WHEN THE DESIGN OUTWEIGHS THE OPERATOR: FRACTIONAL DIFFERENCING, PLUG-IN MEMORY, AND FORECAST EVALUATION FOR HARMONISED INFLATION

CALIN-ADRIAN COMES
Department of Economics, Faculty of Economics and Law, George Emil Palade University of Medicine, Pharmacy, Science and Technology of Târgu Mureş and Institute of National Economy, Romanian Academy

This version: 30 August 2026

How large is the out-of-sample gain from fractional differencing, relative to the design choices made while measuring it? For a nested pair, an autoregression and the same autoregression on a fractionally differenced series with a plug-in memory estimate, the forecast wedge is first-order linear in the plug-in; with fixed coefficients its coefficient is a harmonically weighted sum of the restricted residual history plus a Type-II intercept term, and refitting removes most of the resulting wedge; the loss difference is locally quadratic, with a curvature corrected for the second derivative of the forecast map. On 31 harmonised inflation series over 2000–2025, across 40 configurations and repeated on core inflation, cross-sectional median loss ratios stay within a few per cent of unity while unit-level ratios span 0.63 to 1.53. A pooled model confidence set never separates the pair. The memory estimator moves the verdict further: on core inflation the fractional variant wins a majority of units in 1 of 20 configurations under one plug-in and 17 of 20 under the other, three quarters of that gap being bandwidth; estimated jointly with the autoregression, the memory parameter centres on zero and the median loss ratio exceeds one in every configuration.

KEYWORDS: Clark–West test, long memory, model confidence set, plug-in estimation, specification sensitivity.

1. INTRODUCTION

Long-memory models (Granger and Joyeux, 1980, Hosking, 1981, Beran, 1994, Baillie, 1996) occupy a settled place in the inflation literature. Hassler and Wolters (1995)

Calin-Adrian Comes: calin.comes@umfst.ro

The author thanks seminar participants for helpful comments. No specific funding was received for this work.

document fractional integration in international inflation rates, Baillie, Chung, and Tieslau (1996) embed it in an ARFIMA–GARCH specification, Bos, Franses, and Ooms (1999) weigh it against level shifts, and Gadea and Mayoral (2006) and Kumar and Okimoto (2007) use it to measure the persistence of inflation across countries and over time; persistence is a recurring theme of the inflation-forecasting literature as well (Stock and Watson, 2007, Groen, Paap, and Ravazzolo, 2013, Faust and Wright, 2013). The practical question is narrower: when the goal is to predict next month's or next year's inflation, does replacing integer differencing by a continuous order estimated from the data deliver a better forecast? We answer it for a narrow and exactly nested comparison. The loss difference, whatever its sign, is smaller than the effect of three decisions usually treated as technical detail: how seasonality is removed, where the rolling evaluation begins, and which semiparametric estimator supplies the plug-in memory parameter.

The forecasting value of fractional models has been assessed before, and the answer depends on the setting. In simulations, Crato and Ray (1996) find that simple ARMA models forecast competitively and that only long and strongly persistent series justify an ARFIMA model, in line with the ARMA approximations of long memory studied by Basak, Chan, and Palma (2001), whereas Vera-Valdés (2020) finds that ARFIMA forecasts dominate whatever mechanism generates the memory. On data, Bhardwaj and Swanson (2006) report that ARFIMA models sometimes forecast macroeconomic and financial series significantly better than short-memory alternatives, and Hassler and Pohle (2023) find methods based on fractional integration superior to alternatives that ignore long memory, in applications that include inflation. Baillie, Kongcharoen, and Kapetanios (2012) show that the estimation route matters, multi-step predictions built on a two-step local-Whittle plug-in having larger mean squared error than predictions built on maximum likelihood, and Baillie, Kapetanios, and Papailias (2014) choose the semiparametric bandwidth by out-of-sample forecast performance. Our comparison differs in two respects: the pair is exactly nested and the fractional layer is a plug-in on the same autoregression, so what is measured is the incremental value of adding that layer to the same forecasting procedure; and the evaluation is a design grid rather than a single specification, so that increment can be measured against the variation produced by ordinary evaluation choices.

Those choices have their own literature. Out-of-sample comparisons are known to depend on the estimation window (Pesaran and Timmermann, 2007, Rossi and Inoue, 2012,

Inoue, Jin, and Rossi, 2017) and on the test and reference distribution used (Diebold and Mariano, 1995, Clark and McCracken, 2001, Giacomini and White, 2006, Diebold, 2015), and parameter estimation error enters their limiting distributions (West, 1996). We vary two of these choices, the first origin of the evaluation and the reference distribution, together with the seasonal adjustment and the plug-in estimator, treat all four as design margins, and measure the operator effect against them.

Section 3 gives the reason the operator effect is small. The wedge between the two forecasts is linear in the plug-in memory to first order; with the autoregressive coefficients held fixed its coefficient is exactly a harmonically weighted sum of the restricted model's past residuals plus a Type-II intercept term, so that, under summable residual autocovariances, the projection of the next error on it is a harmonically weighted sum of residual autocorrelations, zero when the restricted residual is covariance white; refitting the coefficients on the filtered series removes most of the resulting wedge, the intercept term being the larger of the two fixed-coefficient components in these data. The loss difference is quadratic to second order, with one root at zero and a second at twice a pseudo-optimal value γ that the data determine. Where the curvature is positive the fractional layer helps between the two roots and hurts outside them, provided the plug-in is far enough from the second root for the quadratic to dominate the neglected cubic term, and two standard estimators at their conventional bandwidths can land on opposite sides of that root.

The empirical side is a design grid: 31 harmonised inflation series over 2000–2025 under 40 configurations—two memory estimators, five rolling-origin rules, four horizons—repeated on core inflation, for 80 in total. The cross-sectional median ratio of mean squared forecast errors stays within a few per cent of unity throughout, while unit-level ratios span 0.63 to 1.53, with roughly a quarter of cells outside five per cent either way. A model confidence set (Hansen, Lunde, and Nason, 2011) over five procedures, pooled by calendar date on a balanced cross-section, keeps the autoregression and its fractional extension together in all sixteen pooled cells and in 142 of 144 configurations obtained by varying the loss normalisation and the block-length rule. The memory estimator moves the verdict much further: on core inflation the fractional variant attains the lower loss in a majority of units in 1 of 20 configurations under the log-periodogram plug-in and in 17 of 20 under the local-Whittle plug-in, and the sign of the small median difference goes with it. The two plug-ins differ in their semiparametric objective and in bandwidth at once; crossing the two estima-

tors with the two conventional bandwidths (Section 6.4) attributes about three quarters of the gap to bandwidth, which shifts the median memory estimate by roughly 0.18 against 0.05 for switching estimator. Estimating the memory parameter jointly with the autoregression by conditional sum of squares, the route Baillie, Kongcharoen, and Kapetanios (2012) recommend over a two-step plug-in, centres the estimate on zero (-0.015 in the median, unit-level estimates spread over $[-0.48, 0.21]$) and puts the cross-sectional median loss ratio between 1.005 and 1.010 in every configuration: the autoregression's own one-step fit sees no memory left to estimate, and the persistence the plug-ins read as 0.2 to 0.5 is a property of the retained frequency band.

Section 6.5 asks whether the theory explains that sensitivity, using the first two derivatives of the forecast map, estimated numerically, together with the restricted model's forecast errors. In sample the sign condition classifies the realised outcome with 19 to 29 percentage points of skill over the base rate at the smaller plug-in and 2 to 18 at the larger, the curvature term being what makes it work at the larger; with the endpoint estimated on the first half of the origins and the rule evaluated on the second, that skill disappears. The interval $(0, 2\gamma)$ does not reliably separate the two plug-ins, and in a fifth of cells the curvature is negative so no such interval exists.

A last design margin is inferential. The same Clark–West statistic (Clark and West, 2007), on the same forecast errors, is referred to four reference distributions: the asymptotic normal, a moving-block bootstrap, and two bootstraps preserving the segmented second-order structure of the loss differential. Across the 80 configurations the asymptotic reference yields at least one unit significant after false-discovery-rate control in 24 and the spectral reference in one; on the unadjusted loss differential the counts are one and none, so on this window the discovery count is the Clark–West adjustment itself. A Monte Carlo study under both plug-ins finds that the calibration of each reference depends on the plug-in: under the log-periodogram estimate the asymptotic reference rejects at roughly twice its nominal rate at the annual horizon while the break-preserving bootstraps hold, and under local Whittle that distortion is absent; none of the four has appreciable power at the magnitudes the data deliver.

The contribution is threefold: a local theory for why a plug-in fractional layer moves forecasts as little as it does, and for what the Clark–West adjustment measures when the nesting parameter is a plug-in (Section 3); a benchmark wide enough to measure the op-

erator effect against the design effect (Sections 4–6); and a diagnostic of the mechanism against numerically estimated derivatives (Section 6.5). Section 8 repeats the exercise on core inflation and Section 9 concludes.

2. SETTING AND NOTATION

2.1. *Series and models*

Let $y_{i,t}$ denote monthly annualised inflation for reporting unit $i = 1, \dots, N$ at date $t = 1, \dots, T$, computed as 1200 times the first difference of the logarithm of a price index. Units are clipped to a common calendar window, but their observed support within it still differs, so T_i depends on i; the subscript is suppressed wherever a statement holds unit by unit. Where a fixed cross-section is needed, as in Section 6.3, it is imposed explicitly.

Two procedures are compared, both built on the same autoregressive order, selected by the corrected Akaike criterion at each order-selection origin, and both on a seasonally adjusted history. Write $\mathcal{S}_t$ for the operator that subtracts month-of-year factors estimated on observations available at t and adds the factor for the target month back to the forecast (Section 5). The restricted procedure is

$$\mathcal{M}_0 : \qquad \phi(L)\,\mathcal{S}_t y_s = c + \varepsilon_s, \qquad s \le t, \tag{1}$$

and the unrestricted procedure applies the same order p to the fractionally differenced series and re-integrates:

$$\mathcal{M}_1(d) : \qquad w_s = (1-L)^d \mathcal{S}_t y_s, \quad \phi(L)\, w_s = c + \varepsilon_s, \quad \hat{y} = (1-L)^{-d}\hat{w}. \tag{2}$$

The binomial weights are $\pi_0(d) = 1$, $\pi_k(d) = \pi_{k-1}(d)(k-1-d)/k$ for the difference and $\psi_k(d) = \psi_{k-1}(d)(k-1+d)/k$ for the inverse (Granger and Joyeux, 1980, Hosking, 1981). At $d = 0$ all weights with $k \ge 1$ vanish, so $\mathcal{M}_1(0) = \mathcal{M}_0$ identically: the pair is nested by construction and the restriction is a single continuous parameter.

2.2. *Plug-in memory*

The memory parameter is not estimated jointly with (ϕ, c); it is supplied by a semiparametric estimator computed on the same information set and inserted. We use two, at their

conventional bandwidths:

$$\hat{d}_{\mathrm{G}} : \text{log-periodogram regression (Geweke and Porter-Hudak, 1983)}, \quad m_{\mathrm{G}} = \lfloor T^{1/2} \rfloor, \tag{3}$$

$$\hat{d}_{\mathrm{W}} : \text{local-Whittle estimator (Robinson, 1995)}, \quad m_{\mathrm{W}} = \lfloor T^{0.65} \rfloor. \tag{4}$$

The second is also known as the Gaussian semiparametric estimator. Both are consistent for the same population memory under their respective regularity conditions, but at finite T they differ in bias through the short-run dynamics contaminating the retained frequency band (Hurvich, Deo, and Brodsky, 1998, Shimotsu and Phillips, 2005).

2.3. *Evaluation design*

Forecasts are generated under an expanding-window, rolling-origin scheme (Tashman, 2000). Let $\kappa \in (0, 1)$ index the origin rule: the first origin is $\max\{R_{\min}, \lfloor \kappa T \rfloor\}$ with $R_{\min} = 120$ months, and the last is $T - h$, giving $P(\kappa)$ origins collected in $\mathcal{O}(\kappa)$. For each such origin and horizon h let $e^{(0)}_{t+h}$ and $e^{(1)}_{t+h}$ be the h-step errors of $\mathcal{M}_0$ and $\mathcal{M}_1(\hat{d})$. The loss differential is the nesting-aware adjustment of Clark and West (2007),

$$\Delta_{t+h} = \left(e^{(0)}_{t+h}\right)^2 - \left[\left(e^{(1)}_{t+h}\right)^2 - \left(\hat{y}^{(0)}_{t+h} - \hat{y}^{(1)}_{t+h}\right)^2\right], \tag{5}$$

studentised by a Bartlett long-run variance at the fixed truncation $L = h - 1$ (Newey and West, 1987). We write $\bar{\Delta}(\kappa) = P(\kappa)^{-1} \sum \Delta_{t+h}$ and ω^2_L for that fixed-lag estimand, reserving ω^2_∞ for the untruncated long-run variance; the gap between the two is the subject of Section 7.

DEFINITION 1—Design cube: Let $\mathfrak{D} = \mathcal{S} \times \mathcal{K} \times \mathcal{E}$ denote the set of analyst choices held fixed within a run but varied across runs: the seasonality treatment $\mathcal{S}$, the origin rule $\kappa \in \mathcal{K}$, and the memory estimator $\mathcal{E} = \{\hat{d}_{\mathrm{G}}, \hat{d}_{\mathrm{W}}\}$. For a scalar summary τ of the comparison (a loss ratio, a rejection count) the *design range* of τ is $\mathrm{range}_{\mathfrak{D}}(\tau) = \max_{\mathfrak{D}} \tau - \min_{\mathfrak{D}} \tau$, and the *operator effect* is $|\tau - \tau_0|$ evaluated at a single design point, where τ_0 is the value under the restriction $d = 0$.

Proposition S3 of the Appendix gives the condition under which the design range exceeds the operator effect.

3. THE FORECAST WEDGE AND ITS SIGN

The results of this section are local expansions around $d = 0$, not limit theorems; they are informative because the plug-in estimates met in practice are of order 0.2 to 0.5 rather than of order one.

ASSUMPTION D1: *(Moments and design.)* The seasonally adjusted history $\{\mathcal{S}_t y_s\}_{s \le t}$ has finite fourth moments and $\sum_{k \ge 1} k^{-1} |\mathcal{S}_t y_{t-k}| = O_p(\log t)$. At every origin the regressor matrix of the autoregression fitted to the filtered series $w(d)$ has full column rank for every d in a neighbourhood of zero, almost surely.

ASSUMPTION D2: *(The procedure as implemented.)* At each origin the two members share the autoregressive order p, selected once on the restricted series by the corrected Akaike criterion over $p \le 13$; each member's coefficients are estimated on its own series, so the unrestricted member is refitted on w. No moving-average terms are estimated anywhere in this paper. Both binomial expansions in (2) are of Type II: the backward filter runs over the observed history from its first observation, and the re-integration runs over the whole extended sample, observed values and forecasts, both with zero pre-sample values.

ASSUMPTION D3: *(The plug-in.)* $\hat{d} = O_p(1)$ and is measurable with respect to the information set at the origin.

Assumption D2 describes the procedure as implemented; two of its features shape the first-order effect, the Type-II truncation, which acts through the intercept, and the refitting of the larger model, which attenuates the wedge. Lemma S1 of the Appendix gives the expansion of the binomial weights near zero, $\pi_k(d) = -d/k + O(d^2)$ and $\psi_k(d) = +d/k + O(d^2)$ for every $k \ge 1$, uniformly on compact neighbourhoods; the harmonic weights $1/k$ below come from there.

LEMMA 1—Sensitivity at the nesting point: *Under Assumptions D1–D3 the map* $d \mapsto \hat{y}^{(1)}_{t+h}(d)$ *is twice continuously differentiable on a neighbourhood of the origin, with*

$\hat{y}^{(1)}_{t+h}(0) = \hat{y}^{(0)}_{t+h}$, *and its first derivative admits the decomposition*

$$\mathcal{G}_{t,h} := \left.\frac{\partial \hat{y}^{(1)}_{t+h}(d)}{\partial d}\right|_{d=0} = \mathcal{G}^{\text{fix}}_{t,h} + \mathcal{G}^{\text{es}}_{t,h}, \tag{6}$$

where $\mathcal{G}^{\text{fix}}$ *is the derivative of the forecast map with the intercept and slopes of the autoregression held at their restricted values, so that only the two binomial filters act, and* $\mathcal{G}^{\text{es}}$ *collects the terms generated by re-estimating those coefficients on the filtered series. Both are* $O_p(\log t)$*, and both are explicit:* $\mathcal{G}^{\text{fix}}$ *in Proposition 1,* $\mathcal{G}^{\text{es}}$ *as the product of the forecast's gradient in the coefficients and the derivative of the least squares coefficients with respect to the filter (Section S1.5 of the Appendix), and their sum reproduces the finite-difference derivative used in Section 6.5 to* $4e-04$ *at every origin. The two parts offset: refitting removes between three quarters and* 85 *per cent of the fixed-coefficient wedge.*

PROPOSITION 1—First-order forecast wedge: *Under Assumptions D1–D3,*

$$\hat{y}^{(1)}_{t+h}(d) - \hat{y}^{(0)}_{t+h} = d\,\mathcal{G}_{t,h} + O_p(d^2), \tag{7}$$

with $\mathcal{G}_{t,h}$ *as in (6). Let* c *and* $\phi_1, \ldots, \phi_p$ *be the intercept and slopes of the restricted autoregression at the origin,* θ_j *its impulse responses (*$\theta_0 = 1$*,* $\theta_j = \sum_{i=1}^{\min(p,j)} \phi_i \theta_{j-i}$*), and* $\hat{\varepsilon}_s = x_s - c - \sum_{i=1}^{p} \phi_i x_{s-i}$ *its one-step residuals on the adjusted history* $x_s = \mathcal{S}_t y_s$*, with* $x_r = 0$ *for* $r < 1$*. The fixed-coefficient component has the explicit form*

$$\mathcal{G}^{\text{fix}}_{t,h} = \sum_{j=1}^{h} \theta_{h-j} \Big(\sum_{k=j}^{t+j-1} \frac{\hat{\varepsilon}_{t+j-k}}{k} + c\, H_{t+j-1} \Big), \qquad \text{so that} \qquad \mathcal{G}^{\text{fix}}_{t,1} = \sum_{k=1}^{t} \frac{\hat{\varepsilon}_{t+1-k}}{k} + c\, H_t, \tag{8}$$

with $H_s = \sum_{k=1}^{s} k^{-1}$ *the harmonic number. It is the first-order term of an exact identity: with the coefficients held fixed,*

$$\hat{y}^{(1)}_{t+h}(d) - \hat{y}^{(0)}_{t+h} = \sum_{j=1}^{h} \theta_{h-j}\, U_{t+j}(d), \qquad U_{t+j}(d) = -\sum_{k=1}^{t+j-1} \pi_k(d) \big(U_{t+j-k} + c \big), \tag{9}$$

where $U_s = \hat{\varepsilon}_s$ *for* $s \le t$*: the unrestricted forecast is the restricted recursion driven by pseudo-innovations that are binomially weighted sums of the restricted residuals.*

REMARK 1—Where the first-order effect comes from: The two parts of $\mathcal{G}_{t,1}^{\mathrm{fix}}$ have different origins. The residual part $\sum_k \hat{\varepsilon}_{t+1-k}/k$ is the memory operator itself: the autoregressive recursion applied to the filtered series differs from the filtered recursion by exactly $-\sum_k \pi_k(d)\hat{\varepsilon}_{t+1-k}$, and this does not vanish without truncation: under summable residual autocovariances, the condition of Section S1.4 of the Appendix, $\sum_{k\geq 1} \varepsilon_{t+1-k}/k$ converges in mean square, to a limit that is degenerate only if the residuals are. The intercept part cH_t is a truncation artefact: the Type-II filter of a constant does not vanish on a finite sample, its first-order coefficient grows like $\log t$, which is where the rate in Lemma 1 comes from, and refitting the intercept on the filtered series removes it. What the operator contributes to the sign of the comparison is the residual part; in these data the intercept part is the larger of the two, with median absolute sizes of 1.41 and 0.76 standard deviations of the adjusted series against 0.21 for the refitted sensitivity (Section S3 of the Appendix). With fixed coefficients, $h = 1$ and the same condition, $B = \mathbb{E}[e^{(0)}\mathcal{G}] = \sum_{k\geq 1} \mathrm{Cov}(\hat{\varepsilon}_{t+1}, \hat{\varepsilon}_{t+1-k})/k$, a harmonically weighted sum of the residual autocovariances of the restricted model, which is zero when that residual is covariance white and equals $\sigma^2 d_0 \pi^2/6 + O(d_0^2)$ when it is fractional noise of order d_0, so that the projection coefficient $B/\mathbb{E}[\mathcal{G}^2]$ is then $d_0 + O(d_0^2)$ (Section S1.4 of the Appendix, which also states the martingale-difference condition under which the curvature is then $\mathbb{E}[\mathcal{G}^2]$). Table S5 of the Appendix verifies the identity (9) numerically and finds the refitted wedge about a third of the fixed-coefficient one.

Because $\mathcal{G}_{t,h}$ is defined as a derivative rather than through a closed form, what follows does not depend on the decomposition in (6).

PROPOSITION 2—Loss differential and the sign condition: *Let $\mathcal{H}_{t,h} := \partial^2 \hat{y}_{t+h}^{(1)}(d)/\partial d^2$ at $d = 0$, which exists by Lemma 1, and suppose in addition that $\hat{y}_{t+h}^{(1)}$ is three times continuously differentiable on a neighbourhood $\mathcal{N} = [-\delta_0, \delta_0]$ of the origin and that $e_{t+h}^{(0)}$, $\mathcal{G}_{t,h}$, $\mathcal{H}_{t,h}$ and $\sup_{d\in\mathcal{N}} |\partial^3 \hat{y}_{t+h}^{(1)}(d)/\partial d^3|$ have finite second moments. Then, with $\mathrm{MSFE}_j = \mathbb{E}(e^{(j)})^2$, there is $C < \infty$ with*

$$\mathrm{MSFE}_1(d) - \mathrm{MSFE}_0 = Q(d) + R(d), \qquad Q(d) := A\, d(d - 2\gamma), \qquad |R(d)| \leq C|d|^3 \tag{10}$$

for $|d| \leq \delta_0$, where $B := \mathbb{E}[e^{(0)}\mathcal{G}]$, $A := \mathbb{E}[\mathcal{G}^2 - e^{(0)}\mathcal{H}]$ and, when $A \neq 0$, $\gamma := B/A$. The curvature is A rather than $\mathbb{E}[\mathcal{G}^2]$, because the second derivative of the forecast map enters

through its covariance with the restricted error. If $A > 0$ *and* $\gamma \neq 0$ *then* Q *is negative exactly between its two roots,*

$$Q(d) < 0 \quad \Longleftrightarrow \quad d \in (0,\, 2\gamma) \ \textit{for}\ \gamma > 0, \qquad d \in (2\gamma,\, 0) \ \textit{for}\ \gamma < 0, \tag{11}$$

and Q *is minimised at* $d = \gamma$ *with value* $-A\gamma^2$*. The sign of* Q *carries over to the loss difference itself wherever the quadratic exceeds the cubic bound, that is wherever*

$$A\,|d - 2\gamma| \;>\; C\,d^2, \qquad 0 < |d| \leq \delta_0. \tag{12}$$

On that set $\mathrm{MSFE}_1(d) < \mathrm{MSFE}_0$ *between the roots and* $\mathrm{MSFE}_1(d) > \mathrm{MSFE}_0$ *outside their closed interval. If* $A \leq 0$ *the second-order part is concave, no interior minimum exists, and* (11) *does not apply.*

In words, (11) says that whether a plug-in fractional layer improves on its short-memory parent does not depend on whether the series has long memory, but on whether the number plugged in falls between zero and twice a pseudo-optimal memory that the restricted model's errors and the wedge determine jointly. Condition (12) carries that reading from Q to the realised loss, and the two roots behave differently under it: at $d = 0$ the quadratic crosses zero with slope $-2A\gamma$ and dominates the cubic remainder for all small enough d, while at $d = 2\gamma$ the remainder bound stays near $C(2\gamma)^3$, so close to the second root the expansion determines nothing. Section 6.5 measures where in the observed range of plug-ins the condition holds.

COROLLARY 1—Estimator-induced divergence: *Suppose* $A > 0$ *and* $\gamma > 0$*, and let* $\hat{d}_{\mathrm{G}} \to_p d_{\mathrm{G}}$*,* $\hat{d}_{\mathrm{W}} \to_p d_{\mathrm{W}}$ *with* $0 < d_{\mathrm{W}} < 2\gamma < d_{\mathrm{G}} \leq \delta_0$*, both limits satisfying the margin condition* (12) *strictly. Then, on the same data, the same models and the same evaluation sample,*

$$\mathbb{P}\big\{\mathrm{MSFE}_1(\hat{d}_{\mathrm{W}}) < \mathrm{MSFE}_0 < \mathrm{MSFE}_1(\hat{d}_{\mathrm{G}})\big\} \longrightarrow 1.$$

The conclusion is probabilistic because the plug-ins are random; the hypotheses fix where their limits *sit relative to the roots. Under those hypotheses the two plug-ins deliver opposite verdicts on the fractional layer while estimating a parameter that is not itself the object of interest.*

The margin condition is restrictive. Section 6.5 finds the quadratic accurate in magnitude only up to about $d = 0.2$, while the log-periodogram plug-in has a median of 0.57 on core inflation, so the corollary describes an outcome the local expansion permits rather than one it establishes for our data; Section 6.5 checks the sign condition against the data instead. The two plug-ins also differ in more than bandwidth: log-periodogram regression and the local-Whittle objective are different estimators, used at $m = \lfloor T^{1/2} \rfloor$ and $m = \lfloor T^{0.65} \rfloor$, and Section 6.4 separates the two differences.

The expansion also says what the two loss differentials used for inference measure when the nesting parameter is a plug-in rather than a coefficient estimated on the evaluation model. Write $f^{\mathrm{DM}} = (e^{(0)})^2 - (e^{(1)})^2$ for the unadjusted differential of Diebold and Mariano (1995) and $f^{\mathrm{CW}} = f^{\mathrm{DM}} + (\hat{y}^{(0)} - \hat{y}^{(1)})^2$ for the adjusted one of (5).

PROPOSITION 3—The adjusted and unadjusted differentials under a plug-in: *Under the conditions of Proposition 2, for* $|d| \le \delta_0$,

$$\mathbb{E}\big[f^{\mathrm{DM}}(d)\big] = 2dB - d^2A + O(d^3), \qquad \mathbb{E}\big[f^{\mathrm{CW}}(d)\big] = 2dB + d^2\,\mathbb{E}[e^{(0)}\mathcal{H}] + O(d^3), \tag{13}$$

so that $\mathbb{E}[f^{\mathrm{CW}}] - \mathbb{E}[f^{\mathrm{DM}}] = d^2\mathbb{E}[\mathcal{G}^2] + O(d^3)$. *In particular, if* $B = 0$ *then* $\mathbb{E}[f^{\mathrm{DM}}] = -d^2A < 0$ *whenever* $A > 0$, *while* $\mathbb{E}[f^{\mathrm{CW}}] = d^2\mathbb{E}[e^{(0)}\mathcal{H}]$ *has no determinate sign; and if the plug-in converges to a limit* $d^* \neq 0$ *as the estimation sample grows, the gap* $d^{*2}\mathbb{E}[\mathcal{G}^2]$ *between the two population differentials does not close.*

The adjustment of Clark and West (2007) adds back the cost of estimating the nesting parameter, $d^2\mathbb{E}[\mathcal{G}^2]$, which in their setting is the sampling noise of a coefficient estimated on the evaluation model and vanishes with the estimation sample. A semiparametric plug-in converges instead to whatever persistence the retained frequency band contains, which need not be zero when the autoregression forecasts as well as its extension, so the two nulls, "the population forecasts coincide" and "the population losses coincide", stay $d^{*2}\mathbb{E}[\mathcal{G}^2]$ apart, and a rejection by the adjusted statistic says that the plug-in is not zero, not that the layer forecasts better; Section 7 measures the consequence.

Three further results are stated and proved in the Appendix, which also collects the proofs of the results above. Proposition S2 concerns the evaluation sample: when the loss differential has different means over sub-periods of the evaluation span, the mean differential is a weighted average of those means with weights that depend on the origin rule

κ, so its sign can change as κ moves, with no change in the models, the data, the statistic or the reference distribution; Section 6.4 shows this in the data. Proposition S3 gives the design range of Definition 1 its content: when the loss ratios at the two plug-in limits lie on opposite sides of unity, the range induced by the choice of estimator equals the sum of the two operator effects and so exceeds either, which happens in six of the eight rows of Table I. Proposition S1 concerns the horizon. With Lemma 1 in closed form the long-horizon behaviour of the sensitivity is explicit: on a demeaned history it converges to the intercept channel of the refit, $\dot{c}/(1-\sum_i \phi_i)$, and on a level history with an intercept it drifts like $\mu \log((t+h)/t)$, because the Type-II re-integration of a constant grows with the horizon; in both cases the projection coefficient $B_h/\mathbb{E}[\mathcal{G}_{t,h}^2]$ tends to zero under a mixing condition, so the interval of improvement should collapse. It does not at the horizons available, where $\log((t+h)/t) \leq 0.06$ and the drift is below a tenth of a percentage point (Section 6.5 and Section S7 of the Appendix), and nothing below builds on it.

4. DATA

The empirical analysis uses the Harmonised Index of Consumer Prices compiled by Eurostat, monthly index series, table `prc_hicp_midx`, in the $2015 = 100$ unit (Eurostat, 2026): the all-items index (`CP00`) for the headline analysis and, in Section 8, the aggregate excluding energy and unprocessed food (`TOT_X_NRG_FOOD`). The archive snapshot covers 1996M01–2025M12 for 38 reporting territories.

Three rules define the estimation sample. The *calendar rule* clips every unit to 2000M01–2025M12. It does not make the observed support identical: within the window the United Kingdom's series ends in 2020M11 and the United States' begins in 2002M01, while the others run throughout; every statistic except the pooled confidence set of Section 6.3 is computed unit by unit and then aggregated, so the difference in support does not affect it. The *inclusion rule* retains a unit with at least 200 monthly observations in the window and no annualised monthly change larger than 100 per cent in absolute value, the second criterion removing episodes whose isolated spikes would dominate the periodogram, and therefore the plug-in, as well as the realised loss; once the calendar rule is applied the first criterion excludes Albania, Kosovo and Montenegro and the second only Turkey, and the two rules retain 34 headline units. The *common-sample rule* produces the number used throughout: because Section 6.4 compares the full window with a pre-2020

sub-window, all reported results use the intersection of the two eligible panels, which drops Switzerland, North Macedonia and Serbia and leaves 31 units on headline and 30 on core.

The 31 headline units are AT, BE, BG, CY, CZ, DE, DK, EE, EL, ES, FI, FR, HR, HU, IE, IS, IT, LT, LU, LV, MT, NL, NO, PL, PT, RO, SE, SI, SK, UK and US. The United States belongs to Eurostat's harmonised comparison set and is retained on headline; the core panel excludes it. The audit of every retained and excluded unit, for every window, is in the replication archive.

5. DESIGN

5.1. *Recursive seasonal adjustment*

The harmonised index is not seasonally adjusted, and leaving it so has a measurable cost: with an unadjusted target the selected autoregressive order sat at or one below the cap $p \leq 13$ at some origin in 31 of the 32 units of the unadjusted baseline configuration, and had a median there in 29, the lag polynomial being spent on the annual cycle. At each origin t we therefore estimate twelve month-of-year factors by least squares on $\{y_s\}_{s \leq t}$ only, subtract them from the history, run both members on the adjusted series, and add the factor for the target month back before computing the loss. No observation dated after the origin enters and both members receive an identical add-back, so the nesting at $d = 0$ is preserved exactly, and the plug-in is estimated on the adjusted history, which keeps the seasonal peak out of the retained frequency band.

5.2. *Configuration grid and benchmarks*

The full grid is $\mathcal{E} \times \mathcal{K} \times \mathcal{H}$ with $\mathcal{E} = \{\hat{d}_{\mathrm{G}}, \hat{d}_{\mathrm{W}}\}$, $\mathcal{K} = \{0.40, 0.50, 0.60, 0.70, 0.80\}$ and $\mathcal{H} = \{1, 3, 6, 12\}$: forty configurations per price index, eighty in all, and within each configuration every unit contributes between 45 and 187 origins. Alongside the nested pair we evaluate three rules requiring no estimation, on the identical origins: the random walk $\hat{y}_{t+h} = y_t$, the Atkeson and Ohanian (2001) rule $\hat{y}_{t+h} = 12^{-1} \sum_{j=0}^{11} y_{t-j}$, and the seasonal naive rule $\hat{y}_{t+h} = y_{t+h-12}$. They show that the fitted models are worth comparing at all.

5.3. *Reference distributions*

The Clark–West statistic of (5) is referred to four references: the asymptotic standard normal; a moving-block bootstrap of the differential (Künsch, 1989); a break-preserving parametric bootstrap that segments the mean of the differential, fits an autoregressive sieve to the within-segment demeaned series and simulates with segment-specific innovation variances under the null; and a spectral variant that reproduces the estimated autocovariance structure rather than an autoregressive approximation to it. In all bootstraps the null is imposed in population rather than by centring each replicate. Multiplicity across units is controlled within configuration by the Benjamini and Hochberg (1995) procedure.

6. RESULTS

6.1. *The memory estimates*

Taking the median over all unit–configuration cells, the plug-in is about 0.44 under $\hat{d}_{\mathrm{G}}$ and 0.22 under $\hat{d}_{\mathrm{W}}$ on headline inflation, and about 0.57 and 0.23 on core. The estimates are not uniformly inside the covariance-stationary region $d < 1/2$: under $\hat{d}_{\mathrm{G}}$, 38 per cent of headline cells and 59 per cent of core cells have medians at or above 0.5, against 7 and 8 per cent under $\hat{d}_{\mathrm{W}}$, so under the log-periodogram plug-in the core panel sits mostly in the non-stationary but mean-reverting region. Both estimators remain usable there: the local-Whittle estimator is consistent for $d < 1$ and asymptotically normal for $d < 3/4$ (Velasco, 1999a), and log-periodogram regression has the same properties (Velasco, 1999b). The two estimators disagree systematically on identical series, by roughly a factor of two in the median, and under Proposition 2 that disagreement can decide the sign of the comparison.

Neither plug-in should be read as an estimate of a stable fractional-integration parameter. Occasional level shifts generate autocorrelation functions that semiparametric estimators read as long memory (Diebold and Inoue, 2001, Granger and Hyung, 2004, Perron and Qu, 2010), inflation over 2000–2025 contains at least one such shift in every unit, and Bos, Franses, and Ooms (1999) document the same confounding for inflation rates. We treat $\hat{d}_{\mathrm{G}}$ and $\hat{d}_{\mathrm{W}}$ as plug-in persistence summaries that the forecasting procedure consumes, and Proposition 2 is stated in terms of whatever number is plugged in.

6.2. *Accuracy*

Table I reports the central accuracy comparison: the ratio $\mathrm{MSFE}_1 \,/\, \mathrm{MSFE}_0$ within the nested pair, and the restricted model against three estimation-free benchmarks on the identical origins.

TABLE I

FORECAST ACCURACY. COLUMNS 3–6: THE RATIO OF MEAN SQUARED FORECAST ERRORS WITHIN THE NESTED PAIR, $\mathrm{MSFE}_1 \,/\, \mathrm{MSFE}_0$, BY PLUG-IN ESTIMATOR. EACH ENTRY IS THE MEDIAN OVER THE FIVE ROLLING-ORIGIN RULES $\kappa \in \{0.40, \ldots, 0.80\}$ OF THE PER-CONFIGURATION CROSS-SECTIONAL MEDIAN, AND BRACKETS GIVE THE RANGE SPANNED BY THOSE FIVE RULES. A VALUE BELOW ONE FAVOURS THE FRACTIONAL VARIANT. COLUMNS 7–9: THE RESTRICTED MODEL AGAINST THE RANDOM WALK, THE ATKESON–OHANIAN RULE AND THE SEASONAL NAIVE RULE, EACH THE MEDIAN OVER ALL UNIT–CONFIGURATION CELLS AT THAT HORIZON. A VALUE BELOW ONE FAVOURS THE RESTRICTED MODEL.

		$\mathrm{MSFE}_1 \,/\, \mathrm{MSFE}_0$, $\hat{d}_{\mathrm{G}}$		$\mathrm{MSFE}_1 \,/\, \mathrm{MSFE}_0$, $\hat{d}_{\mathrm{W}}$		MSFE_0 relative to		
Index	h	med.	range	med.	range	RW	AO	S-nv
Headline	1	1.004	[0.995, 1.011]	0.991	[0.989, 0.993]	0.41	0.69	0.72
	3	1.000	[0.995, 1.016]	0.992	[0.989, 0.992]	0.40	0.73	0.74
	6	1.011	[1.003, 1.025]	0.996	[0.993, 0.998]	0.52	0.71	0.76
	12	1.017	[1.007, 1.027]	1.004	[1.003, 1.007]	0.82	0.67	0.82
Core	1	1.013	[1.010, 1.022]	0.996	[0.989, 0.998]	0.19	0.42	0.84
	3	1.016	[1.011, 1.021]	0.998	[0.991, 1.000]	0.20	0.42	0.84
	6	1.008	[1.004, 1.014]	0.992	[0.981, 0.993]	0.33	0.41	0.86
	12	1.003	[0.999, 1.007]	0.995	[0.993, 1.000]	0.93	0.41	0.93

The cross-sectional median effect is small: no median in columns three to six leaves $[0.989, 1.017]$, and no bracketed range leaves $[0.980, 1.027]$. The two plug-in columns sit on opposite sides of unity almost throughout: under $\hat{d}_{\mathrm{G}}$ the median exceeds one in seven of the eight rows, under $\hat{d}_{\mathrm{W}}$ it falls below one in seven, and the distance between the columns is larger than the distance of either from unity. The benchmark columns show that this flatness is not that of a badly specified comparison: the restricted model attains 0.19 to 0.52 of the random walk's squared error at horizons up to six months and 0.41 to 0.73 of the Atkeson–Ohanian rule's throughout. At $h = 12$ the ratio against the random walk rises to

0.82 on headline and 0.93 on core, so at the annual horizon no fitted procedure reliably improves on a rule requiring no estimation and we report no substantive result there.

The median conceals a good deal of dispersion, which Table S12 sets out. It sits within about a tenth of a per cent of unity on both indices, yet more than half the headline cells lie outside a two per cent band, and the cells outside five per cent split almost evenly: on headline 14 per cent improve and 12 per cent worsen, on core 10 and 9. Country-level changes are often large in both directions and very nearly cancel in the median; what the design margins of Section 6.4 move is that aggregate summary, not the sign a single-country analyst will see. The 2021–23 surge, which the evaluation span at $\kappa = 0.70$ largely covers, shapes the level of these numbers but not their ordering. Tables S15 and S16 of the Appendix repeat Tables I and II on the window ending in 2019M12, whose evaluation span runs from 2014 to 2019. The operator effect is larger and one-sided under $\hat{d}_{\mathrm{G}}$ (median ratio 1.024 to 1.058; a majority of units in 2 of 20 configurations on headline and none on core) and within about one per cent under $\hat{d}_{\mathrm{W}}$ (15 and 1), so the estimator margin predates the surge; the restricted model no longer beats the seasonal naive rule (0.95 to 1.02 of its squared error on headline, 1.12 to 1.14 on core), and the pooled confidence set keeps the pair together in 7 of the 8 headline cells but eliminates both members in all 8 core cells, where the seasonal naive rule survives alone; and discoveries are more frequent, in 53 of the 76 configurations with at least 40 origins under the asymptotic reference and 29 under the spectral bootstrap, every spectral one in a unit where the fractional variant does attain the lower loss, so the shorter window widens the cross-section rather than moving its centre. The range column of Table S15 at $h = 12$ under $\hat{d}_{\mathrm{G}}$, $[0.981, 1.097]$, is the origin-rule reversal of Section 6.4.

6.3. *Model confidence set*

The model confidence set of Hansen, Lunde, and Nason (2011), which builds on the reality check of White (2000), asks which of the five procedures cannot be distinguished from the best once all are compared at once. We compute it under both plug-ins and pool the losses by calendar date rather than by unit, so that dependence between countries at a date stays inside the observation and the resampling runs over time, where a moving-block bootstrap is defensible. Pooling by date requires a fixed cross-section, which the panel does not supply, so we pool over the units sharing an identical set of target dates, taking

the set that maximises units times dates: 29 units in both indices, setting aside the United Kingdom and the United States on headline and the United Kingdom on core for this table only, on a pooled calendar from 2018M03 to 2025M12 at $h = 1$ and starting later as the horizon lengthens. The block length is the larger of the cube root of the series length and the horizon, there are $2{,}000$ replications at the ten per cent level, the origin rule is $\kappa = 0.70$ throughout the table, and the test uses the range statistic, eliminating at each step the model with the largest worst-case standardised loss difference. Elimination stops at the first test that does not reject, and every model still in the set is reported with that test's p-value, so the members of a surviving set share one p-value by construction; the studentised mean loss difference between the two members of the pair on the pooled series, with a two-sided p-value from the same block bootstrap, says how far they are from separation. Section S6 of the Appendix gives the construction in full.

Table II reports the result. The pooled set never separates the two members of the nested pair: in all sixteen cells they survive together, under either plug-in and at every horizon, with p from 0.14 to 0.96, and the studentised loss difference between them ranges from -1.42 to $+1.36$ (positive when the fractional variant has the lower loss) with bootstrap p-values no smaller than 0.14 (Figure S3 of the Appendix), while the estimation-free rules do separate: Atkeson–Ohanian is eliminated everywhere and seasonal naive at every horizon up to six months, surviving at $h = 12$ alongside the random walk, with which it coincides. Balancing is not what drives this: loosening the origin rule to $\kappa = 0.40$ roughly doubles the pooled series without separating the pair in any of the 32 resulting cells. Nor does the result depend on the loss normalisation or the block rule: under three normalisations by three block rules, all formed from identical loss series (Table S11 of the Appendix), the pair survives together in 142 of the 144 configurations, both exceptions occurring in the core cell at $h = 1$ under $\hat{d}_{\mathrm{G}}$ with the $2h$ rule, the shortest block of the three.

The pooled set therefore says that, across countries, the operator margin is too small for this selection procedure to resolve; it does not say the two procedures are equivalent. Unit by unit, where no balancing is needed, the set does sometimes separate the pair, and how often depends on the plug-in: under the log-periodogram estimate the fractional variant is excluded in 6 of 124 headline unit–horizon cells and 15 of 120 core cells, against one cell running the other way, while under local Whittle those counts fall to 3 and 4 and the cells

TABLE II

POOLED MODEL CONFIDENCE SET AT THE TEN PER CENT LEVEL, ON A BALANCED CALENDAR OF 29 UNITS. ENTRIES ARE MCS p-VALUES; SURVIVORS ARE THOSE WITH $p \geq 0.10$ AND ARE SHOWN IN BOLD. THE SMALLEST ATTAINABLE p-VALUE IS REPORTED AS <0.001. AT $h = 12$ THE SEASONAL-NAIVE AND RANDOM-WALK FORECASTS COINCIDE, SINCE THE TWELVE-MONTH LAG OF THE TARGET IS THE LAST TRAINING OBSERVATION, SO THEIR COLUMNS ARE IDENTICAL THERE.

Index	Plug-in	h	dates	AR(p)	frac. AR(p)	random walk	Atk.–Ohanian	seas. naive
Headline	$\hat{d}_{\mathrm{G}}$	1	94	**0.592**	**0.592**	<0.001	<0.001	0.043
		3	92	**0.592**	**0.592**	<0.001	<0.001	0.059
		6	89	**0.961**	**0.961**	<0.001	<0.001	0.085
		12	83	**0.201**	**0.201**	**0.201**	<0.001	**0.201**
	$\hat{d}_{\mathrm{W}}$	1	94	**0.365**	**0.365**	<0.001	<0.001	0.046
		3	92	**0.213**	**0.213**	<0.001	<0.001	0.060
		6	89	**0.329**	**0.329**	<0.001	<0.001	0.087
		12	83	**0.156**	**0.156**	**0.156**	<0.001	**0.156**
Core	$\hat{d}_{\mathrm{G}}$	1	90	**0.145**	**0.145**	<0.001	<0.001	0.005
		3	88	**0.317**	**0.317**	<0.001	<0.001	0.002
		6	85	**0.818**	**0.818**	<0.001	<0.001	0.006
		12	79	**0.317**	**0.317**	**0.317**	<0.001	**0.317**
	$\hat{d}_{\mathrm{W}}$	1	90	**0.452**	**0.452**	<0.001	<0.001	0.004
		3	88	**0.716**	**0.716**	<0.001	<0.001	0.003
		6	85	**0.167**	**0.167**	<0.001	<0.001	0.007
		12	79	**0.238**	**0.238**	**0.238**	<0.001	**0.238**

excluding the autoregression rise from one to three. Separation is the exception either way, and it is more common and more one-sided at the larger plug-in.

6.4. *Three design margins*

6.4.0.1. *Seasonality.* Holding the units, the window and the statistic fixed and switching only the seasonal treatment cuts the raw count of units significant under the asymptotic reference at $h = 1$ from seven to three, and the share of origins at the order cap from 43

to 28 per cent. Part of the apparent evidence in the unadjusted specification was the two members handling the annual cycle differently.

6.4.0.2. *Origin rule.* On the full window the origin rule is benign: sweeping κ from 0.80 to 0.40, which moves the number of origins from 63 to 187 at $h = 1$, leaves the median loss ratio inside the same narrow band. On a shortened pre-2020 window at $h = 12$ it is not benign at all. Under the log-periodogram plug-in, with $\kappa = 0.70$ there are 61 origins, the fractional variant wins in 21 of 31 units and the median ratio is 0.981; with $\kappa = 0.50$ there are 108 origins, it wins in 10 of 31 and the median is 1.097. Only the first origin changes. This is the case described by Proposition S2 of the Appendix, and it is why we report no result at $h = 12$ on short windows.

6.4.0.3. *Memory estimator.* This margin is the largest. Table III counts the configurations, out of twenty per estimator and index, in which the fractional variant attains the lower loss in a majority of units.

TABLE III

CONFIGURATIONS (OF TWENTY) IN WHICH THE FRACTIONAL VARIANT ATTAINS THE LOWER SQUARED FORECAST ERROR IN A MAJORITY OF UNITS. PLUG-IN MEDIANS ARE THE MEDIAN OVER ALL UNIT–CONFIGURATION CELLS OF THE PER-UNIT MEDIAN PLUG-IN.

Index	$\hat{d}_{\mathrm{G}}$	$\hat{d}_{\mathrm{W}}$	median $\hat{d}_{\mathrm{G}}$ / median $\hat{d}_{\mathrm{W}}$
Headline	5 of 20	15 of 20	0.447 / 0.220
Core	**1 of 20**	**17 of 20**	0.571 / 0.229

On core inflation the choice between two standard estimators of the same nuisance parameter moves the verdict from “the fractional layer helps almost nowhere” to “it helps almost everywhere”. They differ in finite-sample bias, and the operator effect is small enough that the gap between them changes the aggregate sign. Section 6.5 asks whether Corollary 1 explains that, with a partly positive answer.

The two plug-ins differ in their objective function and in the number of retained frequencies at once, so comparing them as ordinarily run cannot say which of the two is doing the work. Section S4 of the Appendix crosses the two estimators with the two bandwidth rules on the same units, origins and seasonal adjustment (Table S6) and finds that most of the

gap is bandwidth: between the two plug-ins as conventionally run the median plug-in falls from 0.45 to 0.22 at $h = 1$, of which moving only the bandwidth accounts for about 0.17 and moving only the estimator for 0.06; at $h = 12$ the split is 0.19 against 0.05, and the loss ratios and win counts move the same way by correspondingly small amounts. Bandwidth is therefore itself a design choice in the sense of Definition 1: nothing fixes the exponent at 0.50 rather than 0.65, data-driven rules exist (Henry, 2001, Andrews and Guggenberger, 2003, Andrews and Sun, 2004) and Baillie, Kapetanios, and Papailias (2014) choose the bandwidth by out-of-sample forecast performance, and moving the exponent within the conventional range moves the memory estimate by about three times what switching estimator does. Sweeping the exponent from 0.45 to 0.80 (Section S4, Figure S4 and Table S18 of the Appendix) shows both margins at once: on headline at $h = 1$ the median plug-in falls, not monotonically, from 0.41 to 0.24 under $\hat{d}_{\mathrm{G}}$ and from 0.39 to 0.13 under $\hat{d}_{\mathrm{W}}$, while at a common exponent the two estimators still differ by 0.02 to 0.14, about 0.10 in the median, and the median loss ratio moves between 0.990 and 1.006 under $\hat{d}_{\mathrm{G}}$ and 0.991 and 0.998 under $\hat{d}_{\mathrm{W}}$, at $h = 12$ between 1.001 and 1.084 under $\hat{d}_{\mathrm{G}}$. A rule that selects the exponent at each origin by the cumulative squared error of the fractional forecasts whose targets have been observed by that origin, in the spirit of Baillie, Kapetanios, and Papailias (2014), settles at an average exponent of 0.61 to 0.63 and lands on median ratios of 0.990 to 0.991 at $h = 1$ and 1.014 to 1.043 at $h = 12$: inside the range the fixed exponents span at $h = 1$, and in the upper part of it at $h = 12$, where the score lags the origin by a year; the rule buys nothing over a conventional exponent.

6.4.0.4. *Joint estimation.* The plug-in route is not the only way to supply the memory parameter, and a natural objection is that the verdict is specific to it. Section S9 of the Appendix (Table S21) adds the joint estimate: minimising over d the sum of squared residuals of the autoregression fitted to $(1 - L)^d \mathcal{S}_t y$, the conditional-sum-of-squares objective of Chung and Baillie (1993) concentrated in the autoregressive coefficients, at each origin, with the forecast the same map $\mathcal{M}_1(\hat{d}_{\mathrm{C}})$, so that only the choice of d changes. The joint estimate centres on zero: the median is -0.015 on headline and -0.008 on core, with unit-level estimates spread over $[-0.48, 0.21]$ and 5 per cent of cells at the lower bound of the search interval. With the order selected up to $p \leq 13$ on the same one-step criterion, the autoregression has already absorbed the dependence the conditional sum of squares can see,

as Section S1.4 of the Appendix predicts when the restricted residual is close to white; the persistence the plug-ins read as 0.22 to 0.45 lives in the retained frequency band. The joint variant is then slightly worse in the aggregate: the cross-sectional median ratio lies between 1.005 and 1.010 and exceeds one in all forty configurations, it wins a majority of units in none, and it survives with the pair in 14 of 16 pooled confidence-set cells (Table S13). Three estimators of the same parameter, on the same series, thus deliver medians of -0.01, 0.22 and 0.45 and three aggregate verdicts: the operator effect is small under every route, and its sign is the estimator's.

6.5. *The local approximation and the sign-condition diagnostic*

Proposition 2 expands around $d = 0$, while the plug-ins in the data have medians between 0.22 and 0.57. Section S5 of the Appendix checks the expansion against the exact loss difference, obtained by rerunning the rolling exercise at each point of a d-grid, with $\mathcal{G}_{t,h}$ and $\mathcal{H}_{t,h}$ computed by central finite difference through the forecasting map itself: restoring the curvature term cuts the median relative error from 0.13 to 0.03 at $d = 0.10$ and from 0.32 to 0.13 at $d = 0.20$, but even corrected the expansion is accurate in magnitude only for $d \lesssim 0.2$, which covers the median local-Whittle plug-in and not the log-periodogram one, so claims that depend on the location of 2γ apply to the smaller plug-in and not to the larger.

Solving (10) at the two observed loss ratios cannot test the mechanism: a convex parabola through the origin fitted to two points whose ordinates straddle zero has its second root between the two abscissae by construction. The sensitivity $\mathcal{G}_{t,h}$ is a derivative of the forecast map at the nesting point, computed before any loss is formed, and so supports a diagnostic that is not circular. The sign rule of Proposition 2 predicts that the fractional layer improves accuracy in a unit and horizon when the plug-in lies in $(0, 2\gamma)$ and worsens it otherwise, subject to $A > 0$, the local range and the margin condition. We evaluate the rule with two estimates of the endpoint, each constructed from the forecast-map derivatives and the restricted model's forecast errors, without using the unrestricted forecast or its loss: the projection coefficient $\hat{\gamma}^0 = \hat{B}/\hat{\mathbb{E}}[\mathcal{G}^2]$ and the corrected $\hat{\gamma} = \hat{B}/\hat{A}$. Table S10 of the Appendix compares each prediction with the realised sign of $\mathrm{MSFE}_1 / \mathrm{MSFE}_0 - 1$ over the

124 headline and 120 core cells at $\kappa = 0.70$ on which the estimates are defined; statistics involving the corrected coefficient use the 98 cells in each index with $A > 0$.

In sample the diagnostic succeeds at the smaller plug-in under either curvature coefficient, with accuracy of 0.77 to 0.90 against base rates of 0.57 to 0.61, and at the larger plug-in the corrected curvature raises skill over the base rate from $+0.14$ and $+0.02$ on headline and core to $+0.18$ and $+0.06$. That skill does not survive a split of the evaluation sample: with the endpoint estimated on the first half of each cell's origins and the rule evaluated on the second half with its own median plug-in, skill lies between -0.12 and $+0.02$ in all eight index–coefficient–plug-in cells, whereas the endpoint estimated on the second half itself recovers $+0.09$ to $+0.24$ (Table S14 of the Appendix). The halves are short, about 44 origins, and the split falls at the 2021–23 surge, but on this evidence the in-sample skill is the fit of $\hat{\gamma}$ to the errors it classifies. The diagnostic fails in three further ways (Section S6): in the fifth of cells with $A \leq 0$ no interior optimum exists; strict bracketing of the two plug-ins holds in a seventh to a third of cells; and the horizon prediction of Proposition S1 is unsupported, $\hat{\gamma}^0$ moving only from 0.26 to 0.23 between $h = 1$ and $h = 12$ on headline. The mechanism the data support is the weakest reading: the sign of the realised loss difference is fitted by the sensitivity of the forecast map, closely at small plug-ins and poorly at large ones, not predicted out of sample. Table S5 of the Appendix checks the order of the wedge on twelve units: it is $O(d)$ whether or not the coefficients are refitted, the fixed-coefficient wedge agrees with (9) to round-off, and refitting attenuates it by about a factor of three (0.030 against 0.093 at $d = 0.1$), the most direct explanation available for why the median loss ratios of Table I sit so close to unity.

7. THE INFERENCE MARGIN

The last design margin is the reference distribution. The statistic (5) is normalised by a Bartlett long-run variance truncated at the fixed lag $L = h - 1$, so $\hat{\omega}_L^2 \to_p \omega_L^2$, which equals ω_∞^2 only when the differential has no autocovariance beyond lag L; when it does, the studentised statistic is over-dispersed relative to the standard normal. This resembles the mis-normalisation studied in the fixed-bandwidth literature (Kiefer and Vogelsang, 2005, Hualde and Iacone, 2017, 2019), but there the bandwidth grows with the sample while here $L = h - 1$ is fixed, so those results do not transfer. Counting the configurations that yield at least one discovery after Benjamini–Hochberg control, the asymptotic reference

produces 24 of 80, the moving-block bootstrap 7, the break-preserving bootstrap 2 and the spectral bootstrap 1 (Table S9 of the Appendix), so a significance count depends on the reference as much as on the data. One inferential choice is held fixed throughout: the statistic is the Clark–West adjustment, which asks whether the population forecast of the larger model differs from the smaller one's, whereas the loss ratios of Table I compare realised forecasts, the object of the unadjusted statistic of Diebold and Mariano (1995) in the framework of Giacomini and White (2006). Table S17 of the Appendix sets the two side by side on the same errors, references, bootstrap draws and false-discovery-rate control. The adjusted statistic is the larger in all but one cell (medians 0.66 against -0.02), and on the full window the discovery count is the adjustment: the unadjusted statistic yields a discovery in one configuration under the asymptotic reference against 24, one against 7 under the moving-block bootstrap and none against 2 and 1 under the other two; all 45 cells the asymptotic Clark–West reference flags have a realised loss ratio below one, by too little for the unadjusted statistic, and the adjustment, which adds back $d^2\mathbb{E}[\mathcal{G}^2]$ (Proposition 3; Table S19 checks (13) cell by cell), turns them into rejections. On the window ending in 2019, where the cross-section is wider (Section 6.2), the unadjusted statistic does reject, in 22 of 76 configurations under the asymptotic reference against 53 and in 16 against 29 under the spectral bootstrap. The ordering of the references is the same under both statistics. A significance count therefore depends on the statistic as much as on the reference; the paper's own, 24 of 80, is a count of configurations with a discovery after control within the configuration, and controlling the false-discovery rate once across all 2,436 cells of the full window leaves no discovery under any reference for either statistic (Section S6 of the Appendix).

7.1. *Monte Carlo calibration*

Which of the four references is correctly sized, and whether a reference that never rejects is conservative or merely powerless, is answered by simulation under both plug-ins, since calibration evidence for one says nothing about the other; each replication passes through the same routine used on the data. Series are 310 months with an AR(2) short-run structure, a fixed seasonal pattern and $\kappa = 0.70$; 1,000 replications give a standard error of 0.7 percentage points near five per cent. Four processes: a plain autoregression; the same with

a level shift of two innovation standard deviations at seventy per cent of the sample; the same with the innovation variance tripled over the last fifth; and a fractional process with $d = 0.20$.

Table S2 of the Appendix reports the outcome, and Figure S1 plots the size, power and plug-in distributions behind it. The fixed-lag normalisation distorts at the long horizon, but only under the log-periodogram plug-in: under $\hat{d}_{\mathrm{G}}$ at $h = 12$ the asymptotic normal and the moving-block bootstrap reject 8 to 10 per cent on the memory-free processes while the break-preserving bootstraps stay between 5.0 and 6.7, because the differential has autocovariance beyond $L = h - 1$ and $\hat{\omega}_L^2$ understates ω_∞^2; under $\hat{d}_{\mathrm{W}}$ every reference is within a point and a half of nominal at $h = 12$ and conservative at $h = 1$, that branch delivering a less variable plug-in and hence a differential with less autocovariance beyond L. Only the nuisance estimator changes, so each calibration claim is restricted to the plug-in that produced it.

A level break is read as memory under either plug-in, and no reference protects against it: every reference rejects 88 to 99 per cent of the time, and rightly so about accuracy, since the fractional variant attains the lower loss in 999 and 991 of $1{,}000$ replications, but wrongly about memory, the plug-in reading 0.53 under $\hat{d}_{\mathrm{G}}$ and 0.34 under $\hat{d}_{\mathrm{W}}$ where the truth is zero, which reproduces Diebold and Inoue (2001) and Granger and Hyung (2004) inside this design; the log-periodogram plug-in has a median of 0.57 on core inflation, the magnitude this simulation produces from a memoryless process. Power is low over the whole empirical range (Table S3): at $d = 0.20$ it is 9.5 to 16.1 per cent, and at $d = 0.60$ no reference reaches half. At the panel level no reference controls the discovery rate at both horizons (Table S4): the bootstraps over-reject at $h = 1$ and the asymptotic normal at $h = 12$, with family-wise rates of up to 17 per cent against a nominal five. Section S2 of the Appendix gives the detail and two robustness notes: under the dependence-robust Benjamini and Yekutieli (2001) correction the discovery counts fall to six of eighty under the asymptotic reference and zero under the bootstraps, and the break-preserving bootstraps resample the loss differential and say nothing about whether the plug-in is itself stable. The near-absence of discoveries under the bootstraps in Table S9 is therefore as much a power statement as a calibration statement, and none of this supports a claim that fractional differencing does or does not help harmonised inflation forecasts.

8. CORE INFLATION

Repeating the grid on the index excluding energy and unprocessed food checks that the verdict is not an artefact of the volatile components of the basket, and raises the bar, because core inflation is more forecastable: at horizons up to six months the autoregression attains a median $\mathrm{MSFE}_0 \,/\, \mathrm{MSFE}_{\mathrm{RW}}$ of 0.20 against 0.41 on headline, and 0.42 against 0.71 relative to the Atkeson–Ohanian rule. The verdict is unchanged and the estimator margin is wider: median loss ratios again stay within three per cent of unity, the fractional variant wins a majority of units in 1 of 20 configurations under $\hat{d}_{\mathrm{G}}$ against 17 of 20 under $\hat{d}_{\mathrm{W}}$ (Table III), and the pooled confidence set again leaves the pair together at every horizon.

9. DISCUSSION

On 31 harmonised inflation series over a quarter century, replacing integer differencing by a plug-in continuous order moves the median loss ratio by a few per cent at most, at every horizon and in all eighty configurations, while unit-level ratios run from 0.63 to 1.53 and roughly a quarter of headline cells lie outside a five per cent band, split almost evenly between improvement and deterioration. Three design choices move the aggregate verdict on their own: the seasonality treatment, the rolling-origin rule on short windows, and the memory estimator. The last is the largest: on core inflation the fractional variant wins a majority of units in 1 of 20 configurations under one plug-in and 17 of 20 under the other, and most of that gap is bandwidth, which shifts the median memory estimate by about 0.18 against 0.05 for switching estimator; estimated jointly with the autoregression by conditional sum of squares, the memory parameter centres on zero and the cross-sectional median loss ratio exceeds one in all forty configurations, so three estimators of one parameter give medians of -0.01, 0.22 and 0.45 and three verdicts. The model confidence set leaves the autoregression and its fractional extension together in all sixteen pooled cells and in 142 of the 144 robustness configurations, and unit by unit separates the pair in a minority of cells, more often under the log-periodogram plug-in. The comparison is itself window-specific: on the window ending in 2019 the fractional layer is two to six per cent worse under the log-periodogram plug-in and neutral under local Whittle, and on core inflation neither member of the pair beats the seasonal naive rule.

The theory accounts for part of this. The forecast wedge is first-order linear in the plug-in, with a coefficient that, for fixed coefficients, is a harmonically weighted sum of the

restricted model's residual history plus a truncation term in the intercept, and that refitting cuts to about a third; the loss difference is locally quadratic with curvature $A = \mathbb{E}[\mathcal{G}^2 - e^{(0)}\mathcal{H}]$, whose second term exceeds a tenth of the first in nine cells out of ten. Against numerically estimated derivatives the sign condition classifies the realised outcome with 19 to 29 points of skill over the base rate at the smaller plug-in, in sample; out of sample, with the endpoint estimated on the first half of the origins, it has none. The sharper readings are not supported: the interval $(0, 2\gamma)$ brackets the two plug-ins in a seventh to a third of cells, in a fifth of cells the curvature is negative so no interval exists, near the upper endpoint the expansion is silent, and the horizon prediction of Proposition S1 fails.

Four limitations bound the claims. The expansions are local and, even corrected, accurate in magnitude only for $d \lesssim 0.2$, which covers the local-Whittle plug-in and not the log-periodogram one. Neither plug-in should be read as an estimate of a structural memory parameter: level shifts generate persistence that semiparametric estimators read as long memory (Diebold and Inoue, 2001, Granger and Hyung, 2004, Perron and Qu, 2010, Bos, Franses, and Ooms, 1999), and Section 7.1 shows a level shift in a memoryless process delivering $\hat{d} \approx 0.53$. Power rather than calibration is the binding constraint: at the smaller plug-in's magnitude no reference rejects more than a seventh of the time and at the larger none reaches half. And the evaluation is of point forecasts under squared error on one class of price index. Proposition S3 of the Appendix gives the condition under which the design range exceeds the operator effect; it holds in six of the eight rows of Table I, and there the comparison measures the design at least as much as the operator.

10. DISCLOSURE STATEMENT

The author declares no conflicts of interest.

11. DATA AVAILABILITY STATEMENT

All data are public: the Eurostat table `prc_hicp_midx`, all-items index `CP00` and the aggregate excluding energy and unprocessed food `TOT_X_NRG_FOOD`. Eurostat revises recent months and publishes no vintage URLs, so the replication archive ships the exact snapshot used, gzipped, as `data/estat_prc_hicp_midx.tsv.gz`, with its byte count and SHA-256 in `data/CHECKSUM`; `fetch_data.py` unpacks and verifies it, or downloads the current table if asked. The archive also holds the configuration grid,

all unit-level output, every derived table in machine-readable form, and the code with a pinned environment and fixed seeds. `run_all.py` drives the empirical pipeline behind Sections 4 to 6 in well under an hour on one core; the Monte Carlo study of Section 7.1 is a separate branch taking a few hours under both plug-ins; `code/README.md` maps every table and figure in the paper and its appendix to the function that produces it. The archive is deposited at Zenodo under DOI `10.5281/zenodo.22164867`.

REPLICATION

The proofs of every result stated above, the statements and proofs of three further propositions, the Monte Carlo tables, the order of the wedge, the estimator-by-bandwidth comparison and its continuous sweep, the accuracy of the local approximation, the joint-estimation benchmark, the pre-2020 window, the unadjusted loss differential and the further tables and figures are in the Appendix, Sections S1 to S9, twenty-one tables and four figures in all. The replication archive holds the code, the data snapshot, every archived output, a pinned environment and the fixed seeds.

APPENDIX

This appendix is the online supplement of the journal version of the paper, carried here in full so that the submission is a single document. It contains the proofs of every result stated in the main text; the statements and proofs of three further propositions, on the horizon, the origin rule and the design range, that Section 3 summarises; the tables of the Monte Carlo calibration study of Section 7.1; the estimator-by-bandwidth comparison and the accuracy check of the local approximation that Section 6 summarises; the joint-estimation benchmark; the pre-2020 window; the unadjusted loss differential; and additional tables and figures. Section, table, figure and equation numbers prefixed by S refer to this appendix; unprefixed numbers are those of the main text.

S1. PROOFS AND AUXILIARY DERIVATIONS

This section states and proves Lemma S1, and collects the proofs of Propositions 1–3 and Corollary 1 of the main text and of Propositions S1–S3 of Section S7 below, together with the derivation of the recovery formula used in Section 6.5. The arguments are elementary and are set out in full.

S1.1. *Preliminaries and notation*

Throughout the appendix t denotes a forecast origin, $h \geq 1$ the horizon, and $x_s := \mathcal{S}_t y_s$ the seasonally adjusted history available at t. We suppress the unit index. Table S1 collects the symbols used in the statements.

Two conventions apply throughout.

First, the pair is nested *computationally* and not merely nominally: at $d = 0$ every weight π_k and ψ_k with $k \geq 1$ vanishes, the fractionally differenced series is the original series, and the re-integration is the identity, so $\mathcal{M}_1(0)$ reproduces $\mathcal{M}_0$ exactly rather than in distribution.

Second, Assumption D2 holds the autoregressive order *and* the order common across the two members at each origin, while letting each estimate its own coefficients. This is what allows the perturbation induced by the fractional filter to be propagated through the forecast function without re-deriving it. The expansions below do involve the estimator: re-estimation on the filtered series contributes $\mathcal{G}^{\mathrm{es}}$, and Remark 1 turns on that term.

TABLE S1

NOTATION USED IN THE STATEMENTS AND PROOFS.

Symbol	Meaning
d, $\hat{d}$	memory parameter; its plug-in estimate
$\hat{d}_{\mathrm{G}}$, $\hat{d}_{\mathrm{W}}$	log-periodogram and local-Whittle plug-ins
$\pi_k(d)$, $\psi_k(d)$	binomial weights of $(1-L)^d$ and $(1-L)^{-d}$
$x_s = \mathcal{S}_t y_s$	seasonally adjusted history at origin t
$\hat{y}^{(0)}_{t+h}$, $\hat{y}^{(1)}_{t+h}$	restricted and unrestricted h-step forecasts
$e^{(0)}$, $e^{(1)}$	corresponding forecast errors
$\mathcal{G}_{t,h}$	first-order forecast wedge, equation (7)
γ	pseudo-optimal memory, equation (10)
MSFE_j	mean squared forecast error of member j
κ, $\mathcal{O}(\kappa)$, $P(\kappa)$	origin rule, origin set, number of origins
μ_r, $w_r(\kappa)$	regime mean of the differential; share of origins in regime r
c, ϕ_i, θ_j	intercept, slopes and impulse responses of the restricted autoregression
$\hat{\varepsilon}_s$, H_t	its one-step residuals, zero pre-sample; the harmonic number $\sum_{k\le t} k^{-1}$

S1.2. *Binomial weights near zero*

LEMMA S1—Binomial weights near zero: *For every* $k \ge 1$, $\pi_k(d) = -d/k + O(d^2)$ *and* $\psi_k(d) = +d/k + O(d^2)$, *uniformly on compact neighbourhoods of* $d = 0$.

Iterating the recursion $\pi_k(d) = \pi_{k-1}(d)\,(k-1-d)/k$ from $\pi_0(d) = 1$ gives the closed form

$$\pi_k(d) \;=\; \prod_{j=1}^{k} \frac{j-1-d}{j}, \qquad k \ge 1. \tag{S1}$$

Isolate the first factor. At $j = 1$ the numerator is $-d$ exactly, so (S1) may be written

$$\pi_k(d) \;=\; -d \prod_{j=2}^{k} \frac{j-1-d}{j} \;=\; -d \prod_{j=2}^{k} \frac{j-1}{j} \prod_{j=2}^{k} \Big(1 - \frac{d}{j-1}\Big). \tag{S2}$$

The first product telescopes: $\prod_{j=2}^{k}(j-1)/j = 1/k$. For the second, take logarithms. For $|d| \le \frac{1}{2}$ and $j \ge 2$ we have $|d/(j-1)| \le \frac{1}{2}$, so $\big|\log(1 - d/(j-1)) + d/(j-1)\big| \le$

$d^2/(j-1)^2$, whence

$$\Big|\log\prod_{j=2}^{k}\Big(1-\frac{d}{j-1}\Big)+d\sum_{j=2}^{k}\frac{1}{j-1}\Big| \;\le\; d^2\sum_{j\ge 2}\frac{1}{(j-1)^2} \;=\; \frac{\pi^2}{6}\,d^2. \tag{S3}$$

Since $\sum_{j=2}^{k}(j-1)^{-1}=O(\log k)$, exponentiating (S3) gives $\prod_{j=2}^{k}(1-d/(j-1))=1+O(d\log k)$ uniformly for $|d|\le\frac{1}{2}$. Substituting this and the telescoped product into (S2),

$$\pi_k(d) \;=\; -\frac{d}{k}\big(1+O(d\log k)\big) \;=\; -\frac{d}{k}+O\!\left(\frac{d^2\log k}{k}\right), \tag{S4}$$

which is the claim, the remainder being uniform on $\{|d|\le\frac{1}{2}\}$. The weights ψ_k satisfy the same recursion with $-d$ replaced by $+d$, so (S4) applies verbatim with the sign reversed.

REMARK 2: The remainder in (S4) carries a factor $\log k$, so the approximation degrades slowly in the lag. Summed against a history of length t satisfying Assumption D1 the remainder contributes $O(d^2\log^2 t)$; the expansions of Propositions 1 and 2 are in d at a fixed origin, so this is absorbed into their $O(d^2)$ remainders.

Q.E.D.

S1.3. *Proof of Lemma 1 and Proposition 1*

The argument proceeds in four steps: differentiability, an exact identity for the forecast map with the coefficients held fixed, its first-order term, and the re-estimation term.

Step 1: differentiability. The unrestricted forecast is a finite composition of maps that are smooth in d: the weights $\pi_k(d)$ and $\psi_k(d)$ are polynomials in d of degree k by (S1); the filtered series is a finite linear combination of them; the ordinary least squares coefficients are a smooth function of the filtered series wherever the design matrix has full column rank, which Assumption D1 imposes on a neighbourhood of the origin, almost surely; and the forecast is polynomial in those coefficients. Each of these maps is infinitely differentiable on that neighbourhood, and the composition of finitely many such maps is again infinitely differentiable there; hence $d\mapsto\hat{y}^{(1)}_{t+h}(d)$ is C^∞ on a neighbourhood of the origin, almost

surely. In particular it is C^1, and Taylor's theorem with Lagrange remainder gives (7) with $\mathcal{G}_{t,h}$ the derivative at zero; the C^2 claim of Lemma 1 and the third derivative used in Proposition 2 follow from the same argument. The argument does not deliver a moment bound: pathwise smoothness holds almost surely, whereas the expansion of the expected loss needs the derivatives to be uniformly integrable on a neighbourhood, so Proposition 2 assumes that separately. At $d = 0$ all weights with $k \geq 1$ vanish, the filter and its inverse are the identity, and the refitted coefficients coincide with the restricted ones, so $\hat{y}^{(1)}_{t+h}(0) = \hat{y}^{(0)}_{t+h}$.

Step 2: the fixed-coefficient identity. Hold the intercept c and the slopes $\phi_1, \ldots, \phi_p$ at their restricted values. Write $x = (x_1, \ldots, x_t)$ for the adjusted history, Π_d for the Type-II fractional difference, which acts on a finite sequence with zero pre-sample values as the lower-triangular Toeplitz matrix with entries $\pi_{s-r}(d)$, and Φ for the lower-triangular Toeplitz matrix of $\phi(L) = 1 - \sum_i \phi_i L^i$. Two facts are used. First, $\Pi_{-d}\Pi_d = I$ on finite sequences, because $(1-L)^{-d}(1-L)^d = 1$ as formal power series and the product of two lower-triangular Toeplitz matrices is the Toeplitz matrix of the product series; so the re-integration in (2) reproduces the observed history exactly, and the extended re-integrated sequence $X = \Pi_{-d}W$ satisfies $X_s = x_s$ for $s \leq t$ and $X_{t+j} = \hat{y}^{(1)}_{t+j}(d)$ for $j = 1, \ldots, h$. Second, lower-triangular Toeplitz matrices commute, so $\Phi\Pi_d = \Pi_d\Phi$.

The unrestricted procedure extends $W = \Pi_d X$ by the autoregressive recursion with zero innovations, $(\Phi W)_{t+j} = c$ for $j = 1, \ldots, h$. Define the pseudo-innovation sequence $U := \Phi X - c\mathbf{1}$, so that $U_s = \hat{\varepsilon}_s$ for $s \leq t$ by the definition of the residuals with zero pre-sample values. Then $\Phi W = \Pi_d \Phi X = \Pi_d(U + c\mathbf{1})$, and the recursion says, for $j \geq 1$,

$$c \;=\; \big(\Pi_d(U + c\mathbf{1})\big)_{t+j} \;=\; U_{t+j} + c + \sum_{k=1}^{t+j-1} \pi_k(d)\big(U_{t+j-k} + c\big), \tag{S5}$$

which is the recursion for $U_{t+j}(d)$ in (9). The restricted forecast is the same recursion with zero pseudo-innovations, $(\Phi X^{(0)})_{t+j} = c$ and $X^{(0)}_s = x_s$ for $s \leq t$. Subtracting, $\Phi(X - X^{(0)})$ vanishes at every $s \leq t$ and equals $U_{t+j}(d)$ at $t + j$, so $X_{t+h} - X^{(0)}_{t+h} = \sum_{j=1}^{h} \theta_{h-j} U_{t+j}(d)$ with θ the impulse responses of $\phi(L)^{-1}$, which is (9). Nothing in this step is approximate.

Step 3: the first-order term. By Lemma S1, $\pi_k(d) = -d/k + O(d^2)$ uniformly for $k \leq t+h$, and each $U_{t+j}(d)$ with $j \geq 1$ is $O(d)$, so in (S5) the terms with $t+j-k > t$ contribute

$O(d^2)$ through U_{t+j-k} but $-\pi_k(d)\,c$ through the intercept, and

$$U_{t+j}(d) \;=\; d\Big(\sum_{k=j}^{t+j-1}\frac{\hat{\varepsilon}_{t+j-k}}{k} + c\,H_{t+j-1}\Big) + O(d^2), \qquad H_s = \sum_{k=1}^{s} k^{-1}. \tag{S6}$$

Substituting into (9) gives $\mathcal{G}^{\text{fix}}_{t,h}$ as in (8); at $h=1$ it is $\sum_{k=1}^{t}\hat{\varepsilon}_{t+1-k}/k + cH_t$. Section S3 checks (8) against the derivative of the fixed-coefficient map at $h=1$ and $h=12$, and Section S1.5 below gives $\mathcal{G}^{\text{es}}$ in closed form and checks the sum against the finite-difference derivative of the actual map at every origin. For the rate, $\sum_{k\le t}|\hat{\varepsilon}_{t+1-k}|/k = O_p(\log t)$ under Assumption D1 and $cH_t = O(\log t)$, so $\mathcal{G}^{\text{fix}}_{t,h} = O_p(\log t)$. The residual part is in fact $O_p(1)$ whenever the residuals have summable autocovariances, since then $\operatorname{Var}\big(\sum_{k\le t}\hat{\varepsilon}_{t+1-k}/k\big) \le \sum_{k,j\le t}|\gamma_\varepsilon(k-j)|/(kj) \le \sum_m|\gamma_\varepsilon(m)|\sum_k k^{-2}$; the $\log t$ rate belongs to the intercept part alone. The commutation used in Step 2 is between the two filters. The autoregressive *recursion* does not commute with Π_d: the recursion applied to the filtered series is the filtered recursion driven by $-\sum_k \pi_k(d)(U_{t+j-k}+c)$ rather than by zero, and that is the first-order effect.

Step 4: what re-estimation adds. Let $\hat{\beta}(d)$ denote the least squares coefficients fitted to $w(d)$ and $\hat{\beta}(0)$ those fitted to the adjusted level series. By Step 1 the map is differentiable, and by the chain rule

$$\mathcal{G}^{\text{es}}_{t,h} = \frac{\partial F_h\big(z;\hat{\beta}\big)}{\partial \beta'}\bigg|_{\beta=\hat{\beta}(0)} \frac{\partial \hat{\beta}(d)}{\partial d}\bigg|_{d=0}, \tag{S7}$$

the product of the forecast function's sensitivity to the coefficients and the coefficients' sensitivity to the filter. Both factors are $O_p(1)$ under Assumption D1, so $\mathcal{G}^{\text{es}}_{t,h} = O_p(1)$. Summing (8) and (S7) gives (6) and completes the proof of Lemma 1; substituting into the Taylor expansion of Step 1 gives (7).

REMARK 3: Nothing in Step 4 signs $\mathcal{G}^{\text{es}}$ relative to $\mathcal{G}^{\text{fix}}$. The two could reinforce or offset. Table S5 settles the question empirically for these data: they offset, and the offset is substantial, since holding the coefficients fixed roughly triples the wedge. Refitting the autoregression on the filtered series recovers part of the dynamics the filter removed and, through the intercept, removes the truncation term cH_t, which on these data is the larger part of $\mathcal{G}^{\text{fix}}$ (Section S3).

Q.E.D.

S1.4. *The pseudo-optimal memory under fixed coefficients*

This subsection makes precise the reading of Remark 1. Suppose the coefficients are held fixed, $h = 1$, and the restricted residuals $\{\varepsilon_s\}$ form a mean-zero covariance-stationary sequence with variance σ^2, autocorrelations $\rho(k)$ and summable autocovariances. Then $e^{(0)}_{t+1} = \varepsilon_{t+1}$, and by (8)

$$B = \mathbb{E}\big[e^{(0)}_{t+1}\mathcal{G}^{\mathrm{fix}}_{t,1}\big] = \sigma^2 \sum_{k=1}^{t} \frac{\rho(k)}{k}, \qquad \mathbb{E}\big[(\mathcal{G}^{\mathrm{fix}}_{t,1})^2\big] = \sigma^2 \sum_{k,j\le t} \frac{\rho(|k-j|)}{kj} + c^2 H_t^2, \tag{S8}$$

the intercept term dropping out of B because $\mathbb{E}[\varepsilon_{t+1}] = 0$. Three readings follow. The first needs two conditions, which are worth keeping apart. If the residual is covariance white, $\rho(k) = 0$ for $k \ge 1$, then $B = 0$ by (S8) and the pseudo-optimal memory is zero: this uses only second moments, because $\mathcal{G}^{\mathrm{fix}}_{t,1}$ is a linear functional of the residual history. Whether the curvature is then $\mathbb{E}[\mathcal{G}^2]$ is a different question. With fixed coefficients and $h = 1$, $\mathcal{H}_{t,1}$ is a function of the history alone, so if in addition the restricted one-step error is a martingale difference with respect to the information set at the origin, $\mathbb{E}[e^{(0)}_{t+1} \mid \mathcal{F}_t] = 0$, with independent white noise the sufficient special case, then $\mathbb{E}[e^{(0)}\mathcal{H}] = \mathbb{E}[\mathcal{H}_{t,1}\mathbb{E}(e^{(0)}_{t+1} \mid \mathcal{F}_t)] = 0$ and $A = \mathbb{E}[\mathcal{G}^2] > 0$, so the fractional layer can only hurt, by Ad^2 to second order. Covariance whiteness alone does not give this: a process that is uncorrelated but nonlinearly dependent on its past can be correlated with $\mathcal{H}_{t,1}$, which is a nonlinear functional of the history. Remark 4 makes the corresponding point for the map as implemented, where the error is an h-step error and $\mathcal{H}_{t,h}$ carries the refitting step as well, and Remark 12 reports that in these data $\mathbb{E}[e^{(0)}\mathcal{H}]/\mathbb{E}[\mathcal{G}^2]$ has median -0.13, so the curvature correction is not negligible there. If the residual is fractional noise of order d_0, its autocorrelations are $\rho(k) = \Gamma(k+d_0)\Gamma(1-d_0)/\{\Gamma(k+1-d_0)\Gamma(d_0)\} = d_0/k + O(d_0^2 \log k/k)$, so with $c = 0$ and $t \to \infty$, $B = \sigma^2 d_0 \pi^2/6 + O(d_0^2)$ and $\mathbb{E}[\mathcal{G}^2] = \sigma^2\pi^2/6 + O(d_0)$, whence the projection coefficient $\gamma^0 = B/\mathbb{E}[\mathcal{G}^2]$ equals $d_0 + O(d_0^2)$: to first order, the plug-in that the local expansion favours is the memory of the residual. In a simulation of fractional noise with $t = 300$ and $4{,}000$ replications (`paperD_order.py`, `out_D_order_gamma_check.csv`) the ratio $\hat{B}/\hat{\mathbb{E}}[\mathcal{G}^2]$ is 0.083, 0.154 and 0.181 at $d_0 = 0.10$, 0.20 and 0.30, so the second-order

term is negative and material beyond $d_0 \approx 0.1$. In general, with $c = 0$, γ^0 is the harmonically weighted average $\sum_k \rho(k)/k$ divided by $\sum_{k,j} \rho(|k-j|)/(kj)$: positive when the restricted model leaves positive dependence at long lags, whether that dependence is fractional or generated by level shifts (Diebold and Inoue, 2001, Granger and Hyung, 2004), and small whenever the autoregression has absorbed the dependence at the lags the harmonic weights emphasise. The joint estimate of Section 6.4 of the main text, which minimises the same one-step sum of squares that selects the autoregression, is -0.015 in the median (Table S21, Section S9), as this reading predicts. With refitting the intercept term is removed and the slopes adjust, which attenuates $\mathcal{G}$ (Table S5); the sign-condition diagnostic of Section 6.5 of the main text therefore works with the refitted derivatives rather than with this closed form.

S1.5. *The re-estimation term in closed form*

Step 4 of Section S1.3 wrote $\mathcal{G}^{\text{es}}$ as the product of two derivatives. Both are explicit. Write $\beta(d) = (c(d), \phi(d))$ for the least squares coefficients fitted to $w(d) = \Pi_d x$ and $F_h(x;\beta)$ for the restricted recursion on x with coefficients β. The identity of Step 2 holds for any fixed coefficients, in particular at $\beta = \beta(d)$:

$$\hat{y}^{(1)}_{t+h}(d) \;=\; F_h\big(x;\beta(d)\big) + \sum_{j=1}^{h} \theta_{h-j}\big(\beta(d)\big)\, U_{t+j}\big(d;\beta(d)\big), \tag{S9}$$

where $U(\cdot;\beta)$ is built from the residuals $\hat{\varepsilon}_s(\beta) = x_s - c - \sum_i \phi_i x_{s-i}$. Since $U_{t+j}(0;\beta) = 0$ for every β, differentiating at $d = 0$ gives $\mathcal{G} = \mathcal{G}^{\text{fix}} + \mathcal{G}^{\text{es}}$ with

$$\mathcal{G}^{\text{es}}_{t,h} \;=\; \frac{\partial F_h(x;\beta)}{\partial \beta'}\bigg|_{\beta=\hat{\beta}} \dot{\beta}, \qquad \dot{\beta} \;=\; (X'X)^{-1}\big[\dot{X}'Y + X'\dot{Y} - (\dot{X}'X + X'\dot{X})\hat{\beta}\big], \tag{S10}$$

where $Y = (w_{p+1}, \ldots, w_t)'$ and $X = [\iota,\ w \text{ lagged } 1, \ldots, p]$ are the least squares data at $d = 0$ (that is, x itself), and $\dot{Y}$, $\dot{X}$ are the same arrays built from $\dot{w}_s := \partial w_s/\partial d|_0 = -\sum_{k=1}^{s-1} x_{s-k}/k$, the harmonic sum of the past given by Lemma S1, with a zero column in place of the intercept. Since the harmonic operator and $\phi(L)$ commute on zero-pre-sample sequences, $\dot{Y} - \dot{X}\hat{\beta}$ has entries $-\sum_k(\hat{\varepsilon}_{s-k} + c)/k$, so $\dot{\beta} = (X'X)^{-1}[\dot{X}'\hat{\varepsilon} - X'\mathcal{H}(\hat{\varepsilon} + c)]$ with $\mathcal{H}$ the harmonic-sum operator: the refitted coefficients move by the regression of the harmonic sums of the residual history on the design, which is how refitting absorbs the

intercept term cH_t of $\mathcal{G}^{\text{fix}}$. The forecast gradient follows from differentiating the recursion, $\partial\hat{y}_{t+j}/\partial c = 1 + \sum_i \phi_i\, \partial\hat{y}_{t+j-i}/\partial c$ and $\partial\hat{y}_{t+j}/\partial\phi_i = v_{t+j-i} + \sum_l \phi_l\, \partial\hat{y}_{t+j-l}/\partial\phi_i$, with v the observed or forecast values and zero derivatives for observed ones.

Stage 5.6 of the replication archive (`paperD_wedge_closed.py`) evaluates $\mathcal{G}^{\text{fix}} + \mathcal{G}^{\text{es}}$ from (8) and (S10) at every origin of every unit–horizon cell at $\kappa = 0.70$, 21,562 origins on the two indices, and compares it with the central finite difference of the actual map used in Section 6.5 of the main text and Section S5: the largest absolute discrepancy is $4.3e-04$, the precision of a central difference with step 10^{-3}, so the finite-difference $\mathcal{G}$ of the main text is the closed form. The two parts offset almost entirely: they have opposite signs at 94 per cent or more of the origins in every cell, their correlation across origins is between -0.94 and -0.61 by index and horizon, and the median of $|\mathcal{G}|/|\mathcal{G}^{\text{fix}}|$ lies between 0.15 and 0.28: refitting removes between three quarters and 85 per cent of the fixed-coefficient wedge, leaving a sensitivity of 0.18 to 0.27 standard deviations of the adjusted series against 0.6 to 1.5 for the fixed-coefficient one (`out_D_wedge_closed_summary.csv`).

S1.6. *Proof of Proposition 2*

The proof carries the expansion of the forecast map one term further than Proposition 1.

Step 1: second-order forecast expansion. By Step 1 of Appendix S1.3 the map $d \mapsto \hat{y}^{(1)}_{t+h}(d)$ is twice continuously differentiable near the origin; write its Taylor expansion

$$\hat{y}^{(1)}_{t+h}(d) = \hat{y}^{(0)}_{t+h} + d\,\mathcal{G}_{t,h} + \tfrac{1}{2}d^2\,\mathcal{H}_{t,h} + O_p(d^3), \qquad \mathcal{H}_{t,h} = \left.\frac{\partial^2 \hat{y}^{(1)}_{t+h}}{\partial d^2}\right|_{d=0}. \tag{S11}$$

Step 2: the error and its square. Subtracting (S11) from the target,

$$e^{(1)}_{t+h}(d) = e^{(0)}_{t+h} - d\,\mathcal{G}_{t,h} - \tfrac{1}{2}d^2\,\mathcal{H}_{t,h} + O_p(d^3), \tag{S12}$$

and squaring, retaining every term through d^2,

$$\big(e^{(1)}\big)^2 = \big(e^{(0)}\big)^2 - 2d\,e^{(0)}\mathcal{G} + d^2\mathcal{G}^2 - d^2\,e^{(0)}\mathcal{H} + O_p(d^3). \tag{S13}$$

The fourth term is the one a first-order expansion of the forecast would lose. It arises as $2e^{(0)} \times (-\tfrac{1}{2}d^2\mathcal{H})$: the product of the restricted error with the second-order term of the forecast is of order d^2, not d^3.

Step 3: expectations. Taking expectations of (S13), which exist under the moment condition of Proposition 2 because $e^{(0)}\mathcal{G}$, $\mathcal{G}^2$ and $e^{(0)}\mathcal{H}$ are products of square-integrable variables, and subtracting MSFE_0,

$$\mathrm{MSFE}_1(d) - \mathrm{MSFE}_0 = -2d\,\mathbb{E}\big[e^{(0)}\mathcal{G}\big] + d^2\,\mathbb{E}\big[\mathcal{G}^2 - e^{(0)}\mathcal{H}\big] + O(d^3), \tag{S14}$$

which is (10) with $B = \mathbb{E}[e^{(0)}\mathcal{G}]$ and $A = \mathbb{E}[\mathcal{G}^2 - e^{(0)}\mathcal{H}]$.

Step 4: the sign condition. Suppose $A > 0$ and set $\gamma = B/A$, so that (S14) reads $Q(d) + R(d)$ with $Q(d) = Ad(d - 2\gamma)$. The remainder is uniform on $\mathcal{N} = [-\delta_0, \delta_0]$. Write $\ell(d) := (e^{(1)}(d))^2$ for the loss map and primes for derivatives in d; then $\ell' = -2e^{(1)}\hat{y}'$, $\ell'' = 2(\hat{y}')^2 - 2e^{(1)}\hat{y}''$ and $\ell''' = 6\hat{y}'\hat{y}'' - 2e^{(1)}\hat{y}'''$, with $\ell(0) = (e^{(0)})^2$, $\ell'(0) = -2e^{(0)}\mathcal{G}$ and $\ell''(0) = 2(\mathcal{G}^2 - e^{(0)}\mathcal{H})$. On $\mathcal{N}$ each factor of ℓ''' is dominated by a linear combination of $|e^{(0)}|$, $|\mathcal{G}|$, $|\mathcal{H}|$ and $\sup_{\mathcal{N}}|\hat{y}'''|$ with coefficients that depend only on δ_0, so $\mathbb{E}\sup_{\mathcal{N}}|\ell'''| < \infty$ under the second-moment condition of the proposition, by the Cauchy–Schwarz inequality. Taylor's theorem with Lagrange remainder, applied pathwise to ℓ and then integrated, gives (S14) with $|R(d)| \le \frac{1}{6}\mathbb{E}\sup_{\mathcal{N}}|\ell'''|\,|d|^3 =: C|d|^3$ for $|d| \le \delta_0$. The quadratic Q has roots 0 and 2γ with positive leading coefficient, so $Q < 0$ exactly between them, which is (11); its vertex is at $d = \gamma$ with value $-A\gamma^2$.

Transferring that sign to $\mathrm{MSFE}_1(d) - \mathrm{MSFE}_0$ requires the quadratic to exceed the remainder bound. Since $|Q(d)| = A|d|\,|d - 2\gamma|$, the inequality $|Q(d)| > C|d|^3$ is equivalent to

$$A\,|d - 2\gamma| \;>\; C\,d^2, \tag{S15}$$

which is (12). Wherever (S15) holds and $|d| \le \delta_0$, the loss difference inherits the sign of Q.

The two roots are not symmetric under this condition. At $d = 0$ the left side of (S15) tends to $2A\gamma > 0$ while the right side tends to zero, so the condition holds on a punctured neighbourhood of the origin: there Q crosses zero at rate $2A\gamma$ while the remainder is of strictly higher order. At $d = 2\gamma$ both sides approach 0 and $4C\gamma^2$ respectively, so the condition fails in a neighbourhood of the second root. A radius δ around the origin therefore does not by itself deliver the sign of the loss difference on the whole complement of $[0, 2\gamma]$: for $d = 2\gamma + \epsilon$ we have $Q(d) = A(2\gamma + \epsilon)\epsilon \to 0$ as $\epsilon \to 0$, while the available bound remains of order $C(2\gamma)^3$. Solving the quadratic inequality gives the explicit excluded band around

the second root,

$$|d - 2\gamma| \le C\, d^2/A \qquad \Longrightarrow \qquad \text{the expansion determines nothing at } d,$$

whose half-width near $d = 2\gamma$ is approximately $4C\gamma^2/A$. Estimating C is not possible without a bound on the third derivative, which is why Section S5 measures the usable range directly instead.

If $A \le 0$ the leading behaviour away from zero is $-2Bd$ with concave curvature, there is no interior minimum, and no interval statement is available.

REMARK 4—The curvature term does not vanish by orthogonality: One might expect $\mathbb{E}[e^{(0)}\mathcal{H}] = 0$ from some orthogonality of the fitted model. It does not follow. Least squares makes the *in-sample* residual orthogonal to the regressors at the estimation origin, whereas $e^{(0)}_{t+h}$ is an out-of-sample h-step error and $\mathcal{H}_{t,h}$ is the curvature of the whole rolling map, including the refitting step. The two are generically correlated, and Remark 12 reports that in these data they are: the ratio $\mathbb{E}[e^{(0)}\mathcal{H}]/\mathbb{E}[\mathcal{G}^2]$ has median -0.13 with interquartile range $[-0.96, +0.57]$.

REMARK 5: The neglected $O(d^3)$ term still binds near $d = 2\gamma$, where the quadratic vanishes to first order. Table S7 shows the corrected quadratic is accurate to about three per cent at $d = 0.10$ and thirteen per cent at $d = 0.20$, and deteriorates rapidly beyond $d \approx 0.3$; that range, not the algebra, is what bounds the usable domain of (11).

Q.E.D.

S1.7. *Proof of Corollary 1*

By hypothesis $A > 0$, $\gamma > 0$, $\hat{d}_{\mathrm{W}} \to_p d_{\mathrm{W}}$ and $\hat{d}_{\mathrm{G}} \to_p d_{\mathrm{G}}$ with $0 < d_{\mathrm{W}} < 2\gamma < d_{\mathrm{G}} \le \delta_0$, and both limits satisfy the margin condition (S15).

Step 1: the limits. Evaluate $Q(d) = Ad(d - 2\gamma)$ at the two limits. Since $d_{\mathrm{W}} \in (0, 2\gamma)$ we have $Q(d_{\mathrm{W}}) < 0$, and since $d_{\mathrm{G}} > 2\gamma$ we have $Q(d_{\mathrm{G}}) > 0$. By Step 4 of Appendix S1.6 the margin condition transfers each sign to the loss difference itself, so

$$\mathrm{MSFE}_1(d_{\mathrm{W}}) < \mathrm{MSFE}_0 < \mathrm{MSFE}_1(d_{\mathrm{G}}). \tag{S16}$$

Step 2: from the limits to the estimates. Equation (S16) concerns the two limits, which are constants; the plug-ins are random, so the conclusion about them is a statement about probability and not an inequality that holds path by path at any finite T.

Write $F(d) = \mathrm{MSFE}_1(d) - \mathrm{MSFE}_0$ and let

$$\mathcal{O} \;=\; \{d : |d| \le \delta_0,\; A|d - 2\gamma| > Cd^2\},$$

the set on which the margin condition holds. Both hypotheses are strict, so d_{W} and d_{G} lie in the interior of $\mathcal{O}$, and $\mathcal{O}$ is open because both sides of the defining inequality are continuous in d. Choose $\epsilon > 0$ small enough that the balls $B(d_{\mathrm{W}}, \epsilon)$ and $B(d_{\mathrm{G}}, \epsilon)$ are contained in $\mathcal{O}$, that $B(d_{\mathrm{W}}, \epsilon) \subset (0, 2\gamma)$ and that $B(d_{\mathrm{G}}, \epsilon) \subset (2\gamma, \delta_0]$. On the first ball $Q < 0$ and the margin condition holds, so $F < 0$ there; on the second $Q > 0$ and $F > 0$. Hence

$$\{|\hat{d}_{\mathrm{W}} - d_{\mathrm{W}}| < \epsilon\} \cap \{|\hat{d}_{\mathrm{G}} - d_{\mathrm{G}}| < \epsilon\} \;\subseteq\; \big\{\mathrm{MSFE}_1(\hat{d}_{\mathrm{W}}) < \mathrm{MSFE}_0 < \mathrm{MSFE}_1(\hat{d}_{\mathrm{G}})\big\},$$

and since $\hat{d}_{\mathrm{W}} \to_p d_{\mathrm{W}}$ and $\hat{d}_{\mathrm{G}} \to_p d_{\mathrm{G}}$ the probability of the left-hand event tends to one. Therefore

$$\mathbb{P}\big\{\mathrm{MSFE}_1(\hat{d}_{\mathrm{W}}) < \mathrm{MSFE}_0 < \mathrm{MSFE}_1(\hat{d}_{\mathrm{G}})\big\} \longrightarrow 1,$$

which is the claim. If the plug-ins converge almost surely, the same inclusion gives the inequality for all T beyond a random threshold; convergence in probability alone does not.

Two comments on the hypotheses. The margin condition cannot be replaced by a radius: $d_{\mathrm{G}} \le \delta_0$ alone leaves d_{G} free to sit arbitrarily close to 2γ, where Q vanishes and the cubic bound does not. And C is not available in closed form, so the condition is not checkable from the data without estimating a third derivative. Section S5 therefore measures the range over which the quadratic tracks the exact loss difference, and Section 6.5 of the main text tests the sign condition directly, rather than relying on this corollary.

REMARK 6: The argument requires nothing about which estimator, if either, is consistent for a structural memory parameter, and asserts nothing about whether either is correctly specified. The two differ in their semiparametric objective and in bandwidth—$m_{\mathrm{G}} = \lfloor T^{1/2} \rfloor$ against $m_{\mathrm{W}} = \lfloor T^{0.65} \rfloor$—and therefore in the finite-sample bias induced by short-run dynamics, and, if the persistence is partly break-induced, in what they converge

to at all. The corollary says only that when the operator effect is of the same order as the gap between them, that gap decides the comparison.

Q.E.D.

S1.8. *Proof of Proposition 3*

Expand the unrestricted error and the forecast difference as in Step 2 of Section S1.6: $e^{(1)} = e^{(0)} - d\mathcal{G} - \frac{1}{2}d^2\mathcal{H} + O_p(d^3)$ and $\hat{y}^{(1)} - \hat{y}^{(0)} = d\mathcal{G} + \frac{1}{2}d^2\mathcal{H} + O_p(d^3)$. Then

$$f^{\mathrm{DM}} = (e^{(0)})^2 - (e^{(1)})^2 = 2d\,e^{(0)}\mathcal{G} + d^2\,e^{(0)}\mathcal{H} - d^2\mathcal{G}^2 + O_p(d^3), \tag{S17}$$

$$(\hat{y}^{(0)} - \hat{y}^{(1)})^2 = d^2\mathcal{G}^2 + O_p(d^3), \tag{S18}$$

$$f^{\mathrm{CW}} = f^{\mathrm{DM}} + (\hat{y}^{(0)} - \hat{y}^{(1)})^2 = 2d\,e^{(0)}\mathcal{G} + d^2\,e^{(0)}\mathcal{H} + O_p(d^3). \tag{S19}$$

Taking expectations under the moment conditions of Proposition 2, with the remainders bounded exactly as in Step 4 of Section S1.6, gives (13), and subtracting the two lines gives the gap $d^2\mathbb{E}[\mathcal{G}^2]$. The particular cases follow by substitution. For the last claim, replace d by a plug-in $\hat{d} \to_p d^*$: by the argument of Section S1.7 the expectations at $\hat{d}$ converge to the expressions at d^*, and the gap converges to $d^{*2}\mathbb{E}[\mathcal{G}^2]$, which is zero only if $d^* = 0$ or $\mathcal{G} = 0$ almost surely. *Q.E.D.*

REMARK 7: In the setting of Clark and West (2007) the nesting parameter is estimated by least squares on the evaluation model, its estimate under the null is $O_p(R^{-1/2})$ in the estimation sample size R, and the adjustment removes a term of order $1/R$ that would otherwise bias the unadjusted statistic against the larger model. A semiparametric plug-in on a series whose short-run dynamics leave persistence in the retained frequency band converges to a non-zero d^* even when the fractional layer does not forecast better (the joint estimate of Section 6.4 of the main text is zero while the plug-ins are 0.2 to 0.5), so the term the adjustment removes is not a vanishing bias but the population cost of the layer. The comparison of Table S17 is the empirical face of this: on the full window the adjusted statistic rejects where the unadjusted one does not, in cells whose realised loss ratio is below one but not by enough for the unadjusted statistic.

S1.9. *Proof of Proposition S1*

The first claim is a restatement: by construction $\gamma_h^0 = \mathbb{E}[e_{t+h}^{(0)}\mathcal{G}_{t,h}]/\mathbb{E}[\mathcal{G}_{t,h}^2]$ is the coefficient minimising $\mathbb{E}[(e_{t+h}^{(0)} - d\mathcal{G}_{t,h})^2]$ over d, which is the linear projection of $e_{t+h}^{(0)}$ on $\mathcal{G}_{t,h}$; the pseudo-optimal memory $\gamma_h = B_h/A_h$ differs from it by the factor $\mathbb{E}[\mathcal{G}_{t,h}^2]/A_h$, which is one only when $\mathbb{E}[e^{(0)}\mathcal{H}] = 0$.

(i). By Lemma 1 and Sections S1.3 and S1.5, $\mathcal{G}_{t,h} = \mathcal{G}_{t,h}^{\text{fix}} + \mathcal{G}_{t,h}^{\text{es}}$ with, for $c = 0$, $\mathcal{G}_{t,h}^{\text{fix}} = \sum_{j=1}^{h}\theta_{h-j}S_j$, $S_j = \sum_{m=1}^{t}\hat{\varepsilon}_m/(t+j-m)$, and $\mathcal{G}_{t,h}^{\text{es}} = (\partial F_h/\partial\beta)'\dot{\beta}$ with $\dot{\beta}$ independent of h. Since $|S_j| \le \sum_m|\hat{\varepsilon}_m|/j \to 0$ and the impulse responses are summable, $\mathcal{G}_{t,h}^{\text{fix}} \to 0$. The gradient of the recursive forecast satisfies $\partial F_h/\partial c = \sum_{i<h}\theta_i \to 1/(1-\sum_i\phi_i)$ and $\partial F_h/\partial\phi_i \to \partial\mu/\partial\phi_i = c/(1-\sum_i\phi_i)^2 = 0$ for a stationary model with $c = 0$, so $\mathcal{G}_{t,h}^{\text{es}} \to \dot{c}_t/(1-\sum_i\phi_i)$, which is (S29); the convergence is pathwise, and holds in mean square under the moment conditions of Proposition 2 by dominated convergence, the terms being bounded by multiples of $\sum_m|\hat{\varepsilon}_m|$ and of $|\dot{\beta}|$. Since $\dot{\beta}$ is a ratio of quadratic forms in the history, $\mathcal{G}_t^{\infty}$ is a square-integrable functional of the history up to t.

For the representation, the restricted model is stationary with unconditional mean μ and $\hat{y}_{t+h}^{(0)} \to \mu$ in mean square by assumption, so $e_{t+h}^{(0)} - (y_{t+h}-\mu) \to 0$ in mean square. Since $e_{t+h}^{(0)} - (y_{t+h}-\mu)$ and $\mathcal{G}_{t,h} - \mathcal{G}_t^{\infty}$ both tend to zero in mean square while $y_{t+h}-\mu$ and $\mathcal{G}_{t,h}$ have bounded second moments, the Cauchy–Schwarz inequality gives $B_h = \mathbb{E}[(y_{t+h}-\mu)\mathcal{G}_t^{\infty}] + o(1) = \operatorname{Cov}(y_{t+h},\mathcal{G}_t^{\infty}) + o(1)$, the last step because $\mathbb{E}[y_{t+h}-\mu] = 0$, and $\mathbb{E}[\mathcal{G}_{t,h}^2] = \mathbb{E}[(\mathcal{G}_t^{\infty})^2] + o(1)$, which is (S30). If $\{y_s\}$ is strongly mixing, the covariance inequality for mixing sequences bounds $|\operatorname{Cov}(y_{t+h},\mathcal{G}_t^{\infty})|$ by a multiple of a power of the mixing coefficient at lag h, since $\mathcal{G}_t^{\infty}$ is measurable with respect to the history up to t and both variables have finite moments of order higher than two under the conditions of Proposition 2; hence $\gamma_h^0 \to 0$, and the claim for γ_h follows from $\gamma_h = \gamma_h^0\,\mathbb{E}[\mathcal{G}_{t,h}^2]/A_h$.

(ii). With $c \neq 0$ the intercept part of (8) is $c\sum_{j=1}^{h}\theta_{h-j}H_{t+j-1} = cH_{t+h}\sum_{i<h}\theta_i - c\sum_{i<h}\theta_i\,(H_{t+h} - H_{t+h-i-1})$. The second sum is bounded uniformly in h, because $0 \le H_{t+h} - H_{t+h-i-1} \le (i+1)/(t+h-i-1) \le (i+1)/t$ for $i < h$ and $\sum_i i|\theta_i| < \infty$; the first equals $\mu_t H_{t+h} + O(1)$ since $\sum_{i<h}\theta_i \to 1/(1-\sum\phi_i)$ at a summable rate. With $H_{t+h} = \log(t+h) + O(1)$ and the residual part of $\mathcal{G}^{\text{fix}}$ and $\mathcal{G}^{\text{es}}$ both $O_p(1)$ as in (i), this is (S31). Then $\mathbb{E}[\mathcal{G}_{t,h}^2] = \mu_t^2\log^2((t+h)/t) + O(\log((t+h)/t))$ diverges when-

ever $\mathbb{P}(\mu_t \neq 0) > 0$, while $B_h = \mathbb{E}[e^{(0)}_{t+h}\mathcal{G}_{t,h}] = \log((t+h)/t)\,\mathbb{E}[e^{(0)}_{t+h}\mu_t] + O(1)$ with $\mathbb{E}[e^{(0)}_{t+h}\mu_t] \to \mathbb{E}[(y_{t+h} - \mu)\mu_t] = \mathrm{Cov}(y_{t+h}, \mu_t) \to 0$ under the same mixing condition, so $B_h = O(1)$ and $\gamma^0_h = O(1/\log((t+h)/t))$. The size of the drift at the main text's horizons is arithmetic: $\log(212/200) = 0.058$.

REMARK 8: The drift in (ii) is a property of Type-II fractional integration applied to a series with a non-zero mean, not of the data: $(1-L)^{-d}$ truncated at the first observation turns a constant into a sequence growing like $s^d/\Gamma(1+d)$, so the unrestricted forecast of a level series is not mean-reverting at long horizons. It is invisible at $h \leq 12$ with $t \geq 200$ and does not touch any result of the main text, but it is why the horizon prediction of the proposition cannot be tested on these data, and why the interval of improvement need not narrow at the horizons available (Table S8).

Q.E.D.

S1.10. *Proof of Proposition S2*

Write $\mathcal{O}(\kappa) = \bigcup_r \big(\mathcal{O}(\kappa) \cap \mathcal{R}_r\big)$, where $\mathcal{R}_r$ is the set of dates in regime r, and let $n_r(\kappa) = |\mathcal{O}(\kappa) \cap \mathcal{R}_r|$, so that $w_r(\kappa) = n_r(\kappa)/P(\kappa)$.

Decomposing the mean differential,

$$\bar{\Delta}(\kappa) = \frac{1}{P(\kappa)} \sum_r \sum_{t \in \mathcal{O}(\kappa) \cap \mathcal{R}_r} \Delta_{t+h} = \sum_r w_r(\kappa) \left[\frac{1}{n_r(\kappa)} \sum_{t \in \mathcal{O}(\kappa) \cap \mathcal{R}_r} \Delta_{t+h} \right]. \tag{S20}$$

Within regime r the summands have common mean μ_r and, being h-step loss differentials, are $(h-1)$-dependent up to the estimation effect; a law of large numbers for mixing arrays applies to each bracket, giving $\mu_r + o_p(1)$ as $n_r(\kappa) \to \infty$. Substituting into (S20) yields the first equality of (S32), and the second is the definition of the studentised statistic.

For the sign claim, consider two regimes with $\mu_1 > 0 > \mu_2$ and write $w_2 = 1 - w_1$. Then

$$\bar{\Delta} = w_1\mu_1 + (1-w_1)\mu_2 = \mu_2 + w_1(\mu_1 - \mu_2), \tag{S21}$$

which is continuous and strictly increasing in w_1 because $\mu_1 > \mu_2$. It vanishes at

$$w_1^* = \frac{-\mu_2}{\mu_1 - \mu_2} \in (0, 1). \tag{S22}$$

Since $\kappa \mapsto w_1(\kappa)$ is a step function that increases as κ decreases whenever regime 1 occupies the earlier part of the evaluation span, $w_1(\kappa)$ traverses a neighbourhood of w_1^* for some range of κ, and $\bar{\Delta}(\kappa)$ changes sign there. Neither the models, the data, the statistic nor the reference distribution has changed.

REMARK 9: The construction requires only that the regimes not be nested inside a single tail of the evaluation span; if every date in regime 2 precedes every date in regime 1 and κ can only add earlier origins, then $w_1(\kappa)$ is monotone and the crossing is attained at a unique κ. This is the configuration observed in Section 6.4.

Q.E.D.

S1.11. *Proof of Proposition S3*

Write $a = \tau(d_{\mathrm{G}}) - 1$ and $b = \tau(d_{\mathrm{W}}) - 1$. By Proposition 2 and the hypothesis, $b < 0 < a$, so a and b have opposite signs. Then

$$\mathrm{range}_{\mathcal{E}}(\tau) = |a - b| = |a| + |b| > \max\{|a|, |b|\}, \tag{S23}$$

the inequality being strict because $b \neq 0$. The final expression is the operator effect of Definition 1 evaluated at the more favourable of the two estimators. Only the opposite signs are used, which is why straddling is sufficient but not necessary: any configuration with $\tau(d_{\mathrm{G}}) > 1 > \tau(d_{\mathrm{W}})$ satisfies (S23) regardless of where 2γ lies.

REMARK 10: The converse fails. If both ratios exceed unity—the fractional layer is worse under both plug-ins—then $\mathrm{range} = |a - b| < \max\{|a|, |b|\}$ and the design margin is smaller than the operator effect. That happens at $h = 12$ on headline inflation, where Table I shows both columns above one. The paper's claim is therefore conditional: the design dominates the operator where the operator effect changes sign across the design cube, and not elsewhere.

Q.E.D.

S1.12. *Recovering the pseudo-optimal memory from two loss ratios*

This subsection derives the identity used in Section 6.5. Define the observable loss ratios $r_{\mathrm{G}} = \mathrm{MSFE}_1(\hat{d}_{\mathrm{G}})/\,\mathrm{MSFE}_0$ and $r_{\mathrm{W}} = \mathrm{MSFE}_1(\hat{d}_{\mathrm{W}})/\,\mathrm{MSFE}_0$, and the scaled curvature $c := A/\,\mathrm{MSFE}_0$. Dividing (S14) by MSFE_0 and evaluating at each plug-in gives the pair

$$r_{\mathrm{G}} - 1 = c\big(d_{\mathrm{G}}^2 - 2\gamma\, d_{\mathrm{G}}\big) + O(d^3), \tag{S24}$$

$$r_{\mathrm{W}} - 1 = c\big(d_{\mathrm{W}}^2 - 2\gamma\, d_{\mathrm{W}}\big) + O(d^3). \tag{S25}$$

Two equations, two unknowns (γ, c). Provided $r_{\mathrm{W}} \neq 1$, set $\rho := (r_{\mathrm{G}} - 1)/(r_{\mathrm{W}} - 1)$. Dividing (S24) by (S25) eliminates c:

$$d_{\mathrm{G}}^2 - 2\gamma\, d_{\mathrm{G}} \;=\; \rho\big(d_{\mathrm{W}}^2 - 2\gamma\, d_{\mathrm{W}}\big), \tag{S26}$$

and solving (S26) for γ,

$$\boxed{\gamma \;=\; \frac{d_{\mathrm{G}}^2 - \rho\, d_{\mathrm{W}}^2}{2\big(d_{\mathrm{G}} - \rho\, d_{\mathrm{W}}\big)}} \qquad \text{whenever } d_{\mathrm{G}} \neq \rho\, d_{\mathrm{W}}. \tag{S27}$$

Back-substituting into (S24),

$$c \;=\; \frac{r_{\mathrm{G}} - 1}{d_{\mathrm{G}}\big(d_{\mathrm{G}} - 2\gamma\big)}. \tag{S28}$$

Three conditions must hold for (S27)–(S28) to identify the pair, and each is checked unit by unit in Section 6.5 rather than assumed:

1. $d_{\mathrm{G}} \neq d_{\mathrm{W}}$. The two plug-ins must actually differ; if they agreed, the two equations would coincide and the system would be rank deficient. Empirically they differ by roughly a factor of two in the median, so this is not binding.
2. $d_{\mathrm{G}} \neq \rho\, d_{\mathrm{W}}$, the denominator of (S27). Cases violating it are discarded.
3. $c > 0$. A recovered $c \leq 0$ means either that the curvature is not positive in that cell, which Section S5 finds in a fifth of cells, or that both loss ratios sit within numerical noise of one; in both cases the quadratic recovery is uninformative. We report the

share of unit-configurations satisfying $c > 0$ (69 per cent on headline, 78 on core) and exclude the remainder rather than forcing a solution.

REMARK 11: This recovery is not a test. Because the parabola is forced through the origin and fitted to exactly two points, a pair of loss ratios straddling unity implies a second root between the two plug-ins as a matter of algebra. The test in Section 6.5 instead uses $\mathcal{G}$ and $\mathcal{H}$ estimated from the forecast map by finite differences, which carry information that no loss ratio contains.

Q.E.D.

S2. MONTE CARLO CALIBRATION: SIZE, POWER AND FAMILY-WISE SIZE

Section 7.1 of the main text describes the design and discusses the results; this section carries the size table (Table S2), the power table (Table S3), the family-wise table (Table S4) and Figure S1. Everything is run under both plug-in estimators. The design, the seeds and the timings are documented in `code/README.md` of the replication archive, and `paperD_mc_size.py` reproduces the study.

Four details behind the statements of Section 7.1. First, the long-horizon distortion under $\hat{d}_{\mathrm{G}}$ and its absence under $\hat{d}_{\mathrm{W}}$ go with the dispersion of the studentised statistic: the median Clark–West statistic under the null at $h = 1$ falls from -0.03 under $\hat{d}_{\mathrm{G}}$ to -0.34 under $\hat{d}_{\mathrm{W}}$, the local-Whittle branch delivering a less variable plug-in, hence a less variable wedge and a differential with less autocovariance beyond the truncation lag. Second, under a level break the fractional variant attains the lower loss in 999 of 1,000 replications under $\hat{d}_{\mathrm{G}}$ and 991 under $\hat{d}_{\mathrm{W}}$, with median loss ratios of 0.881 and 0.895; the rejection is about accuracy, and the break-preserving bootstraps cannot undo it because they resample a differential that under a break favours the fractional model. Third, in Table S3 the plug-in tracks the truth closely above $d = 0.2$, so the low power reflects the size of the loss differential rather than a failure of estimation. Fourth, in Table S4 the probability of at least one Benjamini–Hochberg discovery on a memory-free panel, which should be five per cent, lies inside $[3, 7]$ in five of the sixteen cells: at $h = 1$ the bootstraps over-reject, with family-wise rates of 9.2 to 16.8 per cent against 4.8 and 2.5 for the asymptotic reference, and at $h = 12$ the

TABLE S2

MONTE CARLO REJECTION RATES AT THE NOMINAL FIVE PER CENT, 1,000 REPLICATIONS, 310 OBSERVATIONS, UNDER BOTH PLUG-IN ESTIMATORS. THE FIRST THREE PROCESSES HAVE $d = 0$ AND THE ENTRIES ARE SIZES; THE FOURTH IS AN ALTERNATIVE AND THE ENTRIES ARE POWER. $\hat{d}$ IS THE MEDIAN PLUG-IN ACROSS REPLICATIONS. ON THE TWO MEMORY-FREE PROCESSES WITHOUT A BREAK, SIZES OF EIGHT PER CENT OR MORE ARE IN BOLD.

	log-periodogram plug-in $\hat{d}_{\mathrm{G}}$					local-Whittle plug-in $\hat{d}_{\mathrm{W}}$				
Process	$\hat{d}$	$N(0,1)$	MBB	Brk	Spc	$\hat{d}$	$N(0,1)$	MBB	Brk	Spc
Horizon $h = 1$										
AR(2), no memory, no break	$+0.005$	5.7	6.7	6.1	6.8	$+0.033$	2.6	4.7	3.6	3.9
+ level break	$+0.535$	98.9	97.0	96.1	90.5	$+0.347$	97.9	97.1	92.9	88.6
+ variance break	-0.005	5.1	6.2	5.3	5.2	$+0.033$	2.6	3.7	2.6	2.8
fractional, $d = 0.20$	$+0.207$	10.7	13.8	12.5	9.5	$+0.232$	13.7	16.1	13.4	10.5
Horizon $h = 12$										
AR(2), no memory, no break	$+0.007$	**9.8**	**9.1**	6.0	6.7	$+0.034$	6.5	5.5	4.1	5.3
+ level break	$+0.519$	98.9	95.6	91.4	88.1	$+0.335$	97.7	96.6	92.0	89.5
+ variance break	-0.001	**8.4**	**9.3**	5.0	5.4	$+0.033$	6.0	5.6	3.6	4.7
fractional, $d = 0.20$	$+0.206$	12.9	10.0	9.2	8.4	$+0.233$	12.9	10.6	8.9	9.0

ordering reverses, 17.5 and 17.0 for the normal against 3.2 and 4.2 for the break-preserving bootstrap. Under a level break every reference flags 26 to 31 of the 31 units.

S2.1. *Robustness of the discovery counts*

Multiplicity is controlled within configuration by the Benjamini and Hochberg (1995) procedure, whose validity requires a positive-dependence condition that common inflation shocks may violate. Under the dependence-robust Benjamini and Yekutieli (2001) correction, which is valid under arbitrary dependence at the cost of a $\sum_{i \le n} i^{-1}$ penalty, the counts of Table S9 fall from 24 of 80 to 6 of 80 under the asymptotic reference and from 7, 2 and 1 to *zero* under all three bootstraps. The qualitative ordering is unchanged and the discoveries are, if anything, more concentrated in the asymptotic reference—which the calibration study then identifies as the miscalibrated one at the long horizon under $\hat{d}_{\mathrm{G}}$.

The break-preserving bootstraps operate on the Clark–West loss differential. They segment its mean, fit an autoregressive sieve to the within-segment demeaned series, and simulate under the null with segment-specific innovation variances; what they preserve is therefore the second-order structure of the *differential*, not of the underlying series. They say nothing about whether the plug-in memory parameter is itself stable over the evaluation span, and the level-break results in Section 7.1 of the main text show why that distinction matters: when a break makes the fractional variant more accurate, resampling the differential cannot undo the rejection.

TABLE S3

POWER AGAINST A FRACTIONAL ALTERNATIVE AT $h = 1$, 1,000 REPLICATIONS, UNDER BOTH PLUG-IN ESTIMATORS. THE LOG-PERIODOGRAM PLUG-IN IS CLOSE TO UNBIASED THROUGHOUT; THE LOCAL-WHITTLE PLUG-IN IS BIASED UPWARD AT SMALL d AND CLOSE TO UNBIASED ABOVE 0.2. NEITHER REACHES HALF AT ANY POINT OF THE GRID, SO THE LOW POWER IS NOT A FAILURE OF ESTIMATION.

	log-periodogram plug-in $\hat{d}_{\mathrm{G}}$					local-Whittle plug-in $\hat{d}_{\mathrm{W}}$				
true d	$\hat{d}$	$N(0,1)$	MBB	Brk	Spc	$\hat{d}$	$N(0,1)$	MBB	Brk	Spc
0.05	0.056	5.8	6.8	5.3	4.8	0.083	3.1	4.3	3.3	3.2
0.10	0.109	6.3	7.4	6.2	4.7	0.132	6.2	7.3	6.2	4.2
0.20	0.207	10.7	13.8	12.5	9.5	0.232	13.7	16.1	13.4	10.5
0.30	0.309	18.0	21.4	18.1	15.7	0.336	22.8	25.2	22.0	19.6
0.40	0.412	26.3	29.1	26.0	21.9	0.439	32.5	33.8	30.8	25.2
0.60	0.628	40.7	41.0	37.8	31.8	0.645	44.0	45.0	41.4	35.8

TABLE S4

FAMILY-WISE BEHAVIOUR UNDER BOTH PLUG-INS: 31 UNITS, HALF THE INNOVATION VARIANCE COMMON, BENJAMINI–HOCHBERG AT FIVE PER CENT, 400 PANEL REPLICATIONS. "ANY" IS THE SHARE OF PANELS PRODUCING AT LEAST ONE DISCOVERY, WHICH UNDER THE GLOBAL NULL OF THE FIRST BLOCK SHOULD BE FIVE PER CENT; "COUNT" IS THE MEAN NUMBER OF UNITS FLAGGED, OUT OF 31. ENTRIES IN THE FIRST BLOCK OUTSIDE $[3, 7]$ PER CENT ARE IN BOLD.

			$N(0,1)$		MBB		Break-BS		Spec-BS	
Process	Plug-in	h	any	count	any	count	any	count	any	count
AR(2), no memory	$\hat{d}_{\mathrm{G}}$	1	4.8	0.07	**13.5**	0.20	**10.8**	0.14	**16.8**	0.24
		12	**17.5**	0.26	5.0	0.08	3.2	0.04	**8.0**	0.10
	$\hat{d}_{\mathrm{W}}$	1	**2.5**	0.05	**9.5**	0.13	**9.2**	0.14	**13.5**	0.21
		12	**17.0**	0.28	5.2	0.10	4.2	0.09	**9.5**	0.17
+ level break	$\hat{d}_{\mathrm{G}}$	1	100.0	30.68	100.0	30.23	100.0	29.53	99.5	26.59
		12	100.0	30.57	100.0	29.36	99.2	27.84	98.5	25.64
	$\hat{d}_{\mathrm{W}}$	1	100.0	30.08	100.0	29.75	100.0	28.62	100.0	26.54
		12	100.0	30.17	100.0	29.60	100.0	28.06	99.2	26.59

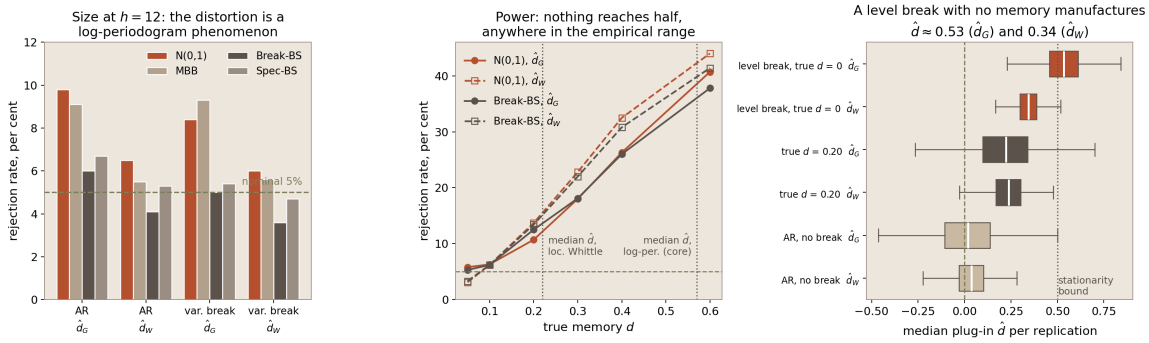

FIGURE S1.—Monte Carlo calibration of the four reference distributions under both plug-in estimators, 1,000 replications of 310 observations through the same routine used on the data. Left: size at $h = 12$ on the two memory-free processes. Under the log-periodogram plug-in $\hat{d}_{\mathrm{G}}$ the asymptotic normal and the moving-block bootstrap reject at roughly twice the nominal rate; under the local-Whittle plug-in $\hat{d}_{\mathrm{W}}$ the same references are close to nominal, so the long-horizon distortion is a property of the plug-in and not of the reference alone. Centre: power against a fractional alternative under each plug-in, with the magnitudes observed in our data marked; the wider bandwidth buys a few points and no reference reaches half anywhere in that range. Right: the distribution of the plug-in by process and estimator—a level shift in a process with *no* memory produces $\hat{d}$ around 0.53 under $\hat{d}_{\mathrm{G}}$ and 0.34 under $\hat{d}_{\mathrm{W}}$, and our own core estimates lie inside that span.

S3. ORDER OF THE FORECAST WEDGE

The main text states the result at the end of Section 6.5; this is the underlying table, produced by `paperD_order.py`, stage 5.5 of the replication archive. For the first twelve units of the common headline panel we take the history available at $\kappa = 0.70$, apply the recursive seasonal adjustment, select the order, and compute the one-step forecast wedge at a sequence of plug-ins that halve, in three ways: re-estimating the autoregressive coefficients on the filtered series, as the procedure of Section 2 does; holding the intercept and slopes at their restricted values, so that only the two binomial filters act; and the first-order term $d\,\mathcal{G}^{\mathrm{fix}}_{t,1}$ of Proposition 1. Wedges are scaled by the standard deviation of the adjusted series. The fixed-coefficient wedge computed through the forecasting map and through the identity (9) agree to 10^{-14} in every cell. In both the re-estimated and the fixed column the ratio converges to the halving ratio 0.50 as d shrinks, so the wedge is $O(d)$ rather than $O(d^2)$. The fixed column is strongly non-linear in d above 0.1, because its intercept part $-c\sum_{k\le t}\pi_k(d) = c\,(1-\pi_t(d-1))$ saturates while its first-order term cdH_t does not, and in several units the two parts cross in sign, which is why the fixed wedge at $d = 0.4$ is smaller than at $d = 0.2$. Re-estimation does not create the first-order effect; it attenuates it, by about a factor of three (0.030 against 0.093 at $d = 0.1$, 0.049 against 0.133 at $d = 0.2$), which is the most direct explanation available for why the median loss ratios of Table I of the main text sit so close to unity. On all 31 units at $h = 1$ the intercept part cH_t of $\mathcal{G}^{\mathrm{fix}}$ has median absolute size 1.41 standard deviations of the adjusted series, the residual part $\sum_k \hat{\varepsilon}_{t+1-k}/k$ 0.76, and the re-estimated sensitivity $\mathcal{G}$ 0.21 (`out_D_order_decomp.csv`): the intercept term is the larger of the two fixed-coefficient components, in 25 of the 31 units, and refitting removes the truncation term and most of the rest. What the decomposition does not identify is how much of each component the re-estimation term removes separately, $\mathcal{G}^{\mathrm{es}}$ being a single quantity, reported against the total in Section S1.5.

S4. ESTIMATOR AND BANDWIDTH

Section 6.4 of the main text summarises this comparison. The two plug-ins are used at conventional settings that differ in the objective function and in the number of retained frequencies at once, so comparing them as ordinarily run cannot say which of the two is doing the work. Table S6 crosses the two estimators with the two bandwidth rules, on the same units, origins and seasonal adjustment.

TABLE S5

ORDER OF THE FORECAST WEDGE. MEDIAN OVER TWELVE UNITS OF $|\hat{y}^{(1)}(d) - \hat{y}^{(0)}|/\mathrm{sd}(y)$ AT $h = 1$, WITH THE AUTOREGRESSIVE COEFFICIENTS RE-ESTIMATED ON THE FILTERED SERIES AND HELD AT THEIR RESTRICTED VALUES, AND THE FIRST-ORDER TERM $d\,\mathcal{G}^{\mathrm{fix}}$ OF PROPOSITION 1. A RATIO NEAR 0.50 AS d HALVES INDICATES $O(d)$; A RATIO NEAR 0.25 WOULD INDICATE $O(d^2)$.

	coefficients re-estimated		coefficients held fixed		first order
d	wedge	ratio	wedge	ratio	$d\,\mathcal{G}^{\mathrm{fix}}$
0.400	0.0748	—	0.0748	—	0.5060
0.200	0.0488	0.652	0.1331	1.780	0.2530
0.100	0.0298	0.611	0.0933	0.701	0.1265
0.050	0.0153	0.514	0.0541	0.579	0.0633
0.025	0.0077	0.505	0.0292	0.540	0.0316

TABLE S6

THE NESTED PAIR EVALUATED WITH THE MEMORY PARAMETER SUPPLIED BY EACH COMBINATION OF ESTIMATOR AND BANDWIDTH, ON THE 31 HEADLINE UNITS AT $\kappa = 0.70$. $\hat{d}$ IS THE MEDIAN ACROSS UNITS OF THE PER-UNIT MEDIAN PLUG-IN; THE RATIO IS THE CROSS-SECTIONAL MEDIAN OF $\mathrm{MSFE}_1 \,/\, \mathrm{MSFE}_0$; "WINS" COUNTS THE UNITS IN WHICH THE FRACTIONAL VARIANT ATTAINS THE LOWER SQUARED ERROR.

		$h = 1$			$h = 12$		
Estimator	m	$\hat{d}$	ratio	wins	$\hat{d}$	ratio	wins
log-periodogram	$T^{0.50}$	0.452	0.9954	17	0.457	1.0169	10
	$T^{0.65}$	0.283	0.9902	20	0.263	1.0087	13
local Whittle	$T^{0.50}$	0.412	0.9982	17	0.409	1.0146	12
	$T^{0.65}$	0.220	0.9920	21	0.215	1.0036	13

Most of the gap is bandwidth. Between the two plug-ins as conventionally run—log-periodogram at $T^{0.50}$, local Whittle at $T^{0.65}$—the median plug-in falls from 0.45 to 0.22 at $h = 1$; holding the estimator fixed and moving only the bandwidth accounts for about 0.17 of that 0.23, and holding the bandwidth fixed and moving only the estimator for the remaining 0.06; at $h = 12$ the split is 0.19 against 0.05. The loss ratios and win counts move the same way and by correspondingly small amounts. The difference between the plug-ins is therefore mostly a choice of how many frequencies to retain rather than a choice

between two semiparametric principles, and bandwidth is itself a design choice in the sense of Definition 1 of the main text: nothing fixes the exponent at 0.50 rather than 0.65, and moving it within the conventional range moves the memory estimate by about three times what switching estimator does. Data-driven bandwidth rules exist (Henry, 2001, Andrews and Guggenberger, 2003, Andrews and Sun, 2004), and Baillie, Kapetanios, and Papailias (2014) choose the bandwidth by out-of-sample forecast performance; the exponents used here are the conventional ones.

Table S18 and Figure S4 sweep the exponent a of $m = \lfloor T^a \rfloor$ from 0.45 to 0.80 for both estimators on the headline panel at $\kappa = 0.70$ (the core panel is in `out_D_bandwidth_sweep_summary.csv`), and add a data-driven rule that, at each origin, uses the exponent whose fractional forecast has the lowest cumulative squared error over the forecasts whose targets have been observed at that origin, the conventional exponent until the first such loss is available; at $h = 12$ the score therefore lags the origin by twelve months, and the choice uses no future information. Over the sweep the median plug-in falls from 0.41 to 0.24 under the log-periodogram estimator and from 0.39 to 0.13 under local Whittle on headline at $h = 1$ (from 0.61 to 0.14 and 0.57 to 0.03 on core), but not monotonically: both paths rise between 0.45 and 0.50, and the log-periodogram path again between 0.70 and 0.75. At a common exponent the two estimators still differ, by 0.02 to 0.14 on headline and 0.02 to 0.13 on core, about 0.10 in the median, so the bandwidth is the larger of the two margins over this range without being the only one; at the conventional pair of settings the decomposition of Table S6 stands. The loss ratio does not move monotonically either, the widest excursion being $\hat{d}_{\mathrm{G}}$ at $h = 12$, where the median ratio is 1.001 at one exponent and 1.084 at another. The forecast-selected exponent averages 0.61 to 0.63 across origins and units; at $h = 1$ its ratios and win counts sit inside the range the fixed exponents span, at $h = 12$, where the selection score lags the origin by twelve months, they sit in the upper part of that range (1.043 against 1.001 to 1.084 under the log-periodogram estimator, 1.014 against 1.003 to 1.015 under local Whittle), so on these data the rule buys nothing over a conventional exponent.

S5. ACCURACY OF THE LOCAL APPROXIMATION

Section 6.5 of the main text summarises this check. Proposition 2 expands around $d = 0$, while the plug-ins in the data have medians between 0.22 and 0.57. For every unit and

horizon at $\kappa = 0.70$ we compute $\mathcal{G}_{t,h}$ and $\mathcal{H}_{t,h}$ by central finite difference through the forecasting map itself, including the refitting step, and compare the exact loss difference, obtained by rerunning the rolling exercise at each point of a d-grid, with the quadratic, with and without the curvature term. Table S7 reports the comparison on the four fifths of cells with $A > 0$, so that the two coefficients are compared on identical data, and the right panel of Figure S2 plots it. Restoring the curvature term cuts the median relative error from 0.13 to 0.03 at $d = 0.10$ and from 0.32 to 0.13 at $d = 0.20$, and improves sign agreement at every point of the grid; beyond $d \approx 0.4$ the corrected version diverges faster in magnitude, as a series correct to second order should once the third-order term dominates, while still agreeing in sign more often. The usable range is narrow: even corrected, the expansion is accurate to a few per cent only for $d \lesssim 0.10$, to about thirteen per cent at $d = 0.20$, and beyond $d \approx 0.3$ its magnitude is not informative. The median local-Whittle plug-in sits at the edge of that range and the median log-periodogram plug-in well outside it, so claims that depend on the location of 2γ apply to the smaller plug-in and not to the larger.

S6. ADDITIONAL TABLES AND FIGURES

Table S8 disaggregates by horizon the pseudo-optimal memory recovered from the two loss ratios and the projection coefficient estimated from the forecast map; Section 6.5 of the main text states what it shows and why the horizon prediction of Proposition S1 is not supported by it. Figure S2 places the core and headline results side by side. Table S9 counts the discoveries by reference distribution, Table S10 reports the sign-condition diagnostic, Figure S3 shows the discoveries, the two plug-ins unit by unit and the pooled confidence set, Table S11 is the confidence-set robustness grid, Table S12 the dispersion of the unit-level loss ratio, Table S13 the pooled confidence set with the joint-estimation variant of Section S9 as a sixth procedure, Table S14 the sign-condition diagnostic of Table S10 with the endpoint estimated on the first half of the origins and the rule evaluated on the second, Tables S15 and S16 the accuracy comparison and the pooled confidence set on the window ending in 2019M12, and Table S17 the Clark–West statistic next to the unadjusted Diebold–Mariano statistic on both windows.

Section 6.5 of the main text says that the sign-condition diagnostic of Table S10 fails in three ways; the details are these. In the fifth of unit–horizon cells where $A \leq 0$ the loss difference is locally concave, no interior optimum exists, and the interval of Proposition 2

TABLE S7

ACCURACY OF THE LOCAL QUADRATIC, WITH AND WITHOUT THE CURVATURE TERM OF PROPOSITION 2. "NAIVE" USES $\mathbb{E}[\mathcal{G}^2]$ AS THE COEFFICIENT ON d^2; "CORRECTED" USES $A = \mathbb{E}[\mathcal{G}^2 - e^{(0)}\mathcal{H}]$. BOTH COLUMNS ARE COMPUTED ON THE SAME CELLS—THE 196 OF 244 UNIT–HORIZON CELLS (80.3 PER CENT) ON WHICH $A > 0$, SO THAT THE TWO COEFFICIENTS ARE COMPARED ON IDENTICAL DATA. ENTRIES ARE MEDIANS OF THE RELATIVE ERROR AGAINST THE EXACT LOSS DIFFERENCE, AND SHARES OF CELLS AGREEING IN SIGN.

	median |rel. error|		sign agreement	
d	naive	corrected	naive	corrected
0.05	0.058	**0.007**	95.9%	**100.0**%
0.10	0.133	**0.030**	96.9%	**99.5**%
0.15	0.213	**0.071**	95.9%	**97.4**%
0.20	0.317	**0.133**	93.4%	**94.4**%
0.30	0.528	**0.331**	85.7%	**88.8**%
0.40	0.769	**0.639**	80.6%	**85.7**%
0.50	**0.916**	1.114	73.0%	**78.1**%
0.60	**1.006**	1.648	70.9%	**73.5**%
0.80	**1.430**	2.566	68.9%	**70.9**%

says nothing at all. Strict bracketing, with $\hat{d}_{\mathrm{W}}$ inside the interval and $\hat{d}_{\mathrm{G}}$ outside it, holds in a seventh to a third of cells depending on index and coefficient, and the projection coefficient $\hat{\gamma}^0$ (medians 0.24 on headline and 0.27 on core) is only weakly related to the pseudo-optimal memory recovered from the two loss ratios (correlations 0.16 and 0.45). And the horizon prediction of Proposition S1 is unsupported: in Table S8 the recovered estimate narrows on headline from 0.20 at $h = 1$ to 0.14 at $h = 12$, while $\hat{\gamma}^0$, which is not constructed from the quantity being explained, moves only from 0.26 to 0.23 and is non-monotone on core. Section S7.1 shows that twelve months is too short for that limit to bite: at $h \leq 12$ with $t \geq 200$ the sensitivity is in its finite-horizon regime, and the two long-horizon regimes of Proposition S1 are below a tenth of a percentage point.

The pooled confidence set of Table II of the main text and of Table S11 is built on a balanced calendar. Pooling by date requires a fixed cross-section, and the panel does not supply one: the United Kingdom's series ends in 2020M11 and the United States' begins in 2002M01, so averaging over whatever units exist at each date gives a series whose compo-

sition drifts, and a block bootstrap over it resamples composition along with dependence. We therefore pool over units sharing an identical set of target dates, taking the set that maximises units times dates. That keeps 29 units in both indices, setting aside the United Kingdom and the United States on headline and the United Kingdom on core for this table only, and gives a pooled calendar from 2018M03 to 2025M12 at $h = 1$, starting later as the horizon lengthens. The block length is the larger of the cube root of the series length and the horizon, so that blocks span the overlap of consecutive h-step forecasts; there are 2,000 replications at the ten per cent level and the origin rule is $\kappa = 0.70$. The test uses the range statistic, and at each step the model with the largest worst-case standardised loss difference is eliminated. Every model still in the set when the first non-rejection occurs is reported with that test's p-value, so the two members of a surviving pair share it; the columns `t_ar_frac` and `p_ar_frac` of the archived table give the studentised pooled loss difference between the autoregression and its fractional extension, positive when the fractional variant has the lower loss, with its two-sided p-value from the same block bootstrap. Across the sixteen cells the statistic lies between -1.42 and $+1.36$ and the p-value is never below 0.14; when the surviving set is exactly the pair, the range statistic is $|t|$ and the two p-values coincide. Loosening the origin rule to $\kappa = 0.40$ roughly doubles the pooled series, to 187 dates on headline inflation at $h = 1$, without separating the nested pair in any of the 32 resulting cells.

Table S13 adds the joint-estimation variant to the set. It is retained with the pair in 14 of the 16 cells and eliminated in the two core cells at $h = 6$; its pooled loss exceeds the autoregression's in every cell, by a studentised margin between 0.40 and 2.09. Table S14 reports the sign-condition diagnostic with the endpoint estimated on the first half of each cell's origins and the rule evaluated on the second: skill over the base rate lies between -0.12 and $+0.02$ and Youden's index below 0.17 in all eight cells, whereas the same rule with the endpoint estimated on the evaluation half itself has skill of $+0.09$ to $+0.24$, and the full-sample block reproduces Table S10 exactly. The halves are short and the split coincides with the 2021–23 surge; on this evidence the in-sample skill is the fit of the endpoint to the errors it classifies.

Tables S15 and S16 take the surge out of the sample altogether. On the window ending in 2019M12 the evaluation span at $\kappa = 0.70$ runs from 2014 to 2019, the years of low and stable inflation before the surge, and the balanced calendar of the pooled set runs from

2014M01 (headline, $h = 1$) to 2019M12 with 58 to 72 dates, keeping 30 units on headline (the United States is set aside; the United Kingdom, whose series ends in 2020M11, is now retained) and all 30 on core. Under $\hat{d}_{\mathrm{G}}$ the fractional layer is two to six per cent worse in the median, wins a majority of units in 2 of 20 headline configurations and in none on core, and the median plug-in is 0.40 on headline and 0.50 on core; under $\hat{d}_{\mathrm{W}}$ the median ratio stays within about one per cent of unity with 15 and 1 majority wins. The estimation-free benchmarks change more than the pair does: the restricted model attains 0.14 to 0.39 of the random walk's squared error at horizons up to six months and 0.30 to 0.58 of the Atkeson–Ohanian rule's, but 0.95 to 1.02 of the seasonal naive rule's on headline and 1.12 to 1.14 on core, and the pooled confidence set eliminates both members of the pair in all eight core cells, where the seasonal naive rule survives alone, while keeping them together in seven of the eight headline cells (the autoregression is eliminated at $h = 6$ under $\hat{d}_{\mathrm{W}}$, $p = 0.092$). Discoveries after false-discovery-rate control appear in 53 of the 76 configurations with at least 40 origins (the rule $\kappa = 0.80$ at $h = 12$ leaves too few on the shorter window) under the asymptotic reference, 58 under the moving-block bootstrap, 35 under the break-preserving bootstrap and 29 under the spectral one, against 24, 7, 2 and 1 of 80 on the full window; all 27 spectral discoveries under $\hat{d}_{\mathrm{G}}$ and all 39 under $\hat{d}_{\mathrm{W}}$ are in unit–configuration cells where the fractional variant attains the lower loss, so the shorter window widens the cross-section of outcomes rather than shifting its centre. The origin-rule reversal at the annual horizon of Section 6.4 of the main text is the range column of Table S15 at $h = 12$ under $\hat{d}_{\mathrm{G}}$.

Table S17 answers the question left open in Section 7 of the main text: how much of the discovery count is the Clark–West adjustment. The unadjusted statistic, on the same errors and with the same bootstrap draws, is smaller than the adjusted one in all but one of the 4,960 unit–configuration cells of the full window, with a median of -0.02 against 0.66, and it yields one discovery under the asymptotic reference, one under the moving-block bootstrap and none under the other two, against 24, 7, 2 and 1 configurations for Clark–West: on the full window the discovery count is the adjustment. The 45 cells the asymptotic Clark–West reference flags all have a realised loss ratio below one; the adjustment does not reject where the larger model forecasts worse, it makes small improvements significant by adding the squared forecast difference to the differential. On the pre-2020 window, where the cross-section of outcomes is wider, the unadjusted statistic rejects in 22, 29, 18 and 16 of 76 configurations against 53, 58, 35 and 29; there 5 of the 197 cells flagged by the

asymptotic Clark–West reference and 29 of the 273 flagged by the moving-block bootstrap have a ratio above one, cells in which the adjusted statistic rejects because the two forecasts differ, not because the larger model forecasts better. The ordering of the four references is the same under both statistics. The counts of Table S17 control the false-discovery rate within each configuration and then count configurations, with no control across them. Stage 7.9 of the archive (`out_D_global_fdr.csv`) applies Benjamini–Hochberg and Benjamini–Yekutieli once across all cells of a window. On the full window, $2,436$ cells, no cell survives under any reference for either statistic. On the pre-2020 window, $2,316$ cells, Benjamini–Hochberg keeps 151 cells in 58 configurations and 20 units under the asymptotic Clark–West reference and 233 under the moving-block bootstrap, 40 under the asymptotic unadjusted statistic, and none under the break-preserving and spectral bootstraps, whose p-values cannot fall below 0.001 with 999 replications and so cannot pass a global threshold of that order; Benjamini–Yekutieli leaves 33 and 10 under the asymptotic references and none elsewhere. Table S19 checks Proposition 3 cell by cell. The expansion has the sign of the squared forecast difference in every cell and the sign of the two differentials in 68 to 94 per cent of cells, but its magnitude only under the smaller plug-in, with median relative errors of 0.24 to 0.42 under $\hat{d}_{\mathrm{W}}$ against 0.80 to 3.65 under $\hat{d}_{\mathrm{G}}$, as Section S5 leads one to expect at plug-ins of 0.45 to 0.56. The qualitative prediction is realised: the adjusted mean is positive while the unadjusted one is negative in 25 per cent of cells under $\hat{d}_{\mathrm{G}}$ and 14 per cent under $\hat{d}_{\mathrm{W}}$, against 30 and 15 predicted, and the adjusted mean is positive in 65 and 75 per cent of cells against 40 and 61 for the unadjusted one.

TABLE S8

RECOVERED PSEUDO-OPTIMAL MEMORY AND PROJECTION COEFFICIENT, BY HORIZON. $\hat{\gamma}^{\,\mathrm{rec}}$ IS RECOVERED FROM THE TWO LOSS RATIOS (MEDIAN OVER UNIT–CONFIGURATIONS SATISFYING THE CONVEXITY CONDITION OF SECTION S1.12); $\hat{\gamma}^{0} = \hat{B}/\hat{\mathbb{E}}[\mathcal{G}^2]$ IS THE PROJECTION COEFFICIENT OF PROPOSITION S1, ESTIMATED FROM THE DERIVATIVE OF THE FORECAST MAP. PLUG-IN MEDIANS ARE AT $\kappa = 0.70$.

Index	h	$\hat{\gamma}^{\,\mathrm{rec}}$	$\hat{\gamma}^{0}$	$2\hat{\gamma}^{0}$	median $\hat{d}_{\mathrm{W}}$	median $\hat{d}_{\mathrm{G}}$
Headline	1	0.203	0.259	0.518	0.219	0.450
	3	0.214	0.246	0.492	0.219	0.455
	6	0.182	0.227	0.455	0.221	0.461
	12	0.136	0.232	0.465	0.216	0.459
Core	1	0.213	0.283	0.567	0.229	0.586
	3	0.219	0.233	0.467	0.216	0.562
	6	0.256	0.346	0.692	0.232	0.581
	12	0.241	0.287	0.575	0.243	0.564

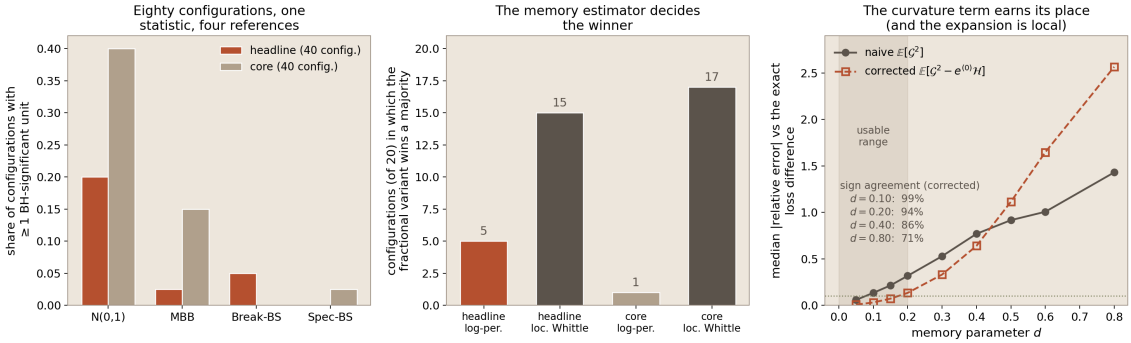

FIGURE S2.—Left: share of configurations yielding at least one discovery, by reference distribution, headline and core. Centre: configurations, of twenty per cell, in which the fractional variant wins a majority of units, by index and memory estimator—the design margin of Table III. Right: the accuracy study of Table S7. Restoring the curvature term $\mathbb{E}[e^{(0)}\mathcal{H}]$ reduces the median relative error by a factor of four at $d = 0.10$ and of two and a half at $d = 0.20$, and improves sign agreement at every point of the grid; beyond $d \approx 0.4$ the corrected expansion diverges faster in magnitude, as a second-order-correct series should.

TABLE S9

CONFIGURATIONS WITH AT LEAST ONE UNIT SIGNIFICANT AT FIVE PER CENT AFTER BENJAMINI–HOCHBERG CONTROL. SAME STATISTIC, SAME FORECAST ERRORS, FOUR REFERENCE DISTRIBUTIONS.

Reference distribution	Headline (of 40)	Core (of 40)	Total (of 80)
Asymptotic $N(0,1)$	8	16	**24**
Moving-block bootstrap	1	6	7
Break-preserving parametric bootstrap	2	0	2
Spectral break-preserving bootstrap	0	1	**1**

TABLE S10

SIGN-CONDITION DIAGNOSTIC WITH TWO ESTIMATES OF THE INTERVAL ENDPOINT. "NAIVE" USES THE PROJECTION COEFFICIENT $\hat{\gamma}^0 = \hat{B}/\hat{\mathbb{E}}[\mathcal{G}^2]$; "CORRECTED" USES $\hat{\gamma} = \hat{B}/\hat{A}$ WITH $A = \mathbb{E}[\mathcal{G}^2 - e^{(0)}\mathcal{H}]$. BOTH COMBINE THE FORECAST-MAP DERIVATIVES WITH THE RESTRICTED MODEL'S FORECAST ERRORS AND NEVER USE THE UNRESTRICTED FORECAST OR ITS LOSS. "BASE RATE" IS THE ACCURACY OF ALWAYS PREDICTING THE MAJORITY OUTCOME; SKILL IS ACCURACY MINUS BASE RATE; J IS YOUDEN'S INDEX, ZERO FOR AN UNINFORMATIVE CLASSIFIER.

Index	Coefficient	Plug-in	n	accuracy	base rate	skill	J
Headline	naive $\mathbb{E}[\mathcal{G}^2]$	$\hat{d}_{\mathrm{W}}$	124	0.806	0.605	$+0.202$	0.595
		$\hat{d}_{\mathrm{G}}$	124	0.669	0.532	$+0.137$	0.343
	corrected A	$\hat{d}_{\mathrm{W}}$	98	0.765	0.571	$+0.194$	0.512
		$\hat{d}_{\mathrm{G}}$	98	**0.776**	0.592	$\mathbf{+0.184}$	**0.566**
Core	naive $\mathbb{E}[\mathcal{G}^2]$	$\hat{d}_{\mathrm{W}}$	120	0.817	0.608	$+0.208$	0.570
		$\hat{d}_{\mathrm{G}}$	120	0.675	0.658	$+0.017$	0.295
	corrected A	$\hat{d}_{\mathrm{W}}$	98	**0.898**	0.612	$\mathbf{+0.286}$	**0.785**
		$\hat{d}_{\mathrm{G}}$	98	0.745	0.684	$+0.061$	0.402

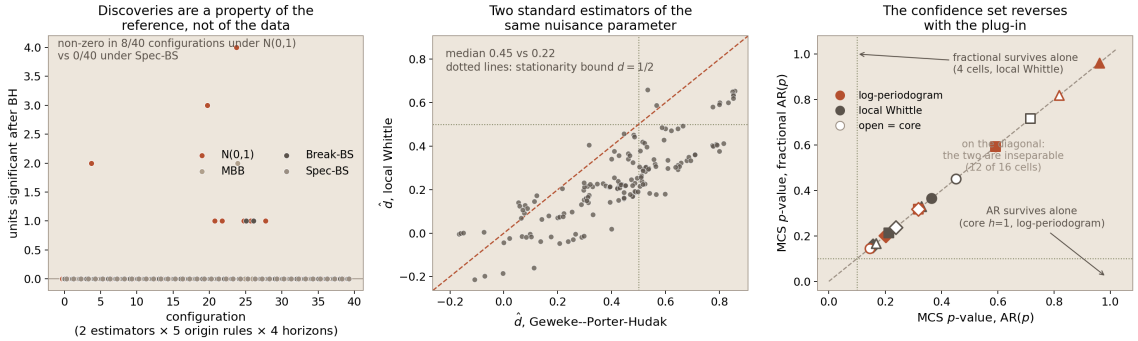

FIGURE S3.—Left: units significant after Benjamini–Hochberg control in each of the forty headline configurations, by reference distribution. Centre: the two semiparametric plug-in estimates of the same memory parameter, unit by unit; dotted lines mark the stationarity bound $d = 1/2$, which the log-periodogram plug-in crosses in a substantial minority of units. Right: the dependence-aware model confidence set of Table II, one point per index—plug-in–horizon cell, plotting the p-value of the autoregression against that of its fractional extension. The pair survives together in every cell, so under the reporting convention of Section 6.3 of the main text the two p-values coincide and every point lies on the diagonal; the studentised loss difference between the two members, which measures their distance, lies between -1.42 and $+1.36$ (`out_D_mcs_pooled_v2.csv`).

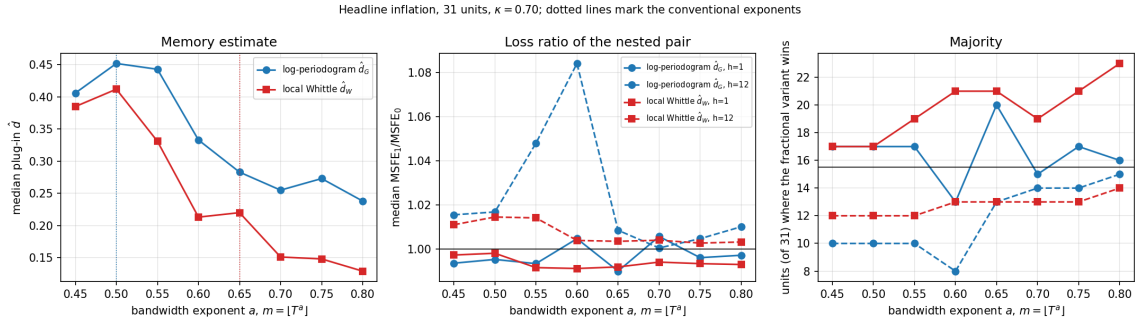

FIGURE S4.—The bandwidth exponent as a continuous design margin, headline inflation, 31 units, $\kappa = 0.70$. Left: the median plug-in against the exponent for the two estimators, with the conventional exponents marked. Centre: the cross-sectional median loss ratio of the nested pair at $h = 1$ (solid) and $h = 12$ (dashed). Right: the number of units in which the fractional variant attains the lower loss; the horizontal line is a majority.

TABLE S11

POOLED MODEL CONFIDENCE SET UNDER NINE NORMALISATION AND BLOCK-LENGTH RULES, SIXTEEN INDEX–PLUG-IN–HORIZON CELLS EACH, ALL ON THE BALANCED CALENDAR OF SECTION 6.3. "PAIR TIED" COUNTS CELLS IN WHICH THE AUTOREGRESSION AND ITS FRACTIONAL EXTENSION ARE BOTH RETAINED; THE NEXT TWO COLUMNS COUNT THE CELLS IN WHICH ONE IS RETAINED AND THE OTHER EXCLUDED. THE FINAL COLUMN COUNTS CELLS WHOSE ENTIRE SURVIVING SET MATCHES THE HEADLINE SPECIFICATION OF TABLE II, THE FIRST ROW.

Loss scaled by	Block	Pair tied	AR kept	frac. kept	Same set as row 1
seasonal naive	$\max\{\lceil n^{1/3}\rceil, h\}$	16	0	0	16
seasonal naive	$\lceil n^{1/3}\rceil$	16	0	0	15
seasonal naive	$2h$	15	1	0	13
none (raw)	$\max\{\lceil n^{1/3}\rceil, h\}$	16	0	0	16
none (raw)	$\lceil n^{1/3}\rceil$	16	0	0	12
none (raw)	$2h$	15	1	0	11
AR(p)	$\max\{\lceil n^{1/3}\rceil, h\}$	16	0	0	16
AR(p)	$\lceil n^{1/3}\rceil$	16	0	0	12
AR(p)	$2h$	16	0	0	13
all 144 configurations		142	2	0	—

TABLE S12

DISTRIBUTION OF THE UNIT-LEVEL LOSS RATIO $\mathrm{MSFE}_1 / \mathrm{MSFE}_0$ ACROSS ALL UNIT–CONFIGURATION CELLS. "HELPS" AND "HURTS" ARE THE SHARES BELOW $1-b$ AND ABOVE $1+b$ FOR THE STATED BAND b.

		quantiles of $\mathrm{MSFE}_1 / \mathrm{MSFE}_0$					outside band b		
Index	cells	min	q_{05}	median	q_{95}	max	$b = 2\%$	$b = 5\%$	$b = 10\%$
Headline	1,238	0.794	0.912	0.9995	1.101	1.525	56.0%	26.7%	8.8%
Core	1,198	0.631	0.906	1.0011	1.072	1.532	47.1%	19.4%	7.4%

TABLE S13

POOLED MODEL CONFIDENCE SET AT THE TEN PER CENT LEVEL WITH THE JOINT-ESTIMATION VARIANT OF SECTION S9 ("CSS") AS A SIXTH PROCEDURE, ON THE BALANCED CALENDAR OF 29 UNITS OF TABLE II, SEED 77, 2,000 REPLICATIONS. ENTRIES ARE MCS p-VALUES UNDER THE REPORTING CONVENTION OF SECTION 6.3 OF THE MAIN TEXT; SURVIVORS IN BOLD. THE LAST COLUMN IS THE STUDENTISED POOLED LOSS DIFFERENCE BETWEEN THE AUTOREGRESSION AND THE JOINT VARIANT, NEGATIVE WHEN THE JOINT VARIANT HAS THE LARGER LOSS, WITH ITS TWO-SIDED BOOTSTRAP p-VALUE.

Index	Plug-in	h	dates	AR(p)	frac. AR(p)	CSS	RW	AO	S-nv	AR vs CSS: t (p)
Headline	$\hat{d}_{\mathrm{G}}$	1	94	**0.664**	**0.664**	**0.664**	<0.001	<0.001	0.062	−0.71 (0.480)
		3	92	**0.758**	**0.758**	**0.758**	<0.001	<0.001	0.085	−0.40 (0.690)
		6	89	**0.140**	**0.140**	**0.140**	<0.001	<0.001	**0.140**	−1.21 (0.226)
		12	83	**0.169**	**0.169**	**0.169**	**0.169**	<0.001	**0.169**	−1.95 (0.054)
	$\hat{d}_{\mathrm{W}}$	1	94	**0.362**	**0.362**	**0.362**	<0.001	<0.001	0.069	−0.71 (0.480)
		3	92	**0.207**	**0.207**	**0.207**	<0.001	<0.001	0.088	−0.40 (0.690)
		6	89	**0.146**	**0.146**	**0.146**	<0.001	<0.001	**0.146**	−1.21 (0.226)
		12	83	**0.145**	**0.145**	**0.145**	**0.145**	<0.001	**0.145**	−1.95 (0.054)
Core	$\hat{d}_{\mathrm{G}}$	1	90	**0.292**	**0.292**	**0.292**	<0.001	<0.001	0.009	−1.06 (0.292)
		3	88	**0.137**	**0.137**	**0.137**	<0.001	<0.001	0.003	−1.83 (0.066)
		6	85	**0.818**	**0.818**	0.078	<0.001	<0.001	0.012	−2.09 (0.035)
		12	79	**0.163**	**0.163**	**0.163**	**0.163**	<0.001	**0.163**	−1.98 (0.057)
	$\hat{d}_{\mathrm{W}}$	1	90	**0.545**	**0.545**	**0.545**	<0.001	<0.001	0.008	−1.06 (0.292)
		3	88	**0.125**	**0.125**	**0.125**	<0.001	<0.001	0.003	−1.83 (0.066)
		6	85	**0.167**	**0.167**	0.064	<0.001	<0.001	0.012	−2.09 (0.035)
		12	79	**0.146**	**0.146**	**0.146**	**0.146**	<0.001	**0.146**	−1.98 (0.057)

TABLE S14

THE SIGN-CONDITION DIAGNOSTIC OUT OF SAMPLE. IN EACH UNIT–HORIZON CELL AT $\kappa = 0.70$ THE ORIGINS ARE SPLIT IN HALF (ABOUT 44 EACH); THE ENDPOINT $\hat{\gamma}^0 = \hat{B}/\hat{\mathbb{E}}[\mathcal{G}^2]$ (NAIVE) OR $\hat{\gamma} = \hat{B}/\hat{A}$ (CORRECTED) IS ESTIMATED ON THE FIRST HALF, AND THE RULE "THE PLUG-IN LIES IN $(0, 2\hat{\gamma})$" IS EVALUATED WITH THE SECOND HALF'S MEDIAN PLUG-IN AGAINST THE SECOND HALF'S REALISED SIGN OF $\mathrm{MSFE}_1 / \mathrm{MSFE}_0 - 1$. THE SECOND BLOCK ESTIMATES THE ENDPOINT ON THE SECOND HALF ITSELF; THE THIRD REPRODUCES TABLE S10. COLUMNS AS IN TABLE S10.

Index	Coefficient	Plug-in	n	accuracy	base rate	skill	J
endpoint on first half, rule on second							
Headline	naive	$\hat{d}_{\mathrm{W}}$	124	0.621	0.597	+0.024	0.170
Headline	naive	$\hat{d}_{\mathrm{G}}$	124	0.508	0.532	−0.024	0.034
Headline	corrected	$\hat{d}_{\mathrm{W}}$	105	0.533	0.610	−0.076	0.103
Headline	corrected	$\hat{d}_{\mathrm{G}}$	105	0.457	0.505	−0.048	-0.090
Core	naive	$\hat{d}_{\mathrm{W}}$	120	0.525	0.650	−0.125	0.082
Core	naive	$\hat{d}_{\mathrm{G}}$	120	0.525	0.575	−0.050	0.010
Core	corrected	$\hat{d}_{\mathrm{W}}$	81	0.506	0.617	−0.111	0.126
Core	corrected	$\hat{d}_{\mathrm{G}}$	81	0.556	0.568	−0.012	0.012
endpoint and rule on second half							
Headline	naive	$\hat{d}_{\mathrm{W}}$	124	0.790	0.597	+0.194	0.571
Headline	naive	$\hat{d}_{\mathrm{G}}$	124	0.669	0.532	+0.137	0.341
Headline	corrected	$\hat{d}_{\mathrm{W}}$	92	0.793	0.554	+0.239	0.565
Headline	corrected	$\hat{d}_{\mathrm{G}}$	92	0.793	0.630	+0.163	0.624
Core	naive	$\hat{d}_{\mathrm{W}}$	120	0.792	0.650	+0.142	0.548
Core	naive	$\hat{d}_{\mathrm{G}}$	120	0.667	0.575	+0.092	0.338
Core	corrected	$\hat{d}_{\mathrm{W}}$	104	0.865	0.644	+0.221	0.767
Core	corrected	$\hat{d}_{\mathrm{G}}$	104	0.769	0.587	+0.183	0.476
full sample (Table S10)							
Headline	naive	$\hat{d}_{\mathrm{W}}$	124	0.806	0.605	+0.202	0.595
Headline	naive	$\hat{d}_{\mathrm{G}}$	124	0.669	0.532	+0.137	0.343
Headline	corrected	$\hat{d}_{\mathrm{W}}$	98	0.765	0.571	+0.194	0.512
Headline	corrected	$\hat{d}_{\mathrm{G}}$	98	0.776	0.592	+0.184	0.566
Core	naive	$\hat{d}_{\mathrm{W}}$	120	0.817	0.608	+0.208	0.570
Core	naive	$\hat{d}_{\mathrm{G}}$	120	0.675	0.658	+0.017	0.295
Core	corrected	$\hat{d}_{\mathrm{W}}$	98	0.898	0.612	+0.286	0.785
Core	corrected	$\hat{d}_{\mathrm{G}}$	98	0.745	0.684	+0.061	0.402

TABLE S15

FORECAST ACCURACY ON THE WINDOW 2000M01–2019M12, THE ANALOGUE OF TABLE I OF THE MAIN TEXT: THE SAME 31 HEADLINE AND 30 CORE UNITS, PLUG-INS, ORIGIN RULES, HORIZONS AND SEEDS, WITH THE ESTIMATION AND EVALUATION SAMPLES ENDING BEFORE 2020. COLUMNS AS IN TABLE I. ON THIS WINDOW THE FRACTIONAL VARIANT WINS A MAJORITY OF UNITS IN 2 (HEADLINE) AND 0 (CORE) OF 20 CONFIGURATIONS UNDER $\hat{d}_{\mathrm{G}}$ AND IN 15 AND 1 UNDER $\hat{d}_{\mathrm{W}}$; THE MEDIAN PLUG-INS ARE 0.40 AND 0.19 ON HEADLINE AND 0.50 AND 0.16 ON CORE.

		$\mathrm{MSFE}_1 / \mathrm{MSFE}_0, \hat{d}_{\mathrm{G}}$		$\mathrm{MSFE}_1 / \mathrm{MSFE}_0, \hat{d}_{\mathrm{W}}$		MSFE_0 relative to		
Index	h	med.	range	med.	range	RW	AO	S-nv
Headline	1	1.038	[1.004, 1.041]	0.998	[0.996, 1.012]	0.30	0.54	0.95
	3	1.050	[1.002, 1.055]	0.994	[0.993, 1.009]	0.24	0.56	0.97
	6	1.034	[0.995, 1.065]	0.995	[0.989, 0.999]	0.39	0.55	0.99
	12	1.058	[0.981, 1.097]	0.995	[0.989, 0.997]	1.02	0.58	1.02
Core	1	1.036	[1.023, 1.038]	1.011	[1.010, 1.020]	0.14	0.30	1.14
	3	1.029	[1.017, 1.033]	1.004	[1.004, 1.008]	0.14	0.31	1.13
	6	1.028	[1.019, 1.031]	1.003	[1.001, 1.008]	0.24	0.31	1.12
	12	1.024	[1.015, 1.026]	1.001	[0.993, 1.005]	1.13	0.30	1.13

TABLE S16

POOLED MODEL CONFIDENCE SET AT THE TEN PER CENT LEVEL ON THE WINDOW 2000M01–2019M12, THE ANALOGUE OF TABLE II OF THE MAIN TEXT, ON A BALANCED CALENDAR OF THE UNITS SHARING AN IDENTICAL SET OF TARGET DATES (FROM 2014-01 TO 2019M12 AT $h = 1$; 58 TO 72 DATES), SEED 77, 2,000 REPLICATIONS, $\kappa = 0.70$. ENTRIES AND CONVENTIONS AS IN TABLE II; THE LAST COLUMN IS THE STUDENTISED POOLED LOSS DIFFERENCE BETWEEN THE TWO MEMBERS OF THE NESTED PAIR, POSITIVE WHEN THE FRACTIONAL VARIANT HAS THE LOWER LOSS, WITH ITS TWO-SIDED BOOTSTRAP p-VALUE.

Index	Plug-in	h	dates	AR(p)	frac. AR(p)	RW	AO	S-nv	AR vs frac.: t (p)
Headline	$\hat{d}_{\mathrm{G}}$	1	72	**0.161**	**0.161**	<0.001	<0.001	**0.161**	−1.55 (0.123)
		3	70	**0.183**	**0.183**	<0.001	<0.001	**0.183**	−0.06 (0.956)
		6	67	**0.274**	**0.274**	<0.001	<0.001	**0.274**	+0.41 (0.701)
		12	61	**0.263**	**0.263**	**0.263**	<0.001	**0.263**	+1.40 (0.177)
	$\hat{d}_{\mathrm{W}}$	1	72	**0.294**	**0.294**	<0.001	<0.001	**0.294**	−0.43 (0.660)
		3	70	**0.269**	**0.269**	<0.001	<0.001	**0.269**	+0.91 (0.360)
		6	67	0.092	**0.231**	<0.001	<0.001	**0.231**	+1.90 (0.050)
		12	61	**0.184**	**0.184**	**0.184**	<0.001	**0.184**	+1.65 (0.087)
Core	$\hat{d}_{\mathrm{G}}$	1	69	0.002	<0.001	<0.001	<0.001	**1.000**	−2.21 (0.031)
		3	67	<0.001	<0.001	<0.001	<0.001	**1.000**	−0.91 (0.369)
		6	64	<0.001	<0.001	<0.001	<0.001	**1.000**	−0.36 (0.724)
		12	58	<0.001	<0.001	**1.000**	<0.001	**1.000**	+0.61 (0.543)
	$\hat{d}_{\mathrm{W}}$	1	69	0.002	0.002	<0.001	<0.001	**1.000**	−0.80 (0.426)
		3	67	<0.001	<0.001	<0.001	<0.001	**1.000**	+0.27 (0.779)
		6	64	<0.001	<0.001	<0.001	<0.001	**1.000**	+0.61 (0.535)
		12	58	<0.001	<0.001	**1.000**	<0.001	**1.000**	+0.95 (0.345)

TABLE S17

THE CLARK–WEST STATISTIC AND THE UNADJUSTED DIEBOLD–MARIANO STATISTIC ON THE SAME FORECAST ERRORS, WITH THE SAME FIXED-LAG STUDENTISATION, REFERENCE DISTRIBUTIONS, BOOTSTRAP DRAWS (SEEDS 4000 PLUS THE UNIT INDEX) AND BENJAMINI–HOCHBERG CONTROL, OVER THE 80 CONFIGURATIONS OF EACH WINDOW. "CONFIGS" COUNTS CONFIGURATIONS WITH AT LEAST ONE DISCOVERY; "CELLS" COUNTS UNIT–CONFIGURATION DISCOVERIES; "BOTH" THE CELLS FLAGGED BY BOTH STATISTICS; ON THE PRE-2020 WINDOW 76 OF THE 80 CONFIGURATIONS HAVE AT LEAST 40 ORIGINS; THE LAST TWO COLUMNS COUNT FLAGGED CELLS WHOSE REALISED LOSS RATIO $\mathrm{MSFE}_1 / \mathrm{MSFE}_0$ EXCEEDS ONE, WHICH THE UNADJUSTED STATISTIC CANNOT PRODUCE.

	configs (of 80)		cells			cells with ratio > 1	
Reference	CW	DM	CW	DM	both	CW	DM
window 2000–2025							
Asymptotic $N(0,1)$	24	1	45	1	1	0	0
Moving-block bootstrap	7	1	17	1	1	2	0
Break-preserving bootstrap	2	0	2	0	0	0	0
Spectral bootstrap	1	0	1	0	0	0	0
window 2000–2019							
Asymptotic $N(0,1)$	53	22	197	72	70	5	0
Moving-block bootstrap	58	29	273	80	80	29	0
Break-preserving bootstrap	35	18	81	28	28	1	0
Spectral bootstrap	29	16	66	32	32	0	0

TABLE S18

THE NESTED PAIR ON THE 31 HEADLINE UNITS AT $\kappa = 0.70$ WITH THE MEMORY PARAMETER SUPPLIED BY EACH ESTIMATOR AT BANDWIDTH $m = \lfloor T^a \rfloor$, $a = 0.45, \ldots, 0.80$, AND BY THE EXPONENT SELECTED AT EACH ORIGIN FROM PAST FORECAST PERFORMANCE. $\hat{d}$ IS THE MEDIAN ACROSS UNITS OF THE PER-UNIT MEDIAN PLUG-IN, "RATIO" THE CROSS-SECTIONAL MEDIAN OF $\mathrm{MSFE}_1 / \mathrm{MSFE}_0$ AND "WINS" THE NUMBER OF UNITS IN WHICH THE FRACTIONAL VARIANT ATTAINS THE LOWER LOSS.

	log-periodogram						local Whittle					
	$h = 1$			$h = 12$			$h = 1$			$h = 12$		
a	$\hat{d}$	ratio	wins	$\hat{d}$	ratio	wins	$\hat{d}$	ratio	wins	$\hat{d}$	ratio	wins
0.45	0.406	0.9937	17	0.440	1.0156	10	0.385	0.9974	17	0.395	1.0111	12
0.50	0.452	0.9954	17	0.457	1.0169	10	0.412	0.9982	17	0.409	1.0146	12
0.55	0.443	0.9935	17	0.402	1.0480	10	0.331	0.9917	19	0.319	1.0142	12
0.60	0.333	1.0048	13	0.300	1.0841	8	0.213	0.9913	21	0.196	1.0040	13
0.65	0.283	0.9901	20	0.263	1.0086	13	0.220	0.9920	21	0.215	1.0036	13
0.70	0.255	1.0058	15	0.216	1.0006	14	0.151	0.9942	19	0.121	1.0041	13
0.75	0.273	0.9962	17	0.261	1.0048	14	0.148	0.9935	21	0.137	1.0028	13
0.80	0.238	0.9973	16	0.265	1.0102	15	0.129	0.9931	23	0.124	1.0033	14
selected by past forecasts	0.324	0.9904	22	0.341	1.0426	11	0.265	0.9913	17	0.238	1.0144	12

TABLE S19

PROPOSITION 3 ON THE DATA. FOR EVERY UNIT–HORIZON CELL AT $\kappa = 0.70$ AND EACH PLUG-IN, THE REALISED CELL MEANS OF THE SQUARED FORECAST DIFFERENCE $(\hat{y}^{(0)} - \hat{y}^{(1)})^2$, OF THE ADJUSTED DIFFERENTIAL f^{CW} AND OF THE UNADJUSTED ONE f^{DM} ARE COMPARED WITH THE EXPANSIONS (S18), (S19) AND (S17) EVALUATED AT THE PER-ORIGIN PLUG-IN, WITH $\mathcal{G}$ IN CLOSED FORM (SECTION S1.5) AND $\mathcal{H}$ BY CENTRAL DIFFERENCE. "REL. ERR." IS THE MEDIAN ABSOLUTE RELATIVE ERROR, "SIGN" THE PERCENTAGE OF CELLS ON WHICH THE EXPANSION HAS THE REALISED SIGN. THE LAST FOUR COLUMNS ARE PERCENTAGES OF CELLS: REALISED $\mathbb{E}f^{\mathrm{CW}} > 0$, REALISED $\mathbb{E}f^{\mathrm{DM}} > 0$, AND REALISED AND PREDICTED $\mathbb{E}f^{\mathrm{CW}} > 0 > \mathbb{E}f^{\mathrm{DM}}$, THE CELLS IN WHICH THE ADJUSTMENT ALONE DECIDES THE SIGN.

				$(\hat{y}^{(0)} - \hat{y}^{(1)})^2$		f^{CW}		f^{DM}		share of cells, %			
Index	Plug-in	cells	med. $\hat{d}$	rel. err.	sign	rel. err.	sign	rel. err.	sign	CW>0	DM>0	CW$>0>$DM	pred.
Headline	$\hat{d}_{\mathrm{G}}$	124	0.46	0.86	100	0.80	82	0.89	85	71	47	24	23
	$\hat{d}_{\mathrm{W}}$	124	0.22	0.46	100	0.24	91	0.28	94	69	60	9	8
Core	$\hat{d}_{\mathrm{G}}$	120	0.56	3.65	100	1.86	68	2.54	76	60	34	26	36
	$\hat{d}_{\mathrm{W}}$	120	0.22	0.77	100	0.31	94	0.42	84	80	61	19	22

S7. THREE FURTHER PROPOSITIONS

Section 3 of the main text summarises three results; their exact statements are recorded here, with the proofs in Sections S1.9–S1.11.

S7.1. *The horizon*

The pseudo-optimal memory is horizon-specific, because $e^{(0)}$ and $\mathcal{G}$ both are. The proposition below says what happens to it as the horizon grows, with the long-horizon sensitivity in closed form from Section S1.5; the main text reports that the prediction is not supported at the horizons available, and Table S20 shows that it could not be.

PROPOSITION S1—Horizon behaviour: *Write $\gamma_h = B_h/A_h$ with A_h, B_h as in (10), suppose $A_h > 0$, and let $\gamma_h^0 := B_h/\mathbb{E}[\mathcal{G}_{t,h}^2]$, so that $\gamma_h = \gamma_h^0\,\mathbb{E}[\mathcal{G}_{t,h}^2]/A_h$. Then γ_h^0 is the population coefficient of the linear projection of the restricted model's h-step error on the sensitivity $\mathcal{G}_{t,h}$. Suppose that the restricted model is stationary with a covariance-stationary error, that its h-step forecast converges in mean square to its unconditional mean μ as h grows, and that $\sum_i i|\theta_i| < \infty$ for its impulse responses.*

(i) Demeaned history. *If the adjusted history is demeaned at each origin, so that the restricted intercept is zero, then $\mathcal{G}_{t,h}^{\mathrm{fix}} \to 0$ and*

$$\mathcal{G}_{t,h} \longrightarrow \mathcal{G}_t^{\infty} = \frac{\dot{c}_t}{1 - \sum_i \phi_i} \qquad (h \to \infty), \tag{S29}$$

in mean square, where $\dot{c}_t$ is the derivative of the refitted intercept with respect to the filter, the first entry of $\dot{\beta}$ in (S10): the long-horizon sensitivity is the intercept channel of the refit, a quadratic functional of the history up to t. Then $e_{t+h}^{(0)}$ approaches the demeaned target and

$$\gamma_h^0 = \frac{\mathrm{Cov}\big(y_{t+h},\, \mathcal{G}_t^{\infty}\big)}{\mathbb{E}\big[(\mathcal{G}_t^{\infty})^2\big]} + o(1) \qquad \text{as } h \to \infty, \tag{S30}$$

the projection of a target on a fixed function of the distant past. If $\{y_s\}$ is strongly mixing with $\mathbb{E}[(\mathcal{G}_t^{\infty})^2] > 0$, that covariance tends to zero, so $\gamma_h^0 \to 0$; if $\mathbb{E}[\mathcal{G}_{t,h}^2]$ stays bounded and A_h stays bounded away from zero, then $\gamma_h \to 0$ as well and the interval of improvement $(0, 2\gamma_h)$ collapses.

(ii) Level history with an intercept, the map as implemented. *If the restricted intercept* c *is not zero, then*

$$\mathcal{G}_{t,h} \;=\; \mu_t \,\log\frac{t+h}{t} + O_p(1) \qquad (h\to\infty), \qquad \mu_t = \frac{c}{1-\sum_i \phi_i}: \tag{S31}$$

the Type-II re-integration of the intercept drifts with the horizon, $\mathcal{G}_{t,h}$ *has no limit,* $\mathbb{E}[\mathcal{G}_{t,h}^2]$ *grows like* $\log^2((t+h)/t)$ *while* B_h *stays bounded, and* $\gamma_h^0 \to 0$ *at the rate* $1/\log((t+h)/t)$. *At the horizons of the main text,* $h \le 12$ *with* $t \ge 200$, $\log((t+h)/t) \le 0.06$ *and the first-order drift of the unrestricted forecast,* $\hat{d}\,\mu_t \log((t+h)/t)$, *is below a tenth of a percentage point.*

The proof is in Section S1.9. A plug-in of fixed magnitude should move from inside the interval of improvement to outside it as the horizon grows. Of the two estimation routes in Section 6.5 of the main text, the one that does not reuse the quantity being explained shows no such narrowing (Table S8), and nothing in the main text builds on the proposition. Table S20 shows why nothing should be visible at twelve months. Stage 5.7 of the replication archive (`paperD_horizon.py`) evaluates $\mathcal{G}_{t,h}$ in closed form at horizons up to 384 months, at the history available at $\kappa = 0.70$, on the level series and on the demeaned series, for every unit. On the level series the median $|\mathcal{G}_{t,h}|$ on headline rises from 0.21 standard deviations at $h = 1$ to 0.75 at $h = 384$, and the increase between $h = 96$ and $h = 384$ is $\mu_t \log((t+384)/(t+96))$ with a median ratio of 1.12 across units (correlation 0.88), which is (S31); on the demeaned series it falls from 0.25 to 0.14 towards the explicit limit (S29), whose median is 0.09 (median gap 0.04 at $h = 384$, 0.14 at $h = 12$, the harmonic weights decaying slowly). The first-order drift at $h = 12$ is 0.11 percentage points per unit of $\hat{d}$ in the median and 0.21 at most, so at the plug-ins of the main text it is below a tenth of a point, and both regimes of the proposition are outside the range of horizons the data allow.

S7.2. *The origin rule*

A further margin concerns the evaluation sample rather than the model.

PROPOSITION S2—Origin-rule dependence: *Partition the evaluation span into regimes* $r = 1, \dots, R^*$ *on which the loss differential has constant means* μ_r, *and let* $w_r(\kappa)$ *be the*

TABLE S20

THE SENSITIVITY AT LONG HORIZONS. MEDIAN ACROSS UNITS OF $|\mathcal{G}_{t,h}|$ SCALED BY THE STANDARD DEVIATION OF THE ADJUSTED HISTORY, AT THE ORIGIN OF $\kappa = 0.70$, IN CLOSED FORM, ON THE LEVEL SERIES (THE MAP AS IMPLEMENTED, WITH AN INTERCEPT) AND ON THE DEMEANED SERIES; THE LAST ROW IS THE EXPLICIT LIMIT (S29) OF THE DEMEANED CASE.

	Headline		Core	
h	level	demeaned	level	demeaned
1	0.214	0.253	0.141	0.142
3	0.244	0.302	0.296	0.156
6	0.309	0.274	0.245	0.229
12	0.185	0.189	0.305	0.261
24	0.262	0.187	0.262	0.182
48	0.260	0.192	0.318	0.168
96	0.328	0.168	0.282	0.142
192	0.528	0.166	0.357	0.173
384	0.747	0.136	0.502	0.147
∞ (demeaned)	—	0.088	—	0.044

share of $\mathcal{O}(\kappa)$ falling in regime r. Then

$$\bar{\Delta}(\kappa) = \sum_r w_r(\kappa)\,\mu_r + o_p(1), \qquad \mathrm{CW}(\kappa) = \sqrt{P(\kappa)}\,\frac{\bar{\Delta}(\kappa)}{\hat{\omega}_L(\kappa)}. \tag{S32}$$

Since w_r depends on κ while μ_r does not, there exist regime means for which $\bar{\Delta}(\kappa)$ changes sign as κ decreases, with no change in the models, the data, the statistic or the reference distribution.

Section 6.4 of the main text shows this in the data: on a short evaluation window a clear majority of units favoured the fractional layer, and halving κ, which only admits earlier origins, reversed both the majority and the sign of the mean loss difference. The proof is in Section S1.10.

S7.3. *Design range and the operator effect*

Proposition 2 of the main text gives the design range of Definition 1 its content.

PROPOSITION S3—Design-range dominance: *Let* $\tau(d) = \mathrm{MSFE}_1(d)/\mathrm{MSFE}_0$ *and suppose the two plug-in limits straddle the interval of improvement,* $d_{\mathrm{W}} \in (0, 2\gamma)$ *and* $d_{\mathrm{G}} \notin [0, 2\gamma]$. *Then*

$$\mathrm{range}_{\mathcal{E}}(\tau) = \big|\tau(d_{\mathrm{G}}) - \tau(d_{\mathrm{W}})\big| = \big|\tau(d_{\mathrm{G}}) - 1\big| + \big|\tau(d_{\mathrm{W}}) - 1\big| \;>\; \max_{e \in \{\mathrm{G},\mathrm{W}\}} \big|\tau(d_e) - 1\big|, \quad \text{(S33)}$$

so the variation induced by the choice of memory estimator strictly exceeds the operator effect measured at either estimator. Straddling is sufficient, not necessary: (S33) *also holds without it whenever* $\tau(d_{\mathrm{G}})$ *and* $\tau(d_{\mathrm{W}})$ *lie on opposite sides of unity.*

Proposition S3 says only that two numbers on opposite sides of one are farther from each other than either is from one. The aggregate loss ratios straddle unity in six of the eight rows of Table I of the main text. The proof is in Section S1.11.

S8. TWO REMARKS

REMARK 12—The curvature term is not negligible: Section S5 estimates $\mathcal{G}$ and $\mathcal{H}$ numerically. In the data, $|\mathbb{E}[e^{(0)}\mathcal{H}]|$ exceeds a tenth of $\mathbb{E}[\mathcal{G}^2]$ in about 91 per cent of unit–horizon cells and half of it in about 63 per cent, and $A \leq 0$, so that no interior optimum exists, in about 20 per cent. Replacing $\mathbb{E}[\mathcal{G}^2]$ by A reduces the median relative error of the quadratic from 0.13 to 0.03 at $d = 0.10$ and from 0.32 to 0.13 at $d = 0.20$.

REMARK 13: Corollary 1 is falsifiable, but not by solving (10) at the two observed loss ratios: a convex parabola through the origin fitted to two points whose ordinates straddle zero has its second root between them by construction, so that exercise restates the sign flip rather than testing it. A diagnostic that is not circular needs $\mathcal{G}$ and $\mathcal{H}$ estimated from the forecast map itself, which Section S5 does by finite differences; Section 6.5 of the main text combines them with the restricted model's forecast errors and compares the resulting sign rule with the realised losses.

S9. JOINT ESTIMATION OF THE MEMORY PARAMETER

Section 6.4 of the main text summarises this comparison. For a given d the sum of squared residuals of the autoregression fitted by least squares to $(1 - L)^d \mathcal{S}_t y$ is the conditional-sum-of-squares objective of Chung and Baillie (1993) concentrated in the autoregressive coefficients; minimising it over $d \in [-0.49, 1.20]$ at each origin gives the joint

estimate $\hat{d}_{\mathrm{C}}$, and the forecast is the same map $\mathcal{M}_1(\hat{d}_{\mathrm{C}})$ as in the main text, so the pair stays nested and only the choice of d changes, from an estimator that never looks at the autoregression to the autoregression's own one-step fit. The grid, the references, the seeds and the Benjamini–Hochberg control are those of Table I of the main text (stage 6.9 of the replication archive).

TABLE S21

THE NESTED PAIR WITH THE MEMORY PARAMETER ESTIMATED JOINTLY BY CONDITIONAL SUM OF SQUARES. COLUMNS 3–4 AS IN TABLE I OF THE MAIN TEXT: THE MEDIAN OVER THE FIVE ORIGIN RULES OF THE CROSS-SECTIONAL MEDIAN OF $\mathrm{MSFE}_1 / \mathrm{MSFE}_0$, AND ITS RANGE. "WINS" COUNTS THE ORIGIN RULES, OF FIVE, UNDER WHICH THE JOINT VARIANT ATTAINS THE LOWER LOSS IN A MAJORITY OF UNITS; $\hat{d}_{\mathrm{C}}$ IS THE MEDIAN OVER UNIT–CONFIGURATION CELLS OF THE PER-UNIT MEDIAN ESTIMATE.

Index	h	med.	range	wins	median $\hat{d}_{\mathrm{C}}$
Headline	1	1.005	[1.001, 1.009]	0 of 5	-0.015
	3	1.008	[1.001, 1.022]	0 of 5	-0.015
	6	1.010	[1.004, 1.023]	0 of 5	-0.014
	12	1.010	[1.008, 1.017]	0 of 5	-0.016
Core	1	1.005	[1.002, 1.010]	0 of 5	-0.008
	3	1.007	[1.002, 1.008]	0 of 5	-0.008
	6	1.010	[1.003, 1.017]	0 of 5	-0.009
	12	1.010	[1.008, 1.031]	0 of 5	-0.012

The joint estimate centres on zero in the median, -0.015 on headline and -0.008 on core, while the unit-level estimates remain dispersed: their 5th to 95th percentiles across unit–configuration cells are -0.48 and 0.21, and 5 per cent of cells sit at the lower bound of the search interval, which is reported rather than widened. The cross-sectional median loss ratio lies between 1.005 and 1.010 in every row and above one in all forty configurations, the joint variant wins a majority of units in none of them, and after false-discovery-rate control no unit is significant under the spectral reference in any configuration or under any reference on headline; on core, at $\kappa \leq 0.50$ only, the asymptotic normal flags at least one unit in six configurations, the moving-block bootstrap in four and the break-preserving bootstrap in three. In the pooled confidence set with the joint variant as a sixth procedure

(Table S13) it survives alongside the pair in 14 of 16 cells and is eliminated in the two core cells at $h = 6$ ($p = 0.078$ and 0.064), with a studentised pooled loss difference against the autoregression between -2.09 and -0.40, negative when the joint variant has the larger loss.

REFERENCES

Andrews, Donald W. K. and Patrik Guggenberger (2003), "A bias-reduced log-periodogram regression estimator for the long-memory parameter." *Econometrica*, 71 (2), 675–712. [20, 51]

Andrews, Donald W. K. and Yixiao Sun (2004), "Adaptive local polynomial Whittle estimation of long-range dependence." *Econometrica*, 72 (2), 569–614. [20, 51]

Atkeson, Andrew and Lee E. Ohanian (2001), "Are Phillips curves useful for forecasting inflation?" *Federal Reserve Bank of Minneapolis Quarterly Review*, 25 (1), 2–11. [13]

Baillie, Richard T. (1996), "Long memory processes and fractional integration in econometrics." *Journal of Econometrics*, 73 (1), 5–59. [1]

Baillie, Richard T., Ching-Fan Chung, and Margie A. Tieslau (1996), "Analysing inflation by the fractionally integrated ARFIMA-GARCH model." *Journal of Applied Econometrics*, 11 (1), 23–40. [2]

Baillie, Richard T., George Kapetanios, and Fotis Papailias (2014), "Bandwidth selection by cross-validation for forecasting long memory financial time series." *Journal of Empirical Finance*, 29, 129–143. [2, 20, 51]

Baillie, Richard T., Chaleampong Kongcharoen, and George Kapetanios (2012), "Prediction from ARFIMA models: Comparisons between MLE and semiparametric estimation procedures." *International Journal of Forecasting*, 28 (1), 46–53. [2, 4]

Basak, Gopal K., Ngai Hang Chan, and Wilfredo Palma (2001), "The approximation of long-memory processes by an ARMA model." *Journal of Forecasting*, 20 (6), 367–389. [2]

Benjamini, Yoav and Yosef Hochberg (1995), "Controlling the false discovery rate: A practical and powerful approach to multiple testing." *Journal of the Royal Statistical Society, Series B*, 57 (1), 289–300. [14, 45]

Benjamini, Yoav and Daniel Yekutieli (2001), "The control of the false discovery rate in multiple testing under dependency." *Annals of Statistics*, 29 (4), 1165–1188. [24, 45]

Beran, Jan (1994), *Statistics for Long-Memory Processes*, Chapman and Hall, New York. [1]

Bhardwaj, Geetesh and Norman R. Swanson (2006), "An empirical investigation of the usefulness of ARFIMA models for predicting macroeconomic and financial time series." *Journal of Econometrics*, 131 (1–2), 539–578. [2]

Bos, Charles S., Philip Hans Franses, and Marius Ooms (1999), "Long memory and level shifts: Re-analyzing inflation rates." *Empirical Economics*, 24 (3), 427–449. [2, 14, 26]

Chung, Ching-Fan and Richard T. Baillie (1993), "Small sample bias in conditional sum-of-squares estimators of fractionally integrated ARMA models." *Empirical Economics*, 18 (4), 791–806. [20, 70]

Clark, Todd E. and Michael W. McCracken (2001), "Tests of equal forecast accuracy and encompassing for nested models." *Journal of Econometrics*, 105 (1), 85–110. [3]

Clark, Todd E. and Kenneth D. West (2007), "Approximately normal tests for equal predictive accuracy in nested models." *Journal of Econometrics*, 138 (1), 291–311. [4, 6, 11, 39]

Crato, Nuno and Bonnie K. Ray (1996), "Model selection and forecasting for long-range dependent processes." *Journal of Forecasting*, 15 (2), 107–125. [2]

Diebold, Francis X. (2015), "Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of Diebold–Mariano tests." *Journal of Business & Economic Statistics*, 33 (1), 1–9. [3]

Diebold, Francis X. and Atsushi Inoue (2001), "Long memory and regime switching." *Journal of Econometrics*, 105 (1), 131–159. [14, 24, 26, 34]

Diebold, Francis X. and Roberto S. Mariano (1995), "Comparing predictive accuracy." *Journal of Business & Economic Statistics*, 13 (3), 253–263. [3, 11, 23]

Eurostat (2026), "HICP – monthly data (index), table `prc_hicp_midx`." Dataset, European Commission, Luxembourg. [12]

Faust, Jon and Jonathan H. Wright (2013), "Forecasting inflation." In *Handbook of Economic Forecasting* (Graham Elliott and Allan Timmermann, eds.), Vol. 2, 2–56, Elsevier, Amsterdam. [2]

Gadea, María Dolores and Laura Mayoral (2006), "The persistence of inflation in OECD countries: A fractionally integrated approach." *International Journal of Central Banking*, 2 (1), 51–104. [2]

Geweke, John and Susan Porter-Hudak (1983), "The estimation and application of long memory time series models." *Journal of Time Series Analysis*, 4 (4), 221–238. [6]

Giacomini, Raffaella and Halbert White (2006), "Tests of conditional predictive ability." *Econometrica*, 74 (6), 1545–1578. [3, 23]

Granger, Clive W. J. and Namwon Hyung (2004), "Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns." *Journal of Empirical Finance*, 11 (3), 399–421. [14, 24, 26, 34]

Granger, Clive W. J. and Roselyne Joyeux (1980), "An introduction to long-memory time series models and fractional differencing." *Journal of Time Series Analysis*, 1 (1), 15–29. [1, 5]

Groen, Jan J. J., Richard Paap, and Francesco Ravazzolo (2013), "Real-time inflation forecasting in a changing world." *Journal of Business & Economic Statistics*, 31 (1), 29–44. [2]

Hansen, Peter R., Asger Lunde, and James M. Nason (2011), "The model confidence set." *Econometrica*, 79 (2), 453–497. [3, 16]

Hassler, Uwe and Marc-Oliver Pohle (2023), "Forecasting under long memory." *Journal of Financial Econometrics*, 21 (3), 742–778. [2]

Hassler, Uwe and Jürgen Wolters (1995), "Long memory in inflation rates: International evidence." *Journal of Business & Economic Statistics*, 13 (1), 37–45. [1]

Henry, Marc (2001), "Robust automatic bandwidth for long memory." *Journal of Time Series Analysis*, 22 (3), 293–316. [20, 51]

Hosking, J. R. M. (1981), "Fractional differencing." *Biometrika*, 68 (1), 165–176. [1, 5]

Hualde, Javier and Fabrizio Iacone (2017), "Fixed bandwidth asymptotics for the studentized mean of fractionally integrated processes." *Economics Letters*, 150, 39–43. [22]

Hualde, Javier and Fabrizio Iacone (2019), "Fixed bandwidth inference for fractional cointegration." *Journal of Time Series Analysis*, 40 (4), 544–572. [22]

Hurvich, Clifford M., Rohit Deo, and Julia Brodsky (1998), "The mean squared error of Geweke and Porter-Hudak's estimator of the memory parameter of a long-memory time series." *Journal of Time Series Analysis*, 19 (1), 19–46. [6]

Inoue, Atsushi, Lu Jin, and Barbara Rossi (2017), "Rolling window selection for out-of-sample forecasting with time-varying parameters." *Journal of Econometrics*, 196 (1), 55–67. [3]

Kiefer, Nicholas M. and Timothy J. Vogelsang (2005), "A new asymptotic theory for heteroskedasticity-autocorrelation robust tests." *Econometric Theory*, 21 (6), 1130–1164. [22]

Kumar, Manmohan S. and Tatsuyoshi Okimoto (2007), "Dynamics of persistence in international inflation rates." *Journal of Money, Credit and Banking*, 39 (6), 1457–1479. [2]

Künsch, Hans R. (1989), "The jackknife and the bootstrap for general stationary observations." *Annals of Statistics*, 17 (3), 1217–1241. [14]

Newey, Whitney K. and Kenneth D. West (1987), "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix." *Econometrica*, 55 (3), 703–708. [6]

Perron, Pierre and Zhongjun Qu (2010), "Long-memory and level shifts in the volatility of stock market return indices." *Journal of Business & Economic Statistics*, 28 (2), 275–290. [14, 26]

Pesaran, M. Hashem and Allan Timmermann (2007), "Selection of estimation window in the presence of breaks." *Journal of Econometrics*, 137 (1), 134–161. [2]

Robinson, Peter M. (1995), "Gaussian semiparametric estimation of long range dependence." *Annals of Statistics*, 23 (5), 1630–1661. [6]

Rossi, Barbara and Atsushi Inoue (2012), "Out-of-sample forecast tests robust to the choice of window size." *Journal of Business & Economic Statistics*, 30 (3), 432–453. [2]

Shimotsu, Katsumi and Peter C. B. Phillips (2005), "Exact local Whittle estimation of fractional integration." *Annals of Statistics*, 33 (4), 1890–1933. [6]

Stock, James H. and Mark W. Watson (2007), "Why has U.S. inflation become harder to forecast?" *Journal of Money, Credit and Banking*, 39 (s1), 3–33. [2]

Tashman, Leonard J. (2000), "Out-of-sample tests of forecasting accuracy: An analysis and review." *International Journal of Forecasting*, 16 (4), 437–450. [6]

Velasco, Carlos (1999a), "Gaussian semiparametric estimation of non-stationary time series." *Journal of Time Series Analysis*, 20 (1), 87–127. [14]

Velasco, Carlos (1999b), "Non-stationary log-periodogram regression." *Journal of Econometrics*, 91 (2), 325–371. [14]

Vera-Valdés, J. Eduardo (2020), "On long memory origins and forecast horizons." *Journal of Forecasting*, 39 (5), 811–826. [2]

West, Kenneth D. (1996), "Asymptotic inference about predictive ability." *Econometrica*, 64 (5), 1067–1084. [3]

White, Halbert (2000), "A reality check for data snooping." *Econometrica*, 68 (5), 1097–1126. [16]