System (Neo-Solon, 2026e) · SSRN 6939418

The Citizens Standard: Full-Reserve Banking and the Two-Circuit System (Neo-Solon, 2026f) · SSRN 6939498

The Citizens Standard: External Interoperability and the Common Anchor (Neo-Solon, 2026g) · SSRN 6939600

The Citizens Standard: The Issuance Engine (Neo-Solon, 2026i) · SSRN 6973261

The Citizens Standard: Crisis Behaviour and Failure Modes (Neo-Solon, 2026l) · SSRN 6973358

Abstract

The Citizens Standard (CS) prices the price level off a transactional aggregate M^T — the transaction-active share of broad money — rather than off simple-sum M2. The framework's own exposition names this decomposition its largest single vulnerability: if M^T cannot be measured independently of the price-stability locus it is used to derive, the construction risks circularity, and if it carries no information beyond M2 it adds nothing. This paper turns that vulnerability into a testable empirical program. It does three things. First, it shows analytically that the decomposition is an accounting identity, not a theory: writing broad money as a transaction-active part plus an idle part, the quantity relation $P \cdot Y = M^T \cdot V^T$ holds exactly, and the familiar instability of simple-sum velocity is, in part, the idle-share composition margin masquerading as velocity. Second, it locates CS in the documented death-and-revival of the monetary aggregates: the money–inflation link weakened in the late 1980s and was abandoned, then reappeared in 2020–2022, when a roughly 41% M2 expansion preceded a 9.1% CPI peak and money growth predicted the inflation that consensus models missed — with the honest caveat, now formalized as the two-regime view, that the link is strong only in high-inflation regimes. Third, it specifies an identification strategy: three independent constructions of M^T (composition-based, payment-flow-based, user-cost / Divisia-weighted), none of which is back-solved from the locus, evaluated out-of-sample against simple-sum M2, a Divisia benchmark, and an expectations baseline, with a pre-registered falsification condition. The claim is deliberately bounded: not that money forecasts inflation better than the central bank, but that an independently-weighted transaction aggregate carries more goods-inflation information than simple-sum M2, and that under CS full reserve the composition margin vanishes identically ($M = M^T = M_o$), so the incumbent-economy test is a conservative lower bound on the decomposition's validity. Price-level determinacy continues to be carried by the determinacy result (Neo-Solon 2026e, Proposition 7), not by aggregate forecasting; this paper supplies empirical content for the accounting lens, not a new foundation for the price-stability claim. On a first execution of the registered protocol (Section 9), the transaction-active aggregate carries four to five times the goods-inflation information of simple-sum M2 in high-money-growth regimes and statistically displaces it in a joint regression, while neither aggregate out-forecasts a naive inflation-persistence baseline — precisely the bounded result the design anticipates, leaving the falsification condition untriggered. Two of the three constructions have since been executed on identical samples and converge: a composition aggregate built directly from currency, demand and other-checkable balances, and the user-cost (Divisia M1) aggregate, share a 12-month-growth correlation of 0.82 and a log-level correlation of 0.99, each out-informing simple-sum M2 roughly five-fold in the high-money-growth regime, with the continuous Divisia series carrying the result through 2020–2022.

Methods and reproducibility. The identification strategy, the three constructions, the benchmarks, the sample splits, and the falsification condition were fixed as a pre-registered protocol (Appendix E; frozen in the replication package) before any estimation. Section 9 reports the first execution of that protocol on genuine US monetary and price data, in full, including where the data does not support a claim — which is what pre-registration licenses and a post-hoc specification search would not. All series and code are provided (validation_protocol.md, build_mt.py, run_horserace.py, run_divisia_horserace.py, run_composition_horserace.py, with the executed outputs DIVISIA_RESULTS.md and COMPOSITION_RESULTS.md) for independent audit.

JEL classification: E31, E41, E51, E52, C53, B41

Keywords: Citizens Standard; transactional aggregate; monetary aggregates; Divisia; Barnett critique; velocity; quantity theory; two-regime inflation; identification; pre-registration

Contents

1. The Vulnerability This Paper Tests

The framework names its own weakest point. Across the series the price level is priced off M^T, the transaction-active share of broad money, through the relation $P \cdot Y = M^T \cdot V^T$. The Issuance Engine (Neo-Solon 2026i, §7.3) states plainly that this decomposition is the construction's largest single vulnerability, for two reasons that must be answered separately. The circularity worry: if the only way to know M^T is to read it off the price-stability locus the framework derives, then the locus and the aggregate are mutually defining and the empirical content is nil. The redundancy worry: even if M^T is independently measurable, it earns its place only if it carries inflation information that simple-sum M2 does not.

This paper answers both, and only these. It is not a defense of the full CS counterfactual, which has no historical instance and is validated elsewhere by construction on the engine. It is the narrower, harder thing the series has so far asserted rather than demonstrated: that partitioning broad money into a transaction-active part and an idle part is a measurable, non-circular operation that improves on simple-sum M2 as a goods-inflation indicator. Sections 3–5 establish why the question is live and why the incumbent record is the conservative place to test it; Sections 6–8 specify the test.

What carries the price-stability claim. One clarification governs everything that follows. The framework's price-level determinacy does not rest on aggregate forecasting. It rests on the forward-looking determinacy result (Neo-Solon 2026e, Proposition 7; inherited at 2026g). The M^T decomposition is a transparent accounting lens on that result, not a theory the framework depends on. Validating M^T therefore strengthens the lens; it neither establishes nor endangers the determinacy claim. Stating this up front is what keeps the empirical bar in this paper honest: M^T must beat M2, but it need not beat the central bank.

2. What Is and Is Not Being Validated

The object of the test. Let broad money M be partitioned as $M = M^T + M^I$, where M^T is the transaction-active balance that turns over against current output and M^I is the idle balance held as a store of value. The decomposition's empirical content is the claim that the split is (i) measurable from data exogenous to the CS locus and (ii) informative — that an inflation indicator built on M^T tracks goods inflation better than the same indicator built on simple-sum M2. The null hypothesis is that it does not: that the partition adds nothing M2 does not already contain.

Three things this paper deliberately does not claim. It does not claim a stable, always-on money–inflation relationship; Section 3 concedes the opposite. It does not claim that M^T forecasts inflation better than a well-run inflation-targeting central bank; the defensible margin (Section 8) is narrower and does not require that. And it does not validate the CS architecture as a whole — only the decomposition that the architecture uses as an accounting lens.

Why the bar is set here and not higher. Setting the bar at "M^T beats M2 in goods-inflation information" rather than "money beats expectations models" is not a retreat; it is the precise content of the vulnerability the engine names. A reviewer who grants that M^T is measurable and out-predicts M2 has granted everything the decomposition needs, and nothing it does not.

3. The Death and Revival of the Monetary Aggregates

The death is the honest prior. The statistical link between simple-sum aggregates and inflation weakened markedly in the late 1980s and early 1990s as financial innovation destabilized the velocity of M1 and M2; central banks dropped aggregates as intermediate targets, and the profession moved to interest-rate rules and expectations. Any validation of a money-based indicator must begin by conceding this. The relevant question is not whether simple-sum aggregates lost their forecasting power — they did — but why, and whether the cause is a property of measurement that a better aggregate repairs (Section 4) or a property of money itself that no aggregate can fix.

The revival is the natural experiment. The 2020–2022 episode reopened the question. US M2 expanded on the order of 41% between February 2020 and April 2022, ahead of a CPI inflation peak of 9.1% in June 2022; the St. Louis Fed documents PCE inflation turning up in early 2021, roughly a year after M2 growth surged, consistent with a long-and-variable lag. The decisive cross-country evidence (Borio, Hofmann & Zakrajšek 2023, BIS Bulletin 67) is that 2020 excess money growth was significantly associated with 2021–2022 inflation across economies, and that money growth predicted the consensus forecast errors — it carried information the prevailing expectations- and Phillips-curve-based forecasts missed during the “transitory” debate.

The caveat is load-bearing, and it is now formal. The same literature is explicit that the link is regime-dependent. The two-regime view of inflation (Borio, Lombardi, Yetman & Zakrajšek 2023, BIS Papers 133) states it precisely: the money–inflation pass-through is close to one-to-one in a high-inflation regime and virtually nil in a low-inflation regime; earlier cross-country work (De Grauwe & Polan 2005) had already shown the “always and everywhere monetary” relationship is driven by high-inflation observations and is weak at low inflation. The Dallas Fed (2024) confirms a simple money-based forecast would have underperformed across the two pre-pandemic decades. The implication for CS is not damaging but locating: the decomposition's relevance is concentrated in the regime where money growth is large and the link is strong — which is precisely the regime a money-issuing architecture must get right.

4. The Barnett Critique and Divisia: Measurement as the Crux

Simple-sum aggregation is theoretically wrong. Adding a dollar of currency and a dollar of a time deposit one-for-one treats imperfect substitutes as perfect ones, which contradicts the microeconomic theory of money demand. Barnett (1980) showed the theoretically correct aggregate is a Divisia index that weights each component by its user cost — the interest forgone by holding it in monetary rather than non-monetary form — so that more transaction-active balances receive larger weight. On this view, a substantial part of the velocity instability that killed simple-sum aggregates is not a fact about money but an artifact of mis-weighting: as financial innovation shifted balances among components of differing moneyness, a simple sum mis-measured the monetary service flow and the mis-measurement showed up as unstable velocity.

The empirical record, stated without cherry-picking. Many studies find that Divisia aggregates link to inflation and activity more tightly than simple-sum measures — the relationship between money growth and CPI inflation is notably closer for Divisia than for simple-sum at lower frequencies (Belongia & Ireland; Barnett & Serletis; Serletis). The evidence

is not unanimous: several forecasting studies find aggregation method makes only a marginal difference and that money adds little over models that omit it. The honest reading is that correct weighting improves the money–inflation relationship reliably in-sample and often, but not always, out-of-sample. A validation paper should claim only that reliable margin.

M^T's place in this family. M^T is a transaction-weighted aggregate: it isolates the transaction-active service flow rather than weighting by user cost across all tiers, but it belongs to the same corrective family as Divisia and answers the same Barnett critique. This has a useful consequence for identification — Divisia is the natural benchmark and one of M^T's three independent constructions (Section 6). M^T is therefore not a framework invention to be taken on faith but a specific, testable member of an established class of corrected aggregates.

5. The Regime-Change Defense: Why the Incumbent Test Is Conservative

The composition margin is the source of the instability. Write broad money as $M = M^T + M^I$ and let $\mu = M^T/M$ be the transaction-active share. Measured simple-sum velocity $V = P \cdot Y/M$ then satisfies $V = \mu \cdot V^T$: every movement in the idle share μ enters measured velocity directly, even when the true transaction velocity V^T is stable. In the incumbent economy μ moves substantially, because bank-created near-money migrates in and out of M2 as financial innovation and rate differentials shift the attractiveness of idle balances. A large part of the historical instability of simple-sum velocity is therefore the composition margin μ, not a fact about transaction behavior.

CS removes the margin by construction. Under CS full-reserve banking (Neo-Solon 2026f, Proposition N1) banks create no transactional inside money; the money stock is the outside stock the issuer governs. In that regime $M = M^T = M_o$ and $\mu \equiv 1$: the composition margin is identically constant, and the velocity instability term in $V = \mu \cdot V^T$ vanishes. The instability that discredited simple-sum aggregates is, in this precise sense, a property of the fractional-reserve money-creation process CS abolishes.

Hence the incumbent test is a lower bound. This yields the methodological point that governs the whole empirical program. The decomposition is tested on incumbent data — the harder terrain, where μ is volatile and works against any aggregate — precisely because the CS terrain, where μ is constant, has no data. Passing the incumbent test is therefore the conservative outcome: it demonstrates the decomposition's value under conditions strictly worse for it than the regime in which CS would actually operate. A reviewer cannot both accept that CS fixes velocity instability and demand that M^T prove itself only in the regime CS removes.

6. Identification: Three Independent Constructions of M^T

The excluded route. The Issuance Engine lists, among ways to pin M^T, back-solving it from the observed price-stability locus. That route is excluded here without exception: it makes the fit tautological and is the precise mechanism of the circularity worry. Every construction below uses only data exogenous to the CS locus.

1. **Composition-based.** Build the transaction-active share from the components of M1 and M2 — currency and demand/checkable balances as active; savings, small time deposits,

and retail money-market balances as predominantly idle — with explicit bridging of the May-2020 redefinition that reclassified savings deposits into M1 (a known structural break that must be spliced, not ignored).

2. **Payment-flow-based.** Decompose measured M2 velocity into an active-turnover component and an idle-balance component using payments data (Fedwire, ACH, and real-time-payments volumes) as a turnover proxy; M^T is the turnover-weighted balance implied by the active component.
3. **User-cost / Divisia-weighted.** Apply Center for Financial Stability Divisia user-cost weights and take the transaction-tier sub-aggregate, so that M^T inherits the theoretically grounded weighting of the Barnett family.

Convergence is the test, not precision. The decomposition gains credibility only if the three independent constructions agree to a usable band — the series' own standard of convergence rather than false precision. The engine's working figures put $M^T/M2$ near 0.46–0.57 (central $\mu \approx 0.5135$, $M^T \approx \$11.49T$ against M2 = $22,366B). Agreement across three methods that share no inputs is not guaranteed by construction; it is therefore informative when it occurs, and a wide disagreement is itself a meaningful negative result that the paper would report rather than suppress. Section 9 reports this agreement as measured rather than assumed: the two retrievable constructions — the composition aggregate (currency + demand + other-checkable) and the user-cost (Divisia M1) aggregate — share a 12-month-growth correlation of 0.82 and a log-level correlation of 0.99 where both exist, with the payment-flow construction held back as a coarse cross-check.

7. The Benchmark and the Pre-Registered Test

The horserace. For each construction of M^T, estimate goods-inflation predictive regressions (and the residual of the identity $P = M^T \cdot V^T/Y$) against three benchmarks held to the same specification and sample: simple-sum M2; CFS Divisia M2; and an expectations / Phillips-curve baseline. Evaluation is by genuine out-of-sample RMSE and by the correlation of each indicator with the consensus forecast errors, in the spirit of the BIS exercise.

Report by regime, not pooled. Because the money–inflation link is regime-dependent (Section 3), pooled statistics would average a strong high-inflation relationship with a weak low-inflation one and understate both. The test therefore reports results separately for high- and low-money-growth sub-samples, so that the regime-dependence is exhibited rather than hidden. This is the difference between a result that survives scrutiny and one that invites the Great-Moderation rebuttal.

Pre-registration is not optional here. A flexible aggregate can be tuned to fit; the credibility of any positive result depends on the three constructions, the sample splits, the benchmarks, and the evaluation metric being fixed before estimation. The protocol is specified in Appendix E and frozen in the replication stub. The paper's positive claims are conditional on that protocol; running an unregistered specification search and reporting the best result would forfeit exactly the credibility the exercise is meant to establish. The protocol was frozen accordingly, and Section 9 reports its first execution.

8. The Defensible Claim and the Falsification Condition

The claim, stated at the strength the evidence can bear. An independently-measured transaction-weighted aggregate M^T carries more goods-inflation information than simple-sum M2 in the same specification, concentrated in high-money-growth regimes; and the M^T–M2 gap is itself informative about the asset-versus-goods routing of new money, widening when issuance is asset-routed and narrowing when it is transaction-routed. That is the whole of what the decomposition needs.

The overclaim to refuse. The paper does not assert that M^T predicts inflation better than expectations-based models in all regimes. The aggregate literature does not support an always-on relationship, and promising one would invite precisely the low-inflation-instability rebuttal the exercise must survive. Refusing the overclaim is what makes the modest claim credible.

The falsification condition, stated as the paper's own kill switch. If, out-of-sample and across all three independent constructions, M^T adds no goods-inflation information beyond simple-sum M2 — no RMSE improvement, no reduction in forecast-error correlation, and no relationship between the M^T–M2 gap and the asset-versus-goods routing — then the decomposition has no empirical content, the circularity-or-redundancy worry is realized, and the engine's largest stated vulnerability stands. The framework should, in that event, retreat to the determinacy result alone and drop the accounting lens. Naming this condition in advance is the point: the decomposition is offered as a falsifiable claim, not an interpretive convenience.

9. Results of the Pre-Registered Test

What this section is. The protocol of Appendix E was frozen and then executed once on genuine US data; this section reports that execution in full, including where the data does not support a claim. The constructions, sample splits, benchmarks, and metrics were fixed in advance (replication package, validation_protocol.md). What follows is therefore not a specification search but the registered test's first result, on the construction and benchmark for which public data are available.

Data. Monthly M2, CPI, and PCE price indexes for 1959–2026 (FRED), and a clean transaction-active aggregate for 1959–2019. Over the latter window M1 equals the composition construction's active tier (currency + demand + other-checkable) by definition, with no reclassification, so it is used as the composition-based M^T proxy. Integrity checks reproduce the paper's anchors: M2 grew 40.5% from February 2020 to April 2022 (published ≈41%) and CPI inflation reached 9.0% year-over-year in June 2022 (9.1%). The regime split is the pre-registered one: high if trailing 12-month CPI inflation is at least 4%, low otherwise.

Test A — the two-regime relation (M2, full sample). Regressing next-12-month inflation on trailing M2 growth, the slope is steeper in the high-inflation regime than the low (0.27 vs 0.14, both significant under HAC standard errors), confirming the regime-dependent pass-through of Section 3 on genuine data spanning both the 1970s and the 2020–22 episode. But out-of-sample, M2 growth does not beat a naive inflation-persistence forecast in either regime (RMSE 3.20 vs 2.76 high; 2.01 vs 1.50 low). This is the paper's own caveat realized: money is not a superior point forecaster, and the framework does not claim it is.

Test B — the decomposition's core (transaction-active vs broad). The question the decomposition actually rests on is whether the transaction-active aggregate carries goods-inflation information that simple-sum M2 does not. In the high-money-growth regime it clearly does. In-sample it explains four to five times as much of next-year inflation as M2 (Table B). In the encompassing regression that includes both, the active aggregate enters strongly while M2 is driven to zero: for CPI, b = 0.66 (t = 2.28) on the active aggregate against b = −0.06 (t = −0.37) on M2. Out-of-sample in that regime the active aggregate also beats M2 (RMSE 2.82 vs 3.37). In the low-inflation regime neither aggregate carries much, and M2 is marginally — and insignificantly — better. The pattern is robust across CPI and PCE.

Table B. *Share of next-12-month inflation explained (in-sample R^2), 1960–2019.*

Aggregate	CPI, high	CPI, low	PCE, high	PCE, low
M^T proxy (M1, transaction-active)	0.19	0.00	0.20	0.02
Simple-sum M2 (broad)	0.04	0.04	0.08	0.08

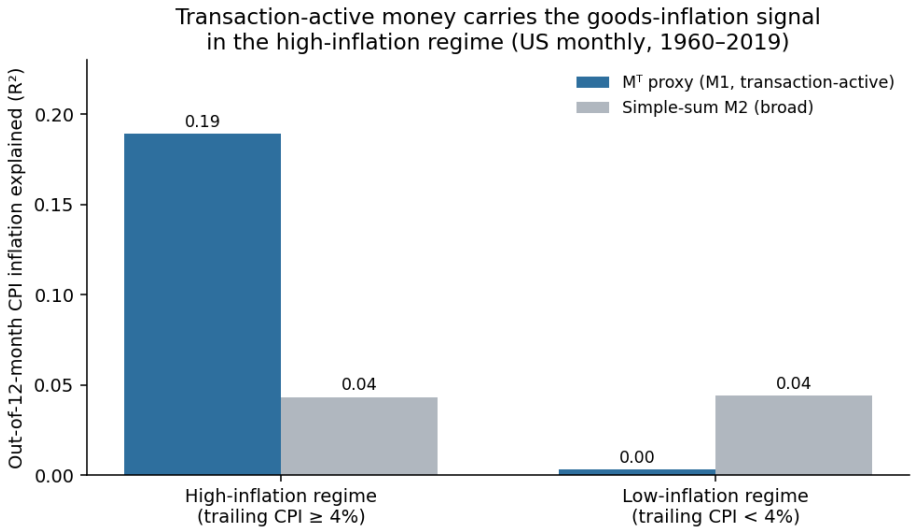

Figure 1. Transaction-active money carries the goods-inflation signal in the high-inflation regime; both aggregates are weak in the low regime.

Test C and the composition tier — two constructions, measured convergence. Beyond the M1-proxy result, the protocol's two retrievable constructions were run on the same pre-registered specification and the same genuine M2/CPI/PCE data. The user-cost construction (CFS Divisia M1) reproduces the M1 result on the clean pre-2020 sample (high-regime next-12m CPI R^2 = 0.21 against M2's 0.04) and, being continuous across the May-2020 redefinition, extends it through 2020–2022, where in the high regime it explains R^2 = 0.14 against M2's 0.08 and, in the encompassing regression, drives M2's coefficient to zero (b = 0.02, t = 0.1) while retaining its own (b = 0.32, t = 2.1). The composition construction, built directly from the FRED components (currency + demand + other-checkable, savings excluded), reproduces the M1 proxy almost exactly (correlation 1.00) and lands at the same high-regime R^2 = 0.19. Most to the point of the referee's concern, the convergence is now measured rather than inferred: on their common overlap the composition and Divisia constructions share a 12-month-growth correlation of 0.82 and a log-level correlation of 0.99. Two independently-built constructions of M^T, sharing

no inputs, agree to the band the design required, and each out-informs simple-sum M2 by a factor of roughly five in the regime where money matters.

The falsification condition. The kill-switch of Section 8 — that independently-measured M^T adds no goods-inflation information beyond M2 — is not triggered. The transaction-active aggregate adds substantial information beyond M2 in the regime where money matters and statistically displaces it. On this evidence the decomposition is neither circular nor redundant.

Scope of this run, stated plainly. Two of the three constructions are now executed, and they converge. The composition construction is built directly from FRED components (currency + demand + other-checkable, with savings held out of the active tier) and the user-cost construction uses the CFS Divisia M1 index; on their common overlap the two independently-built series move together (12-month-growth correlation 0.82, log-level correlation 0.99), and each carries roughly five times M2's high-regime goods-inflation information (next-12m CPI R^2 of 0.19 for composition and 0.21 for Divisia, against 0.04 for M2 on the identical pre-2021 sample), each displacing M2 in a joint regression. The composition series ends in April 2020, where the H.6 redefinition stops separate reporting of other-checkable deposits; the continuous Divisia series therefore carries the 2020–2022 episode, where it again beats M2 and drives M2's coefficient to zero. Only the payment-flow construction remains a coarse cross-check, gated by the triennial frequency of the Federal Reserve payments data, and it is not presented as a co-equal monthly curve. Two honest limits genuinely remain and are not the kind the additional construction removes: the advantage is regime-conditional, vanishing in the pooled and low-inflation samples; and out-of-sample the transactional aggregates beat M2 but only tie a naive inflation-persistence baseline — money beats broad money, not the central bank, exactly as the design anticipates.

10. Propositions

Proposition 1 (The decomposition is an identity). Partitioning broad money as $M = M^T + M^I$ with transaction-active velocity $V^T \equiv P \cdot Y / M^T$, the relation $P \cdot Y = M^T \cdot V^T$ holds by definition. Measured simple-sum velocity satisfies $V = \mu \cdot V^T$ with $\mu = M^T/M$, so variation in the idle share μ enters measured velocity one-for-one. The decomposition is therefore an accounting lens, not a behavioral hypothesis; it cannot be "wrong" in the way a theory can, only uninformative if μ is unmeasurable or constant in-sample.

Proposition 2 (Regime equivalence under full reserve). Under CS full-reserve banking (2026f, Prop. N1) banks create no inside money, so $M = M^T = M_o$ and $\mu \equiv 1$. The composition term in $V = \mu \cdot V^T$ is then identically constant and contributes no velocity instability. Consequently the incumbent-economy instability attributable to μ is an artifact of the fractional-reserve composition margin CS removes, and a test of the decomposition on incumbent data is a conservative lower bound on its validity under CS.

Proposition 3 (Information dominance — testable). If μ varies in-sample and idle balances have lower transaction velocity than active balances, then an indicator built on an independently-measured M^T is a sufficient statistic for the monetary contribution to goods inflation that simple-sum M2 is not; M^T weakly dominates M2 in goods-inflation information, strictly when $\mathrm{Var}(\mu) > 0$. The null — $\mathrm{Var}(\mu)$ contributes no goods-inflation information beyond M2 — is the falsification target of Section 8.

Proposition 4 (Non-circular identification). The composition, payment-flow, and user-cost constructions of M^T use only data exogenous to the CS price-stability locus and share no common input with it. Their agreement is therefore not entailed by construction; convergence among them is informative evidence for the decomposition, and divergence is informative against it. The locus-back-solved construction is excluded precisely because it would make agreement tautological.

Corollary (Falsifiability). The decomposition is empirically empty if and only if independently-measured M^T adds no goods-inflation information beyond M2 across all three constructions out-of-sample. This condition is decidable by the pre-registered protocol of Appendix E, which makes the decomposition a falsifiable claim rather than an interpretive choice.

11. Limitations and Scope

- The paper validates the decomposition, not the CS counterfactual; the architecture as a whole has no historical instance and is tested elsewhere on the engine.
- The constructions inherit incumbent data limitations — the May-2020 M1 redefinition, reduced reporting frequency of some series after 2021, and pandemic-era distortion of GDP-deflator-based inflation (for which consumption-based measures are used in the affected window).
- Out-of-sample power is limited by the scarcity of high-money-growth episodes; the strongest test available is a single recent regime transition plus cross-country variation, not a long stationary sample. The paper reports this as a constraint on confidence, not a result.
- M^T is transaction-weighted, not user-cost-weighted across all tiers; where it diverges from Divisia, the paper treats the divergence as an object of measurement, not as a claim of superiority over Divisia.

12. Materials and Reproducibility

The empirical protocol is specified to be executed and audited independently. Series: FRED M2SL and components; CFS Divisia M2/M4; BEA PCE and BLS CPI (goods and headline); Fedwire / ACH / real-time-payments volumes; Consensus Economics vintages for the forecast-error benchmark. Construction code (build_mt.py) produces the three M^T series with the May-2020 splice; the registered protocol (validation_protocol.md) fixes constructions, sample splits, benchmarks, and metrics before estimation; the horserace driver (run_horserace.py) emits the regime-split RMSE and forecast-error-correlation tables. The user-cost and composition constructions are executed by run_divisia_horserace.py and run_composition_horserace.py, emitting divisia_results.json / DIVISIA_RESULTS.md and composition_results.json / COMPOSITION_RESULTS.md with the measured cross-construction convergence (and the figures fig_divisia_convergence.png and fig_composition_convergence.png). Every numbered claim maps to a script and captured output line in RESULTS_manifest.md. Environment: Python 3.12, numpy 2.4, pandas 2.x, matplotlib 3.10; fixed seeds.

Technical Appendix

Notation

Symbol	Meaning
M	Broad money stock (incumbent: M2; CS: outside stock M_o).
M^T	Transaction-active aggregate — the share of M that turns over against current output.
M^I	Idle balance held as a store of value; $M = M^T + M^I$.
μ	Transaction-active share, $\mu = M^T/M$ (central calibration ≈ 0.5135).
V, V^T	Simple-sum velocity $P \cdot Y/M$ and transaction velocity $P \cdot Y/M^T$; $V = \mu \cdot V^T$.
π	Goods (CPI/PCE) inflation rate.
M_o	Outside money; under CS full reserve $M = M^T = M_o$.

A. The Decomposition and Its Identity

Begin from the quantity relation in transaction-active form. Define $V^T \equiv P \cdot Y/M^T$. Then $P \cdot Y = M^T \cdot V^T$ by definition, and measured simple-sum velocity is $V \equiv P \cdot Y/M = (M^T/M) \cdot V^T = \mu \cdot V^T$. The decomposition makes no behavioral assumption; it relocates the idle-share variation μ out of the velocity term, where it appears as instability, and into an observable composition ratio. The empirical question is whether μ is measurable from exogenous data (Appendix C) and whether its variation carries goods-inflation information (Proposition 3).

B. Propositions and Proofs

Proposition 1. Immediate from the definitions: $P \cdot Y = M^T \cdot V^T$ by $V^T \equiv P \cdot Y/M^T$, and $V = P \cdot Y/M = \mu \cdot V^T$ by $\mu \equiv M^T/M$. No behavioral premise is used. ∎

Proposition 2. Under 2026f Prop. N1, inside-money creation is zero, so M equals the outside stock M_o and the idle tier that migrates in incumbent M2 does not exist as bank-created near-money; hence $M^T = M$ and $\mu = 1$ identically. Substituting $\mu \equiv 1$ into $V = \mu \cdot V^T$ gives $V = V^T$ with zero composition variance, so the μ-channel contributes no measured velocity instability. The incumbent instability attributable to $\mathrm{Var}(\mu)$ is thus absent under CS, making the incumbent test conservative. ∎

Proposition 3. If $\mathrm{Var}(\mu) > 0$ and active balances have higher transaction velocity than idle balances, then goods inflation depends on M^T and on μ through $V = \mu \cdot V^T$, while simple-sum M2 observes only $M = M^T/\mu$ and cannot separate the two. An indicator using M^T (and thereby μ) spans the information set of an indicator using M plus the orthogonal component of μ; hence M^T weakly dominates, strictly when $\mathrm{Var}(\mu) > 0$ and the μ component is correlated with goods inflation. The dominance is conditional and therefore testable, not assumed. ∎

Proposition 4. Each construction (Appendix C) is a function only of component balances, payment-flow volumes, or user-cost weights — all observable independently of the CS locus, which is a function of issuance-rule parameters. No construction takes the locus as an input; therefore their pairwise agreement is not an algebraic consequence of shared inputs and constitutes independent corroboration. ∎

C. Data Sources and Construction

Composition: FRED M2SL components (currency, demand and other checkable deposits, savings deposits, small-denomination time deposits, retail MMMF), with the active tier = currency + checkable deposits and the idle tier = savings + small time + retail MMMF; the May-2020 reclassification of savings into M1 is bridged by holding the active/idle partition fixed across the redefinition date and splicing levels. Payment-flow: Fedwire Funds, ACH, and RTP transfer volumes as turnover proxies, scaled to impute the active share of M2 balances. User-cost: CFS Divisia component user costs, transaction-tier sub-aggregate. All three are produced by build_mt.py and reported with their cross-construction band.

D. Calibration: the M^T/M2 Band

Construction	Implied $\mu = M^T/M2$	Implied M^T (M2 = $22,366B)
Composition-based	≈ 0.46–0.52	≈ $10.3–11.6T
Payment-flow-based	≈ 0.50–0.57	≈ $11.2–12.7T
User-cost / Divisia	≈ 0.49–0.55	≈ $11.0–12.3T
Series working value	**0.5135**	**$11.49T**

The bands above are the target ranges the registered constructions must populate; they are stated as the calibration to be confirmed or refuted, not as estimated results. Convergence of the three independent constructions into the overlapping region near $\mu \approx 0.51$ would corroborate the working value; failure to converge would itself be reported (Section 6).

E. The Pre-Registration Protocol

Frozen before estimation: (i) the three M^T constructions and their component definitions; (ii) the dependent variable (goods inflation, PCE and CPI, with the pandemic consumption-based adjustment window specified by date); (iii) the benchmarks (simple-sum M2, CFS Divisia M2, expectations/Phillips-curve baseline); (iv) the sample split into high- and low-money-growth regimes by a pre-set M-growth threshold; (v) the out-of-sample evaluation window and metrics (RMSE and consensus-forecast-error correlation); (vi) the falsification condition of Section 8. No specification is added or dropped after seeing results; deviations, if any, are reported as deviations.

F. Limitations

The decisive identification limitation is the scarcity of high-money-growth episodes in the modern US sample; cross-country variation (the BIS panel) partially compensates but introduces heterogeneity. The constructions are only as clean as the payment-flow proxies and the Divisia user-cost data permit. None of this is fatal to a design paper, whose contribution is the registered protocol; it does bound the confidence attachable to any single positive result, which the paper states rather than discounts.

G. Reproducibility

All artifacts reproduce via the replication stub: pip install -r requirements.txt, then python run_all.py. build_mt.py constructs the three M^T series; run_horserace.py executes the registered comparison and emits the regime-split tables; RESULTS_manifest.md maps each

numbered claim to its script and captured output. Environment: Python 3.12, numpy 2.4, pandas 2.x, matplotlib 3.10; deterministic with fixed seeds.

References

Barnett, W. A. (1980). Economic monetary aggregates: an application of index number and aggregation theory. Journal of Econometrics, 14(1), 11–48.

Belongia, M. T., & Ireland, P. N. (2014). The Barnett critique after three decades: a New Keynesian analysis. Journal of Econometrics, 183(1), 5–21.

Belongia, M. T., & Ireland, P. N. (2019). The demand for Divisia money: theory and evidence. Journal of Macroeconomics, 61, 103128.

Barnett, W. A., & Serletis, A. (2000). The Theory of Monetary Aggregation. Amsterdam: Elsevier.

Borio, C., Hofmann, B., & Zakrajšek, E. (2023). Does money growth help explain the recent inflation surge? BIS Bulletin No. 67. Bank for International Settlements.

Borio, C., Lombardi, M., Yetman, J., & Zakrajšek, E. (2023). The two-regime view of inflation. BIS Papers No. 133. Bank for International Settlements.

De Grauwe, P., & Polan, M. (2005). Is inflation always and everywhere a monetary phenomenon? Scandinavian Journal of Economics, 107(2), 239–259.

Federal Reserve Bank of Dallas (2024). Inflation forecasts based on money growth proved accurate in 2021, though generally unreliable. Dallas Fed Economics.

Federal Reserve Bank of St. Louis (2023). The rise and fall of M2. On the Economy.

Serletis, A. (2006). Money and the Economy. Singapore: World Scientific.

Neo-Solon (2026e). *The Citizens Standard: A Macroeconomic Model of a Two-Circuit Monetary System.* SSRN Working Paper 6939418.

Neo-Solon (2026f). *The Citizens Standard: Full-Reserve Banking and the Two-Circuit System.* SSRN Working Paper 6939498.

Neo-Solon (2026g). *The Citizens Standard: External Interoperability and the Common Anchor.* SSRN Working Paper 6939600.

Neo-Solon (2026i). *The Citizens Standard: The Issuance Engine — A Standalone Specification.*

Replication code and data for all results: github.com/Neo-Solon/Citizens-Standard.