Local Inference and Stability Testing in Structural Time Series Models

Preliminary and incomplete - please do not circulate

Latest version available here

Robin Braun
Federal Reserve Board
robin.a.braun@frb.gov

George Kapetanios
King's College London
george.kapetanios@kcl.ac.uk

Massimiliano Marcellino
Bocconi University, IGIER, CEPR and Baffi
massimiliano.marcellino@unibocconi.it

March 27, 2026

Abstract

We develop kernel-smoothed GMM statistics for testing parameter stability and conducting local-in-time inference in structural time-series models with potentially weak instruments. The integrated kernel statistic decomposes into a full-sample component and a stability component whose distribution does not depend on identification strength, enabling a sequential workflow: test for stability first, then proceed to full-sample or local inference. Both S-type and K-type versions are available; the latter exploits overidentifying information for additional power. When stability is rejected, pointwise confidence sets for the time-varying parameter are obtained by inverting the local kernel statistics and remain valid regardless of instrument strength. In Monte Carlo experiments calibrated to a New Keynesian Phillips Curve environment, the tests control size and the local confidence sets deliver reliable coverage. In an empirical application using shock-based instruments and System Projections IV (SP-IV) estimation, we find strong evidence of instability, a progressive flattening of the Phillips curve, and increasingly forward-looking inflation dynamics.

JEL classification: C12; C14; C22; E31
Keywords: Structural stability; Weak identification; Kernel-based inference; GMM; New Keynesian Phillips Curve; Instrumental variables

1 Introduction

Structural time-series models of the form

$$y_t = x_t'\beta + u_t, \qquad t = 1, \ldots, T,$$

are a workhorse of empirical macroeconomics. A canonical example is the New Keynesian Phillips Curve (NKPC), which links inflation to real activity and expectations and remains central to modern monetary policy analysis. Estimation of β is complicated by endogeneity: the regressors x_t, such as marginal cost proxies or output-gap measures in the NKPC, are typically correlated with the disturbance u_t. A prominent response has been to incorporate external information in the form of instrumental variables built from identified structural shocks. In particular, sequences of exogenous macroeconomic shocks have been used as instruments to isolate movements in x_t that are orthogonal to u_t (Barnichon and Mesters, 2020; Lewis and Mertens, 2022). These shock-IV strategies are increasingly implemented using limited-information methods such as GMM.

Two empirical features, however, recur in these applications. First, structural relationships may evolve over time: in macroeconomic environments with changing policy regimes, evolving expectations formation, or gradual changes in propagation mechanisms, assuming constant β can be untenable. Second, shock-based instruments are often weak. Weak identification is not a second-order nuisance: it can invalidate conventional asymptotic approximations and, in time-varying environments, can interact with local estimation in ways that distort both stability diagnostics and confidence statements.

Our first contribution is to develop nonparametric stability tests for structural time-series equations that remain valid under weak identification. The starting point is the insight emphasized by Magnusson and Mavroeidis (2014): under weak instruments, size control requires testing stability of the moment conditions rather than stability of the structural parameters themselves. We extend the kernel-based test of Chen and Hong (2012) to locally smoothed orthogonality conditions in GMM settings. The resulting tests are nonparametric in the time dimension and explicitly direct power toward smooth or persistent instability, patterns that are empirically plausible in macroeconomic data. The integrated kernel statistic admits an orthogonal decomposition into a full-sample S-statistic and a stability component, analogous to the decomposition of generalized S-statistics established by Magnusson and Mavroeidis (2014) in the sample-splitting context. In the kernel setting, the stability component has a pivotal limiting distribution—either a functional of a Gaussian process or asymptotically normal—that does not depend on identification strength and directs power toward smooth or persistent departures from stability rather than discrete breaks.

Within this framework, we propose two complementary robust procedures. The first is a kernel S-type test that inherits the robustness properties of S-statistic approaches highlighted by Magnusson and Mavroeidis (2014) and remains valid without requiring

strong first-stage relevance. The second is a kernel K-type test that projects the moment conditions onto the directions identified by the Jacobian, in the spirit of Kleibergen (2005). By filtering moment directions that are weakly informative for β, the K-type approach exploits overidentifying information and can deliver power gains in overidentified models, an empirically relevant case when instruments are constructed as multiple lags of structural shocks.

Our second contribution is local-in-time inference on smoothly time-varying structural parameters. While we report local point estimates constructed from kernel-weighted moments, our emphasis is on inference. In weak-instrument environments, estimator-centered approximations can be unreliable and confidence regions can be ill-behaved. We therefore construct pointwise confidence sets by inverting locally defined robust statistics. Specifically, we develop kernel versions of S- and K-type tests and invert them to obtain confidence sets that remain valid regardless of instrument strength at each time point. These sets naturally reflect both sampling uncertainty and local identification conditions and, via the K-type transformation, can explicitly leverage overidentifying information.

In Monte Carlo experiments calibrated to the NKPC environment of Barnichon and Mesters (2020), the kernel S-based stability test controls size and exhibits reliable power against smooth instability, and the kernel K-based stability test delivers additional power in overidentified designs without sacrificing robustness when identification is weak. The resulting local confidence sets deliver reliable coverage across a range of identification scenarios, including cases where instrument relevance is limited.

We illustrate the empirical relevance of the framework in a re-examination of the NKPC. Following Lewis and Mertens (2022), we use the Main Business Cycle shock of Angeletos et al. (2020) as an instrument for aggregate demand within SP-IV estimation. SP-IV can be interpreted as a restricted GMM implementation in which the NKPC restrictions are imposed across multiple forecast-error or impulse-response horizons.

We begin by applying the proposed robust stability tests and document substantial evidence against stability: both the S- and K-based tests reject, indicating economically meaningful time variation in the underlying moment conditions. Motivated by these findings, we then conduct local-in-time inference on NKPC parameters using kernel-weighted estimation and robust test inversion. The resulting confidence sets point to a pronounced flattening of the Phillips curve over time, together with increasingly forward-looking inflation dynamics. These patterns align with prominent narratives in the literature, but our analysis provides a unified and formally robust assessment by explicitly accounting for instability and weak identification throughout.

Relation to the Literature

Our paper is related to the growing literature in empirical macroeconomics that relies on limited-information econometric methods, such as instrumental variables and GMM,

to estimate structural time-series equations in isolation, rather than as part of fully specified DSGE models. Within this literature, recent work has increasingly turned to external instruments, often narrative shocks or identified structural innovations, to address endogeneity in structural relationships. Earlier limited-information approaches that relied on lagged macroeconomic variables as instruments proved highly sensitive to simultaneity, expectations errors, and specification choices (Galí and Gertler, 1999; Galí et al., 2005; Mavroeidis et al., 2014).

In the NKPC context, Barnichon and Mesters (2020) show how sequences of identified macroeconomic shocks can be repurposed as instruments for modern macroeconomic equations. Lewis and Mertens (2022) refine this approach by proposing the system-projections IV (SP-IV) framework, which exploits impulse responses from local projections or VARs to improve efficiency and effective instrument strength. Despite these advances, weak identification remains an empirical regularity in many structural macroeconomic applications, motivating the use of robust inference. Kleibergen and Mavroeidis (2009) provide a comprehensive overview of such methods for structural time-series models, including Anderson–Rubin type procedures and conditional likelihood ratio tests (Stock and Wright, 2000; Moreira, 2003; Kleibergen, 2005). Our paper builds on this literature by jointly accounting for both parameter instability and robust inference.

The idea to allow for parameter instability in the estimation of macroeconomic equations is not new. Chen (2015) develop two-stage local linear procedures for models with potentially time-varying endogeneity, while Giraitis et al. (2021) propose nonparametric time-varying IV estimation and inference under both deterministic and stochastic coefficient evolution. Antoine and Boldea (2018) exploit time variation in the information content of instruments to improve efficiency in the NKPC context. Relative to these contributions, our emphasis is on a unified GMM framework that (i) delivers global stability tests valid under weak identification and (ii) provides pointwise local-in-time confidence sets via inversion of robust statistics, with explicit accommodation of overidentification typical when multiple lagged shocks or impulse-response components are used as instruments.

An important exception within the time-varying IV/NKPC literature is the recent work by Inoue et al. (2025), who propose a time-varying instrumental-variable approach to the NKPC that allows for both robust inference under weak instruments and time-varying parameters. Their analysis adopts local instability asymptotics for drifting parameters (Müller and Petalas, 2010) and connects to their companion work on inference in unstable environments (Inoue et al., 2024). Methodologically, inference is conducted through reduced-form objects and confidence sets are constructed using weighted-average-risk minimizing intervals designed to remain reliable when identification is weak. Our approach differs in four respects. First, we develop robust stability tests for moment conditions, which are not a focus of Inoue et al. (2025). Second, we formulate both global stability testing and local-in-time inference directly within a GMM framework, facilitating

application to overidentified macro-shock IV settings and restricted-GMM implementations such as SP-IV. Third, our local confidence sets are obtained by inverting kernel versions of the S- and K-statistics, yielding pointwise robust inference that remains valid at each time point and that can explicitly exploit overidentifying information through the K-type transformation. Fourth, our decomposition into full-sample and stability components provides a natural sequential workflow: the researcher first tests for stability and, conditional on the outcome, proceeds to either full-sample or local-in-time inference.

Given that our framework covers global stability testing and local-in-time inference, it is also related to the literature on structural change and stability testing (Andrews, 1993; Bai and Perron, 1998). In macroeconomic settings, gradual shifts in policy regimes, expectations formation, and propagation mechanisms naturally motivate alternatives characterized by smooth or persistent parameter change. Elliott and Müller (2006) develop efficient tests against general persistent time variation in a regression framework and propose the popular quasi local-level (qLL) statistic. Similarly, Chen and Hong (2012) propose kernel-based nonparametric tests for smooth structural change that explicitly direct power toward gradual departures from stability. Chen (2015) and Li et al. (2024) extend this work further to cover endogeneity in the regressor and a general moment-condition framework, respectively.

A key difference relative to our setting is that most of this literature is developed under the assumption of strong identification, and therefore does not generally guarantee correct size when the structural parameters are weakly identified or even unidentified. Notable exceptions are Caner (2007), who develops breakpoint-based structural change tests in continuous-updating GMM that remain valid under weak identification and derives conservative nuisance-parameter-free bounds for their limiting distributions, and Caner (2011), who proposes pivotal tests for linear simultaneous equation models with weak identification. In a closely related contribution, Magnusson and Mavroeidis (2014) establish an orthogonal decomposition of generalized S-statistics into full-sample and stability components and show that the stability component is pivotal under weak identification. Their framework uses sample splitting and considers several weighting functionals: the qLL statistic, which targets persistent or local-level parameter variation in the spirit of Elliott and Müller (2006), and exponential and average functionals that are particularly powerful against discrete structural breaks. We extend this logic to kernel-smoothed statistics, which direct power toward smooth or persistent instability, and complement it with K-type transformations in the sense of Kleibergen (2005) that exploit overidentifying information. We further show that the stability component remains pivotal when evaluated at a weakly identified full-sample estimator, rather than at the hypothesized value. Finally, we use the local kernel statistics for pointwise-in-time inference for structural parameters.

Finally, our empirical application speaks to the debate on whether the Phillips curve has flattened and why inflation appeared increasingly disconnected from real activity in

the pre-COVID U.S. economy (see Bergholt et al. (2024) for a recent overview). A central difficulty in this literature is distinguishing genuine structural change from artifacts of weak identification and misspecification: reduced-form evidence of flattening (Atkeson et al., 2001; Stock and Watson, 2007) can reflect a declining structural slope of the Phillips curve (Ball and Mazumder, 2011; Blanchard, 2016; Stock and Watson, 2020; Del Negro et al., 2020; Hazell et al., 2022), shifts in monetary-policy behavior (McLeay and Tenreyro, 2020), or changes in the composition and importance of supply shocks that can weaken unconditional inflation–activity comovement without requiring a flatter Phillips curve (Galí and Gambetti, 2019; Coibion and Gorodnichenko, 2015; Bergholt et al., 2024). Our framework contributes by making it possible to test for and characterize structural instability directly within the shock-IV environment used in modern NKPC estimation, while controlling for weak identification. The empirical results, which point to a progressive flattening of the slope and increasingly forward-looking dynamics, thus rest on stability tests and confidence sets whose validity does not depend on instrument strength.

Roadmap

Section 2 introduces kernel-based inference and stability testing under weak identification. Section 3 extends the analysis to the K-statistic. Section 4 evaluates finite-sample properties. Section 5 presents the empirical application. Section 6 concludes. Proofs are collected in Appendix A-B, Appendix C provides implementation details for the SP-IV estimator, and Appendix D describes critical-value simulation for the proposed stability test.

2 Kernel-Based Inference and Stability Testing Under Weak Identification

This section develops a unified framework to conduct weak-identification robust inference for time-varying parameters in GMM models, including formal tests for parameter stability. Throughout the section, we employ standard Nadaraya–Watson kernel smoothing for local inference. For integrated (stability) statistics, we adopt a boundary correction that permits integration over the full interval $(0, 1)$ without trimming; details are given in Section 2.3.

We proceed in three steps. First, we establish a kernel-weighted local S-statistic for conducting inference at a specific point in time $\tau \in (0, 1)$ under the null hypothesis $\beta(\tau) = \beta_0$, without relying on consistency or asymptotic normality of a local GMM estimator. Second, we construct an integrated kernel statistic to test the global stability hypothesis $\beta(\tau) = \beta_0$ for all $\tau \in (0, 1)$, extending the moment-based stability framework of Magnusson and Mavroeidis (2014) to the kernel-smoothing context. Third, we decompose

this integrated statistic into two orthogonal components: a standard (full-sample) S-statistic capturing global fit, and a stability component that isolates time variation in the moment conditions. The latter serves as a test for stability that remains robust to weak identification. Extensions to the K-statistics are discussed separately in section 3.

The central idea is to evaluate the moment conditions $E[g_t(\beta_0)] = 0$ *locally* at each point in rescaled time $\tau \in (0,1)$ rather than only over the full sample. When the structural parameter β is constant and the instruments are valid, these moment conditions hold uniformly in τ. Three objects emerge from this approach:

(i) A *local S-statistic* $S_T^{\text{loc}}(\tau, \beta_0)$ that asks whether the model fits at a particular point in time τ. Inverting this statistic over β_0 yields pointwise, weak-identification robust confidence sets for the time-varying parameter $\beta(\tau)$.

(ii) An *integrated (generalized) S-statistic* $\mathcal{S}_T^{\text{gen}}(\beta_0)$ that pools the local fit across all $\tau \in (0,1)$ and asks whether β_0 is the common parameter throughout the sample.

(iii) A *stability component* $\mathcal{S}_T^{\text{stab}}(\beta_0)$ that filters out the average fit (already captured by the full-sample S-statistic of Stock and Wright (2000)) and isolates time variation in the moment conditions. This is the object used for formal stability testing, and we show that the distribution of it remains pivotal if evaluated at a (potentially weakly identified) full-sample estimator $\hat{\beta}$.

Crucially, all three objects have pivotal limiting distributions that do not depend on the strength of identification. The applied researcher can therefore use them regardless of whether instruments are strong, weak, or somewhere in between.

2.1 Notation and Assumptions

Let the dependent variable be a scalar $y_t \in \mathbb{R}$, the vector of regressors be $x_t \in \mathbb{R}^K$, and $z_t \in \mathbb{R}^L$ denote the vector of instruments, with $L \geq K$. The structural equation is

$$y_t = x_t'\beta_t + u_t, \qquad t = 1, \ldots, T, \tag{1}$$

where $\beta_t \in \mathbb{R}^K$ may vary with t and u_t is an unobserved disturbance. We interpret discrete-time variation through a rescaled-time function $\beta(\tau)$ via the link

$$\beta_t \equiv \beta(t/T), \qquad \tau = t/T \in (0,1]. \tag{2}$$

Under the null of parameter stability we assume that

$$H_0^{\text{stab}}: \quad \beta_t \equiv \beta_0 \in B \subset \mathbb{R}^K \qquad \text{for all } t, \tag{3}$$

for some compact parameter space B, and that the instruments satisfy the exogeneity

condition

$$E[z_t u_t] = 0, \qquad t = 1, \ldots, T. \tag{4}$$

Define the linear IV–GMM moment function

$$g_t(\beta) \equiv z_t\big(y_t - x_t'\beta\big) \in \mathbb{R}^L. \tag{5}$$

Under H_0^{stab} the identifying moment condition is

$$E\big[g_t(\beta_0)\big] = 0, \qquad t = 1, \ldots, T. \tag{6}$$

We will construct a kernel-based S test of stability from kernel averages of $g_t(\beta_0)$ over time.

Normalize calendar time by $\tau = t/T \in (0, 1]$. Let $K(\cdot)$ be a bounded, symmetric kernel function with compact support $[-1, 1]$ and let $h_T > 0$ denote a bandwidth. For each $\tau \in (0, 1)$ define the Nadaraya–Watson weights

$$w_{t,T}(\tau) = \frac{K\left(\frac{t/T-\tau}{h_T}\right)}{\sum_{s=1}^{T} K\left(\frac{s/T-\tau}{h_T}\right)}. \tag{7}$$

The corresponding local kernel average of the score at time τ is

$$\bar{g}_T(\tau; \beta_0) = \sum_{t=1}^{T} w_{t,T}(\tau)\, g_t(\beta_0). \tag{8}$$

The sum is effectively finite due to compact support of K.

We now state high-level assumptions on the score process $\{g_t(\beta_0)\}$, the kernel, and the long-run covariance that are sufficient for our test statistics to converge to pivotal distributions under H_0^{stab}. Our asymptotic framework is designed to deliver size control that is uniform in identification strength while allowing for rich forms of time variation under the alternative. The assumptions below play three distinct roles: (i) they formalize a weak-identification environment in which conventional local-GMM asymptotics need not apply; (ii) they provide a Gaussian-process approximation for kernel-smoothed moment averages; and (iii) they isolate a stability component whose limiting law is pivotal once an appropriate (time-invariant) normalization is imposed.

Assumption 1 (Local-to-zero drift / weak identification)**.** *Let $\{g_t(\beta_0)\}_{t=1}^T$ denote the score sequence under H_0^{stab}. For each β in a neighborhood of β_0 and every $\tau \in [0, 1]$, define the partial-sum process*

$$S_T(\tau, \beta) = \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor T\tau \rfloor} g_t(\beta).$$

We assume that under sequences of parameters $\{\beta_T\}$ *with* $\beta_T \to \beta_0$*, the mean of* $S_T(\tau, \beta_0)$ *satisfies*

$$E_{\beta_T}[S_T(\tau, \beta_0)] \longrightarrow \int_0^\tau m(r)\, dr \qquad \text{uniformly in } \tau \in [0,1], \tag{9}$$

for some measurable function $m : [0,1] \to \mathbb{R}^L$ *that depends on the sequence* $\{\beta_T\}$*. Under the null (*$\beta_T = \beta_0$ *for all* T*) we have* $m(\tau) \equiv 0$ *for* $\tau \in [0,1]$*.*

Assumption 1 is a high-level weak-IV condition that coincides with Assumption 1 of Magnusson and Mavroeidis (2014). In the linear IV example,

$$\mathrm{E}[g_t(\beta_0)] \;=\; \mathrm{E}[z_t x_t'](\beta_t - \beta_0),$$

so the assumption accommodates (i) weak or drifting first-stage strength (small $\mathrm{E}[z_t x_t']$), (ii) local deviations $\beta_t - \beta_0$ of order $T^{-1/2}$, or combinations thereof. This is precisely the regime under which identification-robust statistics (such as S/AR- and K-type tests) remain well behaved while Wald-type procedures need not. Under strong identification—when $\mathrm{E}[z_t x_t']$ has full column rank and $\beta_t \equiv \beta_0$—the assumption holds trivially with $m(\tau) \equiv 0$.

Evaluating (9) at $\tau = 1$ yields $T^{-1/2}\sum_{t=1}^T \mathrm{E}[g_t(\beta_0)] \to \int_0^1 m(r)\, dr$, which is exactly the Stock–Wright local-to-zero nesting (Stock and Wright, 2000). The additional content relative to Stock–Wright is that Assumption 1 also governs the *subsample* behavior of the partial-sum process: the function $m(\tau)$ captures how the mean of the moment conditions varies over time, not just in total. This generalization is essential because our kernel-smoothed statistics use localized (partial-sample) information. For the local S-statistic (Proposition 1), the assumption ensures that the kernel average $\bar{g}_T(\tau; \beta_0)$ is centered under $H_0^{\mathrm{loc}}(\tau)$, which delivers the pivotal χ^2_L limit. For the stability decomposition (Theorem 1), it ensures that the centering operation $\tilde{g}_T(\tau; \beta_0) = \bar{g}_T(\tau; \beta_0) - \bar{g}(\beta_0)$ eliminates τ-invariant components under H_0^{stab}, while preserving τ-varying departures under the alternative—which is the source of power for the stability test.[1]

Assumption 2 (Functional central limit theorem (FCLT) for the score)**.** *Let*

$$S_T(\tau) \equiv \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor T\tau \rfloor} g_t(\beta_0), \qquad \tau \in [0,1].$$

We assume that, under H_0^{stab} *and under the local-to-zero sequences in Assumption 1,*

$$S_T(\cdot) - E\,[S_T(\cdot)] \Rightarrow \Omega^{1/2} W(\cdot) \quad \text{in } D[0,1], \tag{10}$$

[1] Specifically, the assumption enters the proof of Proposition 1 through the condition $\int_0^\tau m(r)\, dr = 0$, which centers the local kernel average, and the proof of Theorem 1(i) through the identity $m(\tau) \equiv 0$ for all τ, which ensures that both $\bar{g}_T(\tau; \beta_0)$ and $\bar{g}(\beta_0)$ share the same (zero) mean under H_0^{stab}; see Appendix A for details.

where $W(\cdot)$ is an L-dimensional standard Brownian motion and Ω is a finite, symmetric, positive-definite long-run covariance matrix of $g_t(\beta_0)$.

Assumption 2 provides a functional central limit theorem for the score process, which is the key input for deriving the limiting distribution of kernel-weighted averages $\bar{g}_T(\tau;\beta_0)$. Intuitively, kernel smoothing is a local linear functional of the partial-sum process, so the weak convergence of $S_T(\cdot)$ implies that $\sqrt{Th_T}\,\bar{g}_T(\tau;\beta_0)$ is asymptotically Gaussian (after suitable normalization). This yields the chi-square limit for pointwise statistics and, when integrated over τ, leads to quadratic forms in Gaussian bridges used for global stability testing. The assumption is satisfied under standard strong-mixing or near-epoch-dependence conditions on (y_t, x_t, z_t), and it permits conditional heteroskedasticity in both the structural errors and the instruments, provided the long-run covariance Ω exists and is finite. It rules out processes for which the partial sums have non-Gaussian limits, such as heavy-tailed or long-memory innovations.

Assumption 3 (Kernel and bandwidth)**.**

1. *The kernel $K : \mathbb{R} \to \mathbb{R}$ is bounded, symmetric, with compact support $[-1, 1]$, and satisfies*

$$\int_{-1}^{1} K(u)\,du = 1, \qquad \int_{-1}^{1} u\,K(u)\,du = 0, \qquad R(K) \equiv \int_{-1}^{1} K(u)^2\,du < \infty.$$

2. *(*Shrinking bandwidth.*) The bandwidth sequence h_T satisfies*

$$h_T \to 0, \qquad Th_T \to \infty \quad \text{as } T \to \infty.$$

3. *(*Fixed bandwidth.*) Instead of (ii), $h_T \to h \in (0, \infty)$ as $T \to \infty$.*

Assumption 3(i) is standard for kernel estimation; symmetry and compact support simplify boundary handling and yield the normalization constant $R(K)$. The rate conditions in (ii) ensure the local window contains many observations ($Th_T \to \infty$) while still shrinking in rescaled time ($h_T \to 0$) for pointwise inference. Shrinking bandwidths are the natural choice for pointwise local inference (Proposition 1). For integrated (stability) statistics, we additionally consider the fixed-bandwidth regime (iii), which keeps the stability component non-degenerate and admits a clean Gaussian-process representation (Theorem 1(iii)(b)). In what follows, we refer to the two cases as Assumption 3(shrinking) and Assumption 3(fixed), respectively. Common kernels satisfying (i) include the Epanechnikov kernel $K(u) = \frac{3}{4}(1-u^2)\mathbf{1}(|u| \le 1)$, the biweight (quartic), and the uniform kernel.

Assumption 4 (Long-run covariance and weighting)**.** *Let $g_t(\beta_0) \in \mathbb{R}^L$ denote the score process under H_0^{stab}.*

1. ***(For pointwise inference.)*** *There exists a smooth, positive-definite matrix function $\Omega(\tau)$ on $(0,1)$, interpreted as the long-run covariance of the locally stationary approximation to $\{g_t(\beta_0)\}$ at rescaled time τ, together with a consistent estimator $\widehat{\Omega}_T(\tau) \xrightarrow{p} \Omega(\tau)$ that is invertible with probability approaching one.*

2. ***(For global stability testing.)*** *Under H_0^{stab} the long-run covariance is time-invariant: $\Omega(\tau) \equiv \Omega$ for all $\tau \in (0,1)$. We use a full-sample estimator $\widehat{\Omega}_T \xrightarrow{p} \Omega$, so that the weighting matrix in the stability statistic is $\widehat{\Omega}_T^{-1}$.*

Assumption 4 separates the requirements for pointwise inference from those for global stability testing. Local inference can be studentized using a time-varying $\Omega(\tau)$ without affecting the χ^2_L limit. Global stability testing, however, requires a time-invariant weighting matrix: if $\Omega(\tau)$ varied with τ, the integrated statistic would confound time variation in the *mean* of the moments—the signal of parameter instability—with time variation in their *second moments*, and a rejection could arise from heteroskedasticity in $g_t(\beta_0)$ alone. Part (b) rules this out. This type of restriction is standard in the stability testing literature: it is shared by all tests of structural change proposed in Andrews (1993), Elliott and Müller (2006) and the generalized S-tests of Magnusson and Mavroeidis (2014), and it appears difficult to dispense with while retaining pivotal limit distributions.[2]

When the score variance exhibits persistent shifts, the stability test may conflate mean instability with heteroskedasticity. In the SP-IV implementation, the VAR absorbs much of the predictable variation in volatility, and the Kronecker structure $\Omega = \Sigma_{u_H^\perp} \otimes Q$ concentrates remaining heteroskedasticity into the forecast-error block. The fact that both the S- and K-statistics reject—and that the K-statistic rejects more strongly—suggests that mean instability drives the rejection. Developing a heteroskedasticity-robust extension remains an important direction for future work.

We now define two related testing problems and their corresponding S-statistics. The first addresses local inference: testing whether the parameter equals a hypothesized value β_0 at a specific point in time τ. The second addresses global stability: testing whether the parameter remains constant at β_0 throughout the entire sample period.

2.2 Local hypothesis: point-in-time inference

For a given $\tau \in (0,1)$, consider the null hypothesis

$$H_0^{\text{loc}}(\tau): \quad \beta(\tau) = \beta_0. \tag{11}$$

To test $H_0^{\text{loc}}(\tau)$, we require a consistent estimator of the local long-run covariance $\Omega(\tau)$. When the score process $g_t(\beta_0)$ is serially uncorrelated—as in the SP-IV implementation

[2]The finite-sample consequences of moderate violations appear benign. Magnusson and Mavroeidis (2014) report that, in Monte Carlo experiments calibrated to the Great Moderation with variance breaks of up to six standard errors, size distortions for all generalized S-type tests remain within a few percentage points of the nominal level.

where VAR conditioning whitens the instrument—the contemporaneous covariance equals the long-run covariance, and a natural estimator is

$$\widehat{\Omega}_T(\tau) = \sum_{t=1}^{T} w_{t,T}(\tau)\big(g_t(\beta_0) - \bar{g}_T(\tau;\beta_0)\big)\big(g_t(\beta_0) - \bar{g}_T(\tau;\beta_0)\big)'. \tag{12}$$

[3]

The local kernel S-statistic at τ is then defined as

$$S_T^{\text{loc}}(\tau,\beta_0) = \frac{Th_T}{R(K)} \cdot \bar{g}_T(\tau;\beta_0)'\,\widehat{\Omega}_T^{-1}(\tau)\,\bar{g}_T(\tau;\beta_0), \tag{13}$$

where $R(K) = \int K(u)^2\,du$ is the roughness of the kernel.[4]

Proposition 1 (Asymptotic distribution of the local S-statistic)**.** *Suppose Assumptions 1–4(a) hold and h_T satisfies Assumption 3(shrinking). Then, under $H_0^{\text{loc}}(\tau) : \beta(\tau) = \beta_0$, for any $\tau \in (0,1)$,*

$$S_T^{\text{loc}}(\tau,\beta_0) \xrightarrow{d} \chi_L^2 \quad \text{as } T \to \infty, \tag{14}$$

where L is the number of moment conditions (dimension of $g_t(\beta_0)$). Moreover, the limit distribution does not depend on the strength of identification, i.e., it holds uniformly over the local-to-zero sequences in Assumption 1.

Practical interpretation and confidence sets. The local S-statistic can be evaluated for each τ on a grid over $(0,1)$. By inverting the test, one obtains pointwise weak-identification robust confidence sets for $\beta(\tau)$:

$$CS_T(\tau) = \big\{\beta_0 \in B : S_T^{\text{loc}}(\tau;\beta_0) \le \chi^2_{L,1-\alpha}\big\},$$

where $\chi^2_{L,1-\alpha}$ denotes the $(1-\alpha)$ quantile of the χ^2_L distribution. Computation is straightforward: evaluate $S_T^{\text{loc}}(\tau,\beta)$ on a grid over $\beta \in B$ for each τ. The confidence set $CS_T(\tau)$ may be empty (indicating local misspecification at τ), a single connected region, or a disconnected set (reflecting multimodality from weak identification). The χ^2_L critical value means the set is wider when L is large relative to K—this is the cost of not exploiting Jacobian information in the S-statistic, and motivates the K-statistic extension in Section 3.

To interpret $CS_T(\tau)$ as inference on the true time-varying object at rescaled time τ, we follow the locally stationary framework of Li et al. (2024). For each $\tau \in (0,1)$, let $\{(y_t(\tau), x_t(\tau), z_t(\tau))\}_{t\in\mathbb{Z}}$ denote a strictly stationary approximation valid in a shrinking

[3] When the score $g_t(\beta_0)$ exhibits serial dependence $\widehat{\Omega}_T(\tau)$ may be replaced by a kernel-weighted Newey–West estimator.

[4] The normalization $Th_T/R(K)$ equals $(\sum_{t=1}^T w_{t,T}(\tau)^2)^{-1}$ at interior points under boundary reflection, and is asymptotically equivalent at interior points without boundary reflection. In practice, one may compute the exact effective sample size $(\sum_t w_{t,T}(\tau)^2)^{-1}$ for each τ when boundary reflection is not used; the resulting statistic has the same asymptotic distribution.

neighborhood of τ, and define the local population moment

$$G(\tau, \beta) \equiv \mathrm{E}\big[z_t(\tau)\big(y_t(\tau) - x_t(\tau)'\beta\big)\big] .$$

Assumption 5 (Local identification and smooth pseudo-true path)**.** *For each $\tau \in (0,1)$ there exists a unique $\beta(\tau) \in B$ such that $G(\tau, \beta(\tau)) = 0$. Moreover, $\tau \mapsto \beta(\tau)$ is twice continuously differentiable on compact subsets of $(0,1)$ with uniformly bounded derivatives.*

Under Assumption 5, kernel localization targets the point-valued pseudo-true parameter $\beta(\tau)$ (rather than an average of regimes), and the drift within a window of width h_T is of order h_T^2 after centering. The local stationarity framework of Li et al. (2024) provides an appropriate set of regularity conditions under which this interpretation is formalized. In their framework, the data-generating process $\{X_t\}$ is locally stationary in the sense of Dahlhaus (1997): there exists, for each rescaled time point $u \in [0,1]$, a strictly stationary process $\{X_t(u)\}$ that approximates $\{X_t\}$ in a neighborhood of u with an asymptotically negligible error. Under their conditions, the local GMM estimator is consistent for $\beta(\tau)$ and asymptotically normal with rate $\sqrt{Th_T}$, with asymptotic bias of order h_T^2 arising from the curvature of $\beta(\cdot)$.

An alternative approach is that of Giraitis et al. (2021), who allow the time-varying parameters β_t to follow either smooth deterministic paths or persistent stochastic processes (such as random walks), with smoothness controlled by conditions on the increments $|\beta_t - \beta_{t-1}|$ rather than by assuming $\beta_t = \beta(t/T)$ for a smooth function $\beta(\cdot)$. Their framework yields consistency at a slower rate—the estimation error $\hat{\beta}(\tau) - \beta(\tau)$ is $o_p(1)$ but converges more slowly than $O_p((Th_T)^{-1/2})$—compared to the $O_p((Th_T)^{-1/2})$ rate obtainable under the local stationarity framework, but accommodates a broader class of parameter paths, including stochastically time-varying coefficients.

For our purposes, the distinction between these two frameworks affects primarily the interpretation of the confidence set $CS_T(\tau)$ and the required bandwidth conditions. Under the Li et al. (2024) conditions, $CS_T(\tau)$ targets the smooth pseudo-true function $\beta(\tau)$ and the undersmoothing condition $\sqrt{Th_T}\, h_T^2 \to 0$ ensures the kernel bias is asymptotically negligible. Under the Giraitis et al. (2021) framework, $CS_T(\tau)$ targets the realized parameter path β_t at time $t \approx \tau T$, and the bandwidth conditions ensure that the local average is close to β_t with high probability. The S-statistic and its pivotal χ^2_L limit (Proposition 1) remain valid under both frameworks, since they depend only on the high-level FCLT (Assumption 2) and not on the specific structural model for parameter variation.

Proposition 2 (Pointwise coverage for $\beta(\tau)$)**.** *Suppose Assumptions 1–4(a), the local stationarity conditions underlying Assumption 2, and Assumption 5 hold. If*

$$h_T \to 0, \qquad Th_T \to \infty, \qquad \sqrt{Th_T}\, h_T^2 \to 0,$$

then for any fixed interior $\tau \in (0,1)$,

$$\liminf_{T\to\infty} \Pr\big(\beta(\tau) \in CS_T(\tau)\big) \geq 1-\alpha.$$

The undersmoothing condition $\sqrt{Th_T}\, h_T^2 \to 0$ requires $h_T = o(T^{-1/5})$, which is faster than the MSE-optimal rate $h_T \asymp T^{-1/5}$ for kernel regression. In practice, a plug-in rule such as $h_T = 0.8\, T^{-1/5}$ technically violates this condition, though the resulting bias is typically negligible in the samples considered in our Monte Carlo experiments (Section 4).

When local identification fails—i.e., when uniqueness of $\beta(\tau)$ is not imposed in Assumption 5—one may define the local identified set $\mathcal{I}(\tau) = \{\beta \in B : G(\tau,\beta) = 0\}$. Standard test inversion then yields pointwise coverage for any fixed $\beta_0 \in \mathcal{I}(\tau)$: $\liminf_T \Pr(\beta_0 \in CS_T(\tau)) \geq 1-\alpha$. Covering the entire identified set $\mathcal{I}(\tau) \subseteq CS_T(\tau)$ with probability at least $1-\alpha$ would require a uniform argument over $\beta \in \mathcal{I}(\tau)$, which we do not pursue here. The framework nevertheless degrades gracefully from point to partial identification.

2.3 Global hypothesis: testing parameter stability

We now turn to testing whether the parameter remains constant throughout the entire sample period. The global stability null hypothesis is

$$H_0^{\text{stab}}: \quad \beta(\tau) = \beta_0 \quad \text{for all } \tau \in (0,1). \tag{15}$$

To obtain a uniform normalization for the integrated statistic, we follow Chen and Hong (2012) and adopt the boundary reflection method of Hall and Wehrly (1991): the sample is augmented with pseudodata constructed by reflection at the endpoints, and $\bar{g}_T(\tau;\beta_0)$ is computed on the extended sample. This makes the NW denominator (approximately) constant across $\tau \in (0,1)$, so that the variance normalization is $R(K)$ uniformly and no boundary-specific constants appear. Boundary reflection is not required for the local S-statistic (Proposition 1), which is evaluated at interior points. In practice, one may alternatively omit reflection and simulate the Gaussian-process limit with its boundary-adjusted covariance. We describe the simulation algorithm in Appendix D.

The key mechanism underlying the stability test is centering. The intuition is as follows: if we evaluate the moment conditions at an estimator $\hat{\beta}$ rather than the true β_0, the estimation error $\hat{\beta} - \beta_0$ shifts all local averages $\bar{g}_T(\tau;\hat{\beta})$ by approximately the same τ-invariant amount, because the Jacobian $\mathrm{E}[z_t x_t']$ does not vary much within the kernel window. Differencing out the full-sample average removes this common shift, so that the centered object $\tilde{g}_T(\tau;\hat{\beta}) = \bar{g}_T(\tau;\hat{\beta}) - \bar{g}(\hat{\beta})$ depends on $\hat{\beta}$ only through higher-order terms. This is conceptually analogous to the orthogonalization/projection argument used in weak-instrument structural change tests (Caner, 2011).

Following the generalized S-statistic framework of Magnusson and Mavroeidis (2014),

define the generalized kernel S-statistic

$$\mathcal{S}_T^{\text{gen}}(\beta_0) = \int_0^1 \frac{Th_T}{R(K)} \cdot \bar{g}_T(\tau;\beta_0)' \,\widehat{\Omega}_T^{-1}\, \bar{g}_T(\tau;\beta_0)\, d\tau. \tag{16}$$

To isolate instability, define the centered local scores

$$\tilde{g}_T(\tau;\beta_0) = \bar{g}_T(\tau;\beta_0) - \bar{g}(\beta_0), \tag{17}$$

where $\bar{g}(\beta_0) = T^{-1}\sum_{t=1}^{T} g_t(\beta_0)$ is the full-sample average. The corresponding stability component is

$$\mathcal{S}_T^{\text{stab}}(\beta_0) = \int_0^1 \frac{Th_T}{R(K)} \cdot \tilde{g}_T(\tau;\beta_0)' \,\widehat{\Omega}_T^{-1}\, \tilde{g}_T(\tau;\beta_0)\, d\tau, \tag{18}$$

which in practice can be approximated by

$$\mathcal{S}_T^{\text{stab}}(\beta_0) \approx \frac{h_T}{R(K)} \sum_{t=1}^{T} \tilde{g}_T(t/T;\beta_0)' \,\widehat{\Omega}_T^{-1}\, \tilde{g}_T(t/T;\beta_0).$$

Define the full-sample (Stock–Wright) S-statistic

$$S_T^{\text{full}}(\beta_0) \equiv T\,\bar{g}(\beta_0)' \,\widehat{\Omega}_T^{-1}\, \bar{g}(\beta_0). \tag{19}$$

The following theorem summarizes key results on the generalized kernel S-statistic.

Theorem 1 (Asymptotic distribution of the generalized kernel S-statistic)**.** *Suppose Assumptions 1–4(b) hold. Under the null hypothesis* $H_0^{\text{stab}} : \beta(\tau) = \beta_0$ *for all* $\tau \in (0,1)$*:*

(i) ***Orthogonal decomposition:*** *The integrated statistic admits the decomposition*

$$\mathcal{S}_T^{\text{gen}}(\beta_0) = \underbrace{\frac{h_T}{R(K)}\, S_T^{\text{full}}(\beta_0)}_{\text{rescaled full-sample S-statistic}} + \underbrace{\mathcal{S}_T^{\text{stab}}(\beta_0)}_{\text{stability component}} + o_p(1). \tag{20}$$

The first term is a rescaled version of the standard full-sample S-statistic of Stock and Wright (2000); the stability component isolates time variation.

(ii) ***Asymptotic independence:*** *The full-sample component and the stability component are asymptotically independent.*

(iii) ***Limiting behavior:***

(a) Full-sample component: *Under* H_0^{stab} *and Assumption 3(shrinking),* $S_T^{\text{full}}(\beta_0) = O_p(1)$ *(and in particular* $S_T^{\text{full}}(\beta_0) \Rightarrow \chi_L^2$ *under standard regularity). Since* $h_T \to 0$ *under Assumption 3(shrinking), the rescaled full-sample term* $\frac{h_T}{R(K)} S_T^{\text{full}}(\beta_0) = o_p(1)$.

(b) Stability component (fixed bandwidth): *If Assumption 3(fixed) holds, then*

$$\mathcal{S}_T^{\text{stab}}(\beta_0) \Rightarrow \int_0^1 \|\tilde{G}(\tau)\|^2 \, d\tau, \tag{21}$$

where $\tilde{G}(\tau) = G(\tau) - \int_0^1 G(s)\, ds$ *is an* L*-dimensional Gaussian bridge process with covariance (componentwise)*

$$Cov\big(G(\tau_1), G(\tau_2)\big) = \rho\left(\frac{\tau_2 - \tau_1}{h}\right), \tag{22}$$

with $\rho(u) = \int_{-1}^{1} K(v)K(v+u)\, dv$ *and* $h = \lim h_T$*. Critical values for the Gaussian-process representation can be obtained quickly via simulation; see Appendix D.*

(c) Stability component (shrinking bandwidth): *If Assumption 3(shrinking) holds, then*

$$\frac{\mathcal{S}_T^{\text{stab}}(\beta_0) - L}{\sqrt{h_T \cdot 4L\, C_B\, R(K)^{-2}}} \Rightarrow N(0,1), \tag{23}$$

and

$$C_B = \int_0^2 \left[\int_{-1}^{1} K(u)K(u+v)\, du\right]^2 dv.$$

The proof is outlined in Appendix A. The asymptotic normality result for the stability component closely follows Chen and Hong (2012).

The decomposition (20) parallels that of the generalized S statistic in Magnusson and Mavroeidis (2014). Their qLL-S statistic admits the representation qLL-$S_T^{\tilde{c},\bar{c}}(\beta_0) =$ qLL-$S_T^{\tilde{c}}(\beta_0) + \frac{\bar{c}}{1+\bar{c}} S_T(\beta_0)$, where qLL-$S_T^{\tilde{c}}(\beta_0)$ is the stability component and $S_T(\beta_0)$ is the full-sample S-statistic. Our kernel-based decomposition achieves an analogous separation, with the scaling $h_T/R(K)$ playing a role analogous to $\bar{c}/(1+\bar{c})$.

The stability statistic $\mathcal{S}_T^{\text{stab}}$ is closely related to the generalized Hausman test of Chen and Hong (2012). Specializing their framework to the intercept-only case ($X_t = 1$), their statistic takes the form $\hat{Q} = T^{-1} \sum_{t=1}^{T} \big(\bar{g}_T(t/T; \beta_0) - \bar{g}(\beta_0)\big)^2$, which coincides with our $\mathcal{S}_T^{\text{stab}}$ up to scaling and multivariate weighting. Chen and Hong (2012) establish a standardized normal limit in the univariate case; Theorem 1(iii)(c) extends this to the multivariate setting ($L \geq 1$) with the weighting matrix $\widehat{\Omega}_T^{-1}$.

In practice, we recommend the following two-step procedure. First, evaluate $\mathcal{S}_T^{\text{stab}}(\hat{\beta})$ using GP critical values from Theorem 1(iii)(b) (or normal critical values under Theorem 1(iii)(c)). If the stability null is not rejected, full-sample S- or K-based inference proceeds as in Stock and Wright (2000) or Kleibergen (2005). If rejected, the researcher proceeds to local inference using the pointwise statistics from Proposition 1, which provide time-varying confidence sets for $\beta(\tau)$ at each τ.

Choice of asymptotic regime. The bandwidth may be data-dependent or chosen by rule of thumb. Our Monte Carlo evidence shows GP-based critical values from the fixed-bandwidth representation provide accurate finite-sample calibration even with a shrinking rule such as $h_T = 0.8\,T^{-1/5}$, whereas normal approximation underrejects at typical sample sizes. We recommend using GP-based critical values as the default.

It is important to recognize that the stability test $\mathcal{S}_T^{\text{stab}}$ detects time variation in the moment conditions $\mathrm{E}[g_t(\beta_0)] = \mathrm{E}[z_t(y_t - x_t'\beta_0)]$, not directly in the structural parameter β_t. Under the maintained assumption that the instruments are valid throughout the sample—i.e., $\mathrm{E}[z_t u_t] = 0$ for all t—moment instability and parameter instability coincide: $\mathrm{E}[g_t(\beta_0)] \neq 0$ for some t if and only if $\beta_t \neq \beta_0$ for those periods. However, if instrument validity is itself time-varying—for instance, if $\mathrm{E}[z_t u_t] \neq 0$ for t in some subperiod due to policy regime changes, measurement shifts, or misspecification of the exclusion restriction—then the test may reject even when parameters are stable, because the moment conditions are violated locally.

A rejection of H_0^{stab} therefore has two possible interpretations: (i) parameter instability under maintained instrument validity, or (ii) local instrument invalidity with possibly stable parameters. Disentangling these two sources of rejection requires additional economic reasoning or auxiliary tests. For instance, one may compare results across alternative instrument sets: if rejection occurs with one set of instruments but not another, this suggests the issue lies with instrument validity rather than structural instability. Alternatively, one may apply the local S-statistic (Proposition 1) to diagnose where in the sample the moment conditions fail, which may help distinguish between economic structural change and specification problems.

For further discussion of local instrument validity in a time-varying GMM context, we refer the reader to Giraitis et al. (2021), who develop time-varying Hausman tests that compare OLS and IV estimates at each point in time and can thereby detect periods of endogeneity, and to Li et al. (2024), who formalize local moment validity under the locally stationary framework and propose tests for both parameter constancy and parametric functional forms.

The following proposition establishes that the stability test remains valid even when the value of β used to evaluate the moments is replaced by a weakly identified estimator.

Proposition 3 (Weak-identification robustness of the stability statistic)**.** *Under Assumptions 1–4(b), let $\hat{\beta}$ be any estimator satisfying $\hat{\beta} - \beta_0 = O_p(1)$ under weak identification. In addition, assume the (appendix-only) integrated Jacobian stability condition stated in Appendix A (Assumption A.1) holds. Then:*

*(i) (*Fixed bandwidth.*) The feasible stability statistic $\mathcal{S}_T^{\text{stab}}(\hat{\beta})$ has the same asymptotic distribution as $\mathcal{S}_T^{\text{stab}}(\beta_0)$ in Theorem 1(iii)(b), i.e.,*

$$\mathcal{S}_T^{\text{stab}}(\hat{\beta}) \Rightarrow \int_0^1 \|\tilde{G}(\tau)\|^2 \, d\tau \quad \textit{under Assumption 3(fixed).}$$

(ii) (Shrinking bandwidth.) *Under the shrinking-bandwidth regime (Assumption 3(shrinking)) and the additional quadratic-form regularity conditions required for Theorem 1(iii)(c),*

$$\frac{\mathcal{S}_T^{\mathrm{stab}}(\hat{\beta}) - L}{\sqrt{h_T \cdot 4L\, C_B\, R(K)^{-2}}} \Rightarrow N(0,1).$$

In particular, the limiting distribution does not depend on the strength of identification or on the specific estimator $\hat{\beta}$ used.

3 Generalization to the K-statistic

This section extends the framework to local inference and stability testing based on the K-statistic of Kleibergen (2005). The S-statistic developed in Section 2 uses all L moment directions to test $\beta(\tau) = \beta_0$, yielding a χ^2_L null limit. The K-statistic projects the moment conditions onto the K directions identified by the Jacobian of the moment function, trading breadth for sharper power along identified directions. The resulting statistic is χ^2_K under the null, which can offer substantial power gains when L is much larger than K. As with the S-statistic, weak-identification robustness is inherited from the orthogonalization construction of Kleibergen (2005).

We maintain the linear IV–GMM setup of Section 2: the score is $g_t(\beta) = z_t(y_t - x_t'\beta) \in \mathbb{R}^L$ as defined in (5), and we additionally use the Jacobian

$$H_t(\beta) = \frac{\partial g_t(\beta)}{\partial \beta'} = -z_t x_t' \in \mathbb{R}^{L\times K}. \tag{24}$$

The K-statistic is built from an orthogonalized score that removes the component of Jacobian fluctuations predictable from fluctuations in the moments. This requires a joint weak convergence result for $(g_t(\beta_0), \mathrm{vec}(H_t(\beta_0)))$ and consistent estimation of the associated long-run covariance blocks—the kernel analogue of the regularity conditions in Kleibergen (2005).

Define the stacked innovation $\phi_t(\beta_0) = (g_t(\beta_0)', \mathrm{vec}(H_t(\beta_0) - J(\beta_0))')' \in \mathbb{R}^{L+LK}$, where $J(\beta_0) = \mathrm{E}[H_t(\beta_0)]$ is the expected Jacobian.

Assumption 6 (Joint FCLT for score and Jacobian)**.** *Under the local-to-zero sequences in Assumption 1, the partial-sum process of $\phi_t(\beta_0)$ satisfies*

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor sT \rfloor} \phi_t(\beta_0) \;\Rightarrow\; V^{1/2} W(s) \quad \text{in } D[0,1], \tag{25}$$

where $W(s)$ is an $(L+LK)$-dimensional standard Brownian motion and V is the long-run

covariance of $\phi_t(\beta_0)$ with partition

$$V = \begin{pmatrix} \Omega & V_{gh} \\ V_{hg} & V_{hh} \end{pmatrix}, \tag{26}$$

with Ω the $L \times L$ long-run covariance of $g_t(\beta_0)$, $V_{gh} = V_{hg}'$ the $L \times LK$ cross-covariance, and V_{hh} the $LK \times LK$ long-run covariance of the demeaned Jacobian. We assume Ω is positive definite and V is finite.

Assumption 7 (Covariance estimation for K-statistics)**.**

1. ***(For pointwise inference.)*** *For each $\tau \in (0,1)$, the stacked process $\{\phi_t(\beta_0)\}$ admits a locally stationary approximation with long-run covariance $V(\tau)$, partitioned conformably with (26). There exists a kernel-based estimator $\widehat{V}_T(\tau) \xrightarrow{p} V(\tau)$, with $V(\tau)$ positive definite for all $\tau \in (0,1)$.*

2. ***(For stability testing.)*** *Under H_0^{stab}, the long-run covariance is time-invariant: $V(\tau) \equiv V$ for all $\tau \in (0,1)$. A full-sample estimator satisfies $\widehat{V}_T \xrightarrow{p} V$.*

3.1 Local K-inference at a point in time

For a given $\tau \in (0,1)$, consider the null hypothesis $H_0^{\text{loc}}(\tau) : \beta(\tau) = \beta_0$. Using the NW weights $w_{t,T}(\tau)$ from (7), define the local kernel averages of the score and Jacobian,

$$\bar{g}_T(\tau;\beta_0) = \sum_{t=1}^{T} w_{t,T}(\tau)\, g_t(\beta_0), \qquad \bar{H}_T(\tau;\beta_0) = \sum_{t=1}^{T} w_{t,T}(\tau)\, H_t(\beta_0), \tag{27}$$

and the orthogonalized local Jacobian

$$\operatorname{vec}\big(\bar{D}_T(\tau;\beta_0)\big) = \operatorname{vec}\big(\bar{H}_T(\tau;\beta_0)\big) - \widehat{V}_{hg,T}(\tau)\,\widehat{\Omega}_T(\tau)^{-1}\,\bar{g}_T(\tau;\beta_0), \tag{28}$$

where $\widehat{\Omega}_T(\tau)$ and $\widehat{V}_{hg,T}(\tau)$ are the local covariance estimators from Assumption 7(a). By construction, $\bar{D}_T(\tau;\beta_0)$ is asymptotically uncorrelated with $\bar{g}_T(\tau;\beta_0)$ under the null—this is the kernel analogue of the orthogonalization in Kleibergen (2005).

The local kernel K-statistic is the projected quadratic form

$$K_T^{\text{loc}}(\tau;\beta_0) = \frac{Th_T}{R(K)}\,\bar{g}_T(\tau;\beta_0)'\,\widehat{\Omega}_T(\tau)^{-1/2}\,P_{\bar{D}_T(\tau)}\,\widehat{\Omega}_T(\tau)^{-1/2}\,\bar{g}_T(\tau;\beta_0), \tag{29}$$

where $P_{\bar{D}_T(\tau)}$ is the orthogonal projection onto $\operatorname{span}(\widehat{\Omega}_T(\tau)^{-1/2}\bar{D}_T(\tau;\beta_0))$. This is equivalent to Kleibergen's formulation $\hat{\kappa}_T'\hat{A}_T^{-1}\hat{\kappa}_T$ with $\hat{\kappa}_T = \sqrt{Th_T}\,\bar{D}_T(\tau)'\widehat{\Omega}_T(\tau)^{-1}\bar{g}_T(\tau)$ and $\hat{A}_T = \bar{D}_T(\tau)'\widehat{\Omega}_T(\tau)^{-1}\bar{D}_T(\tau)$.

Proposition 4 (Local chi-square limit of the kernel K-statistic)**.** *Suppose Assumptions 1–3(shrinking) and Assumptions 6–7(a) hold. Fix $\tau \in (0,1)$ and consider $H_0^{\text{loc}}(\tau) : \beta(\tau) =$*

β_0. *Then*

$$K_T^{\mathrm{loc}}(\tau;\beta_0) \xrightarrow{d} \chi^2_K. \tag{30}$$

The limit holds uniformly over the local-to-zero sequences in Assumption 1.

Inverting the test yields pointwise, weak-identification-robust confidence sets for $\beta(\tau)$: $CS_T^K(\tau) = \{\beta_0 \in B : K_T^{\mathrm{loc}}(\tau;\beta_0) \leq \chi^2_{K,1-\alpha}\}$. Because $K < L$ in the overidentified case, these sets are typically tighter than those based on the S-statistic. The same coverage guarantee extends to the true time-varying parameter $\beta(\tau)$ under the conditions of Proposition 2: if the bandwidth satisfies $\sqrt{Th_T}\,h_T^2 \to 0$ and local identification holds (Assumption 5), then $\liminf_T \Pr(\beta(\tau) \in CS_T^K(\tau)) \geq 1-\alpha$ by the same undersmoothing argument, with L replaced by K throughout.

3.2 Stability testing based on the kernel K-statistic

To construct a stability K-statistic, we need a projection that reduces the L-dimensional centered moment process to its K most informative directions. The natural choice is to project onto the column space of the full-sample orthogonalized Jacobian. Let

$$\mathrm{vec}(\bar{D}) = \mathrm{vec}\Big(\frac{1}{T}\sum_{t=1}^{T} H_t(\beta_0)\Big) - \widehat{V}_{hg}\,\widehat{\Omega}_T^{-1}\,\bar{g}(\beta_0), \tag{31}$$

where $\widehat{V}_{hg}$ and $\widehat{\Omega}_T$ are full-sample estimators (Assumption 7(b)), and define the rank-K projection

$$P_{\bar{D}} = \widehat{\Omega}_T^{-1/2}\bar{D}\big(\bar{D}'\widehat{\Omega}_T^{-1}\bar{D}\big)^{-1}\bar{D}'\widehat{\Omega}_T^{-1/2}. \tag{32}$$

Because $P_{\bar{D}}$ is computed from the full sample, it is a fixed (data-dependent but τ-invariant) matrix. No assumption on the time-invariance of the Jacobian subspace is required: the construction is a choice of test statistic, not a restriction on the data-generating process.[5]

The integrated kernel K-statistic is

$$\mathcal{K}_T^{\mathrm{int}}(\beta_0) = \int_0^1 \frac{Th_T}{R(K)}\,\bar{g}_T(\tau;\beta_0)'\,\widehat{\Omega}_T^{-1/2}\,P_{\bar{D}}\,\widehat{\Omega}_T^{-1/2}\,\bar{g}_T(\tau;\beta_0)\,d\tau. \tag{33}$$

Because $P_{\bar{D}}$ is τ-invariant, the same algebraic decomposition as in Theorem 1(i) applies. Writing $\tilde{g}_T(\tau) = \bar{g}_T(\tau;\beta_0) - \bar{g}(\beta_0)$, we obtain

$$\mathcal{K}_T^{\mathrm{int}}(\beta_0) = \frac{h_T}{R(K)}\,K_T^{\mathrm{full}}(\beta_0) + \mathcal{K}_T^{\mathrm{stab}}(\beta_0) + o_p(1), \tag{34}$$

[5]This contrasts with the local K-statistic in Section 3.1, which uses $\bar{D}_T(\tau;\beta_0)$ and therefore adapts the projection to each τ. For stability testing, a τ-invariant projection is needed to ensure that the integrated statistic detects time variation in the moments rather than time variation in the projection itself.

where the full-sample K-statistic and stability component are

$$K_T^{\text{full}}(\beta_0) = T\,\bar{g}(\beta_0)'\,\widehat{\Omega}_T^{-1/2}\,P_{\bar{D}}\,\widehat{\Omega}_T^{-1/2}\,\bar{g}(\beta_0), \tag{35}$$

$$\mathcal{K}_T^{\text{stab}}(\beta_0) = \int_0^1 \frac{Th_T}{R(K)}\,\tilde{g}_T(\tau)'\,\widehat{\Omega}_T^{-1/2}\,P_{\bar{D}}\,\widehat{\Omega}_T^{-1/2}\,\tilde{g}_T(\tau)\,d\tau. \tag{36}$$

The cross term vanishes by the same integrated-weights argument as in Lemma 1.

The null distribution of $\mathcal{K}_T^{\text{stab}}$ follows directly from Theorem 1. Under H_0^{stab}, the centered process $\sqrt{Th_T/R(K)}\,\widehat{\Omega}_T^{-1/2}\tilde{g}_T(\tau)$ converges to an L-dimensional Gaussian bridge $\tilde{G}(\tau)$. Projecting onto the fixed rank-K subspace $P_{\bar{D}}$ yields a K-dimensional Gaussian bridge, and the integrated quadratic form inherits the same distributional structure with L replaced by K.

Corollary 1 (Null distribution of the stability K-statistic)**.** *Suppose Assumptions 1–4(b) and Assumptions 6–7(b) hold.*

(i) Under Assumption 3(fixed) with $h_T \to h \in (0,\infty)$,

$$\mathcal{K}_T^{\text{stab}}(\beta_0) \;\Rightarrow\; \int_0^1 \|\tilde{\xi}(\tau)\|^2\,d\tau, \tag{37}$$

where $\tilde{\xi}(\tau) = \xi(\tau) - \int_0^1 \xi(s)\,ds$ is a K-dimensional centered Gaussian process, with $\xi(\tau)$ having covariance $\mathrm{Cov}(\xi(\tau_1),\xi(\tau_2)) = \rho\big((\tau_2-\tau_1)/h\big)I_K$ and $\tilde{\xi}$ inheriting the corresponding bridged covariance $\mathrm{Cov}(\tilde{\xi}(\tau_1),\tilde{\xi}(\tau_2)) = \big[\rho\big((\tau_2-\tau_1)/h\big) - \bar{\rho}(\tau_1) - \bar{\rho}(\tau_2) + \bar{\bar{\rho}}\big]I_K$, where $\bar{\rho}(\tau) = \int_0^1 \rho\big((\tau-s)/h\big)\,ds$ and $\bar{\bar{\rho}} = \int_0^1\int_0^1 \rho\big((s_1-s_2)/h\big)\,ds_1\,ds_2$.

(ii) Under Assumption 3(shrinking) with $h_T \to 0$ and $Th_T \to \infty$,

$$\frac{\mathcal{K}_T^{\text{stab}}(\beta_0) - K}{\sqrt{4K\,h_T\,C_B\,R(K)^{-2}}} \;\xrightarrow{d}\; N(0,1). \tag{38}$$

In both cases, the limit does not depend on identification strength. Critical values are obtained exactly as for the S-statistic: by simulation of the Gaussian process in case (i) (Appendix D, with L replaced by K), or from the standard normal table in case (ii).

The stability K-statistic concentrates the test on the K directions that are, on average, informative about β_0. When the identification subspace is approximately stable over time, this yields sharper power than the S-statistic because the null distribution involves a K-dimensional (rather than L-dimensional) Gaussian process, producing smaller critical values under both fixed-bandwidth GP calibration and the shrinking-bandwidth normal approximation. If, however, the Jacobian subspace rotates substantially across the sample—for instance, because different instruments are relevant in different subperiods—the full-sample projection averages over these directions, potentially diluting power against localized alternatives. In such settings, the unprojected S-statistic

$\mathcal{S}_T^{\text{stab}}$, which uses all L moment directions, may be preferable. Comparing the two statistics can itself be informative: if $\mathcal{S}_T^{\text{stab}}$ rejects but $\mathcal{K}_T^{\text{stab}}$ does not, this suggests instability may be concentrated in directions orthogonal to the average Jacobian.

Robustness to time-varying Jacobians. The stability K-statistic's null distribution does not depend on the rank-K subspace defined by P_{J_0}. Under the null, projecting the whitened scores onto any K orthonormal directions yields components whose distribution depends only on K and bandwidth h. Even if the true local Jacobian varies over time, the full-sample projection converges to a fixed rank-K projection whose identity is immaterial for the null distribution. Time variation in the Jacobian therefore does not threaten validity of $\mathcal{K}_T^{\text{stab}}$—only its power. This stands in contrast to the long-run covariance Ω, whose time-invariance is essential for pivotality of both the S- and K-statistics.

Corollary 2 (Plug-in evaluation under weak identification)**.** *Suppose, in addition, that $\hat{\beta}$ satisfies $\hat{\beta} - \beta_0 = O_p(1)$ under H_0^{stab}. Since $g_t(\beta)$ is affine in β in the linear IV model, the centering operation $\tilde{g}_T(\tau; \hat{\beta}) = \bar{g}_T(\tau; \hat{\beta}) - \bar{g}(\hat{\beta})$ removes the τ-invariant estimation error by the same argument as in Proposition 3. Hence the limits in Corollary 1 continue to hold with β_0 replaced by $\hat{\beta}$.*

4 Finite sample properties

This section studies the finite-sample behavior of the kernel-based structural break tests proposed in Section 2 and 3. The analysis is based on Monte Carlo simulations designed to closely mirror the simulation design in Barnichon and Mesters (2020) who study constant-parameter weak-instrument-robust inference procedures in forward-looking New Keynesian Phillips Curve (NKPC) settings.

4.1 Size

First, we focus on size control under the null of parameter stability, both for local-in-time inference and for the integrated stability tests. Following Barnichon and Mesters (2020), we simulate data from the forward-looking NKPC given by

$$y_t = \gamma_b y_{t-1} + \gamma_f E_t(y_{t+1}) + \lambda x_t + e_t, \tag{39}$$

where the forcing variable x_t follows an AR(2) process,

$$x_t = \rho_1 x_{t-1} + \rho_2 x_{t-2} + \varepsilon_t^i + \nu e_t. \tag{40}$$

where the structural parameters are fixed at $(\gamma_b, \gamma_f, \lambda) = (0.6, 0.3, 0.4)$, while $(\rho_1, \rho_2) = (1.2, -0.4)$ and $\nu = -1$. The economy is driven by two shocks, e_t and ε_t^i, where $z_t = \varepsilon_t^i$ is observed by the econometrician and serves as the external instrument. To match

empirical estimates for the importance of monetary policy shocks in driving output and inflation, Barnichon and Mesters (2020) set $\varepsilon_t \sim \mathcal{N}(0, \sigma_i^2)$ with standard deviation $\sigma_i = \{0.1, 0.25, 0.5, 1\}$ and $e_t = \rho e_{t-1} + \sqrt{1-\rho^2}\xi_t$ with $\xi_t \sim \mathcal{N}(0,1)$. This covers cases where the instrument explains between roughly 1% and 50% of the variance of y_t. We also consider the case where the instrument is noise ($z_t \sim \mathcal{N}(0,1)$) and therefore completely uninformative. The parameter $\rho = 0.5$ is set to allow for serial correlation in the disturbance e_t.

We consider sample sizes $T \in 250, 500$ and generate 5,000 Monte Carlo replications for each value of σ_i as well as the completely unidentified case where $z_t \sim \mathcal{N}(0,1)$. All simulations are conducted under the null hypothesis of parameter stability, $\beta(t) = \beta_0$ for all t, where $\beta = (\gamma_b, \gamma_f, \lambda)'$.

We report rejection frequencies at the nominal 5% level for two classes of tests:

1. *Local inference at a fixed time point* τ. For a given $\tau \in (0,1)$, we evaluate the kernel-based local S-statistic $S_T^{loc}(\tau, \beta_0)$ and K-statistic $K_T^{loc}(\tau, \beta_0)$ under the null $\beta(\tau) = \beta_0$, treating β_0 as known. We report results at $\tau_0 = \{0.05, 0.15, 0.5\}$ to study the behavior also near the boundary. We compare it to a full-sample Wald test obtained under strong-instrument asymptotics, as well as the full-sample S-Statistic and K-statistic.

2. *Global stability inference.* We consider the integrated kernel stability test $\mathcal{S}_T^{\text{stab}}(\beta_0)$, $\mathcal{K}_T^{\text{stab}}(\beta_0)$, $\mathcal{S}_T^{\text{stab}}(\hat{\beta})$ and $\mathcal{K}_T^{\text{stab}}(\hat{\beta})$ where $\hat{\beta}$ is the plug-in estimator obtained from full-sample GMM estimation. For both of the tests, we consider size based on the Gaussian process representation as well as asymptotic normality.

To estimate the parameters, we use the SP-IV GMM estimator by Lewis and Mertens (2022) based on $H = 20$ forecast horizon of which we take every second in the SP-IV regression.[6] Given a scalar instruments, this yields 10 moments for three parameters, which resembles the case of over-identification encountered when using structural shocks to identify the NKPC. The results remain little changed if we choose every second forecast error up to 10 steps ahead, yielding a total of 5 moments. Throughout the section, we rely on the Epanechnikov Kernel and a simple plug-in rule of $h_T = 0.8 \cdot T^{-1/5}$.

[6] We use a VAR(4) to create forecast errors and a (block diagonal) weighting matrix $\Sigma_{u_H^\perp}^{-1} \otimes E(z_t z_t')^{-1}$ where $\Sigma_{u_H^\perp}$ is the covariance of the forecast errors.

Table 1: Finite-sample size of local kernel-based tests ($\alpha = 5\%$, $T = 250$)

	Local inference at fixed τ						Global reference		
Instrument strength	$S^{loc}_{0.05}$	$K^{loc}_{0.05}$	$S^{loc}_{0.15}$	$K^{loc}_{0.15}$	$S^{loc}_{0.50}$	$K^{loc}_{0.50}$	S	K	Wald
Uninformative	0.035	0.061	0.039	0.052	0.044	0.054	0.076	0.077	0.924
$\sigma_i = 0.1$	0.023	0.055	0.029	0.046	0.047	0.070	0.070	0.064	0.955
$\sigma_i = 0.25$	0.032	0.057	0.047	0.062	0.039	0.057	0.073	0.062	0.798
$\sigma_i = 0.5$	0.032	0.044	0.055	0.057	0.036	0.038	0.059	0.061	0.445
$\sigma_i = 1$	0.029	0.031	0.033	0.035	0.035	0.043	0.068	0.058	0.078

Table 2: Finite-sample size of local kernel-based tests ($\alpha = 5\%$, $T = 500$)

	Local inference at fixed τ						Global reference		
Instrument strength	$S^{loc}_{0.05}$	$K^{loc}_{0.05}$	$S^{loc}_{0.15}$	$K^{loc}_{0.15}$	$S^{loc}_{0.50}$	$K^{loc}_{0.50}$	S	K	Wald
Uninformative	0.044	0.047	0.038	0.047	0.048	0.062	0.068	0.062	0.980
$\sigma_i = 0.1$	0.043	0.054	0.047	0.050	0.037	0.048	0.068	0.056	0.978
$\sigma_i = 0.25$	0.034	0.039	0.048	0.046	0.042	0.035	0.052	0.055	0.888
$\sigma_i = 0.5$	0.060	0.056	0.061	0.054	0.043	0.038	0.053	0.055	0.472
$\sigma_i = 1$	0.043	0.045	0.036	0.047	0.045	0.048	0.057	0.066	0.049

Results for local inference Tables 1–2 report rejection frequencies for local kernel-based inference at $\tau \in \{0.05, 0.15, 0.50\}$ and for three global benchmark tests (full-sample S, full-sample K, and a conventional Wald test based on strong-instrument asymptotics). Overall, the kernel-based local tests exhibit good size control across all instrument-strength configurations, including the uninformative-instrument design. For $T = 250$, the rejection frequencies of the local S and K tests range between 2.3% and 7.0% across the three evaluation points and all values of σ_i, with no systematic deterioration as instruments weaken. For $T = 500$, the corresponding rejection frequencies range between 3.4% and 6.2%, indicating modest improvement in finite-sample calibration as T increases.

The global benchmark statistics reinforce the weak-identification interpretation of the design. The full-sample S and K statistics remain well-calibrated, with rejection frequencies close to the nominal 5% level in both sample sizes. In contrast, the Wald test massively overrejects under weak instruments: for $T = 250$ it rejects with probability 0.924 in the uninformative case and 0.955 at $\sigma_i = 0.1$, only declining to 0.078 at $\sigma_i = 1$. For $T = 500$, Wald overrejection is similarly severe (e.g., 0.980 in the uninformative case and 0.978 at $\sigma_i = 0.1$), falling to 0.049 at $\sigma_i = 1$. These results confirm that the simulation design induces a pronounced weak-identification problem for standard asymptotic approximations, while the proposed kernel-based local procedures retain stable size.

Results for stability tests Tables 3 and 4 report rejection frequencies for the integrated stability tests based on Gaussian-process (GP) critical values for $\mathcal{S}_T^{\text{stab}}$ and $\mathcal{K}_T^{\text{stab}}$,

together with the reference stability statistic $\tilde{q}_{LL}$ from Magnusson and Mavroeidis (2014). Each test is evaluated both at the true parameter value β_0 (Panel A) and at the plug-in estimator $\hat{\beta}$ (Panel B). Across all instrument strengths, the GP-calibrated stability tests are close to nominal size. For $T = 250$, $\mathcal{S}_T^{\text{stab}}$ rejects between 3.0% and 5.5% and $\mathcal{K}_T^{\text{stab}}$ between 3.6% and 6.1%, with similarly well-calibrated behavior for $\tilde{q}_{LL}$. For $T = 500$, rejection frequencies move closer to 5% overall, with $\mathcal{S}_T^{\text{stab}}$ between 3.5% and 4.8% and $\mathcal{K}_T^{\text{stab}}$ between 3.8% and 5.1%. Importantly, evaluating the stability statistics at $\hat{\beta}$ rather than β_0 does not generate a systematic size distortion in either sample size.

Table 3: Finite-sample size of stability tests ($\alpha = 5\%$, $T = 250$)

Panel A: evaluation at β_0

Instrument	$\mathcal{S}^{\text{stab}}$	$\mathcal{K}^{\text{stab}}$	$\tilde{q}_{LL}$
Uninformative	0.037	0.038	0.060
$\sigma_i = 0.1$	0.047	0.053	0.063
$\sigma_i = 0.25$	0.055	0.061	0.073
$\sigma_i = 0.5$	0.054	0.057	0.070
$\sigma_i = 1$	0.033	0.047	0.050

Panel B: evaluation at $\hat{\beta}$

Instrument	$\mathcal{S}^{\text{stab}}$	$\mathcal{K}^{\text{stab}}$	$\tilde{q}_{LL}$
Uninformative	0.032	0.039	0.049
$\sigma_i = 0.1$	0.036	0.046	0.061
$\sigma_i = 0.25$	0.034	0.037	0.058
$\sigma_i = 0.5$	0.041	0.061	0.063
$\sigma_i = 1$	0.030	0.036	0.048

Notes: Monte Carlo rejection frequencies under the null of parameter stability based on 5,000 replications. Panel A evaluates all statistics at the true parameter value β_0, while Panel B uses the plug-in estimator $\hat{\beta}$. The statistic $\tilde{q}_{LL}$ denotes the stability component of the generalized S test of Magnusson and Mavroeidis (2014).

Table 4: Finite-sample size of stability tests ($\alpha = 5\%$, $T = 500$)

Panel A: evaluation at β_0

Instrument	$\mathcal{S}^{\text{stab}}$	$\mathcal{K}^{\text{stab}}$	$\tilde{q}_{LL}$
Uninformative	0.039	0.051	0.057
$\sigma_i = 0.1$	0.041	0.051	0.052
$\sigma_i = 0.25$	0.044	0.042	0.051
$\sigma_i = 0.5$	0.048	0.048	0.063
$\sigma_i = 1$	0.038	0.040	0.052

Panel B: evaluation at $\hat{\beta}$

Instrument	$\mathcal{S}^{\text{stab}}$	$\mathcal{K}^{\text{stab}}$	$\tilde{q}_{LL}$
Uninformative	0.039	0.051	0.051
$\sigma_i = 0.1$	0.036	0.038	0.052
$\sigma_i = 0.25$	0.042	0.050	0.051
$\sigma_i = 0.5$	0.045	0.045	0.060
$\sigma_i = 1$	0.035	0.042	0.051

Notes: Monte Carlo rejection frequencies under the null of parameter stability based on 5,000 replications. Panel A evaluates all statistics at the true parameter value β_0, while Panel B uses the plug-in estimator $\hat{\beta}$. The statistic $\tilde{q}_{LL}$ denotes the stability component of the generalized S test of Magnusson and Mavroeidis (2014).

In contrast, Tables 5 and 6 show that using asymptotic normal critical values for the standardized stability statistics leads to substantial size distortions. At $T = 250$, rejection frequencies for $\mathcal{S}_T^{\text{stab}}$ are between 0.4% and 1.2% and for $\mathcal{K}_T^{\text{stab}}$ between 1.6% and 3.7%, depending on whether the statistic is evaluated at β_0 or $\hat{\beta}$. These distortions remain pronounced at $T = 500$, with $\mathcal{S}_T^{\text{stab}}$ rejecting between 0.4% and 0.9% and $\mathcal{K}_T^{\text{stab}}$ between 2.1% and 2.7%. This comparison underscores the practical value of the Gaussian-process representation for finite-sample implementation: GP-based critical values deliver accurate

size in empirically relevant sample sizes, while the normal approximation leads to severe underrejection. We note that the asymptotic approximation does get considerably more accurate for bandwidth choices that are smaller, but rejection frequencies still remain below the nominal size.[7]

Table 5: Finite-sample size of stability tests based on asymptotic normal distribution ($\alpha = 5\%$, $T = 250$)

Panel A: evaluation at β_0

Instrument	$\mathcal{S}^{\text{stab}}$	$\mathcal{K}^{\text{stab}}$
Uninformative	0.006	0.022
$\sigma_i = 0.1$	0.008	0.027
$\sigma_i = 0.25$	0.008	0.037
$\sigma_i = 0.5$	0.012	0.034
$\sigma_i = 1$	0.007	0.022

Panel B: evaluation at $\hat{\beta}$

Instrument	$\mathcal{S}^{\text{stab}}$	$\mathcal{K}^{\text{stab}}$
Uninformative	0.004	0.016
$\sigma_i = 0.1$	0.004	0.023
$\sigma_i = 0.25$	0.006	0.021
$\sigma_i = 0.5$	0.009	0.029
$\sigma_i = 1$	0.006	0.023

Notes: Monte Carlo rejection frequencies under the null of parameter stability based on 5,000 replications. Panel A evaluates all statistics at the true parameter value β_0, while Panel B uses the plug-in estimator $\hat{\beta}$. Critical values are obtained from the asymptotic normal approximation in Theorem 1(iii)(c) and Corollary 1(ii).

Table 6: Finite-sample size of stability tests based on asymptotic normal distribution ($\alpha = 5\%$, $T = 500$)

Panel A: evaluation at β_0

Instrument	$\mathcal{S}^{\text{stab}}$	$\mathcal{K}^{\text{stab}}$
Uninformative	0.005	0.025
$\sigma_i = 0.1$	0.007	0.025
$\sigma_i = 0.25$	0.008	0.021
$\sigma_i = 0.5$	0.008	0.024
$\sigma_i = 1$	0.004	0.025

Panel B: evaluation at $\hat{\beta}$

Instrument	$\mathcal{S}^{\text{stab}}$	$\mathcal{K}^{\text{stab}}$
Uninformative	0.008	0.027
$\sigma_i = 0.1$	0.005	0.024
$\sigma_i = 0.25$	0.009	0.025
$\sigma_i = 0.5$	0.008	0.025
$\sigma_i = 1$	0.006	0.023

Notes: Monte Carlo rejection frequencies under the null of parameter stability based on 5,000 replications. Panel A evaluates all statistics at the true parameter value β_0, while Panel B uses the plug-in estimator $\hat{\beta}$. Critical values are obtained from the asymptotic normal approximation in Theorem 1(iii)(c) and Corollary 1(ii).

Overall, these results suggest that the proposed kernel-based procedures achieve reliable finite-sample size control in empirically relevant sample sizes, even in designs with severely weak instruments. This finding complements the asymptotic weak-identification theory developed in Sections 2 and 3 and motivates the power analysis reported next.

[7] Acknowledging the same finite-sample problems, Chen and Hong (2012) suggest a Bootstrap procedure. Given that the fixed h asymptotics work very well we abstract from exploring this avenue in this paper.

4.2 Power

We next study the ability of the proposed kernel-based stability tests to detect smooth time variation in the NKPC slope. The power simulations are based on the same structural environment as in the size analysis, except that we set the model to a purely backward-looking specification ($\gamma_f = 0$) and allow the slope coefficient to evolve deterministically over the sample while keeping its average level comparable to the constant-parameter benchmark. The restriction to a backward-looking NKPC is not for computational convenience: in a forward-looking model with time-varying λ_t, the rational-expectations equilibrium depends on agents' information about the future path of λ_t and different model choices yield different reduced-form dynamics (Kulish and Pagan, 2017). By setting $\gamma_f = 0$ we circumvent the problem since our focus is on the statistical properties of the stability tests.

Data generating process under the alternative. We maintain the NKPC (39), the AR(2) forcing process (40), and the cost-push shock dynamics from the size design, but simplify the model to a purely backward looking. Specifically,

$$y_t = \gamma_b y_{t-1} + \lambda_t x_t + e_t, \tag{41}$$

$$x_t = \rho_1 x_{t-1} + \rho_2 x_{t-2} + \varepsilon_t^i + \nu e_t, \tag{42}$$

$$e_t = \rho_e e_{t-1} + \sqrt{1-\rho_e^2}\,\xi_t, \qquad \xi_t \sim \mathcal{N}(0,1), \tag{43}$$

with $\gamma_b = 0.6$, $(\rho_1, \rho_2) = (1.2, -0.4)$, $\nu = -1$, and $\rho_e = 0.5$ as before. Instrument strength is varied by $\varepsilon_t^i \sim \mathcal{N}(0, \sigma_i^2)$ with $\sigma_i \in \{0.1, 0.25, 0.5, 1\}$, and we also consider the completely uninformative case in which $z_t \sim \mathcal{N}(0,1)$ is independent of all structural shocks. The departure from the null is introduced through a smoothly time-varying Phillips-curve slope λ_t. Let $t = 1, \ldots, T$ index the sample after burn-in. We set

$$\lambda_t \;=\; 0.7 \;-\; \frac{0.6}{1+\exp(-\kappa_t)}, \qquad \kappa_t \equiv -5 + 10 \cdot \frac{t-1}{T-1}, \tag{44}$$

which generates a monotone and smooth decline in λ_t from approximately 0.7 at the beginning of the sample to approximately 0.1 at the end. The path is centered so that its average is approximately equal to the constant-parameter value used in the size analysis,

$$\frac{1}{T}\sum_{t=1}^{T} \lambda_t \approx 0.4, \tag{45}$$

thereby isolating power against parameter *instability* rather than against a level shift in the average slope.

Similar to the size study, we consider $T \in \{250, 500\}$ and generate 5,000 replications for each value of σ_i and the uninformative-instrument case. All other elements of the

estimation procedure are as in the size simulations: we use the SP-IV GMM estimator of Lewis and Mertens (2022) with horizons up to $H = 20$ and thinning $h_{\text{thin}} = 2$, and we construct forecast errors using a VAR(4). For the kernel-based statistics, we use the Epanechnikov kernel and the plug-in bandwidth $h_T = 0.8\,T^{-1/5}$. The null hypothesis for stability is $\beta(t) = \beta_0$ for all t with $\beta_0 = (\lambda_{ss}, \gamma_b, \gamma_f)'$, while the alternative is the smooth time variation in λ_t implied by (44). We report rejection frequencies for the integrated kernel stability tests $\mathcal{S}_T^{\text{stab}}$ and $\mathcal{K}_T^{\text{stab}}$ evaluated both at β_0 and at the plug-in estimator $\hat{\beta}$, and we compare them to the stability statistic $\tilde{q}_{LL}$ of Magnusson and Mavroeidis (2014). Unless noted otherwise, critical values for $\mathcal{S}_T^{\text{stab}}$ and $\mathcal{K}_T^{\text{stab}}$ are obtained from the Gaussian-process representation corresponding to the fixed-h asymptotics.

Table 7: Finite-sample power of local kernel-based tests ($\alpha = 5\%$, $T = 500$)

	Local inference at fixed τ						Global reference		
Instrument strength	$S^{loc}_{0.05}$	$K^{loc}_{0.05}$	$S^{loc}_{0.15}$	$K^{loc}_{0.15}$	$S^{loc}_{0.50}$	$K^{loc}_{0.50}$	S	K	Wald
Uninformative	0.037	0.058	0.052	0.059	0.048	0.044	0.063	0.064	0.134
$\sigma_i = 0.1$	0.063	0.079	0.067	0.089	0.027	0.040	0.071	0.051	0.100
$\sigma_i = 0.25$	0.131	0.148	0.178	0.210	0.042	0.038	0.054	0.056	0.012
$\sigma_i = 0.5$	0.362	0.434	0.526	0.623	0.042	0.044	0.058	0.063	0.000
$\sigma_i = 1$	0.779	0.876	0.909	0.967	0.040	0.038	0.059	0.057	0.000

Results for local inference under the alternative. Table 7 reports rejection frequencies for the local kernel-based tests under the smooth-break alternative for $T = 500$. As expected, the local tests exhibit power only at evaluation points τ where the time-varying parameter λ_t deviates substantially from its steady-state value. At $\tau = 0.05$ and $\tau = 0.15$—near the beginning of the sample where λ_t is elevated relative to λ_{ss}—the local S and K statistics show increasing power as instrument strength improves: rejection frequencies rise from around 5%–6% in the uninformative case to 77.9% and 87.6% at $\tau = 0.05$, and to 90.9% and 96.7% at $\tau = 0.15$, for $\sigma_i = 1$. In contrast, at $\tau = 0.50$—where λ_t passes through its average value close to λ_{ss}—rejection frequencies remain near the nominal size across all instrument configurations, as expected under a centered alternative. The full-sample S and K statistics show no meaningful power against this smooth alternative, reflecting their insensitivity to parameter paths that average out to the null. The Wald test exhibits apparent "power" under weak instruments, but this reflects severe size distortion rather than genuine detection of instability.

Table 8: Finite-sample power of stability tests ($\alpha = 5\%$, $T = 500$)

Panel A: evaluation at β_0

Instrument	$\mathcal{S}^{\text{stab}}$	$\mathcal{K}^{\text{stab}}$	$\tilde{q}_{LL}$
Uninformative	0.060	0.104	0.099
$\sigma_i = 0.1$	0.065	0.101	0.092
$\sigma_i = 0.25$	0.118	0.233	0.151
$\sigma_i = 0.5$	0.301	0.605	0.297
$\sigma_i = 1$	0.862	0.991	0.802

Panel B: evaluation at $\hat{\beta}$

Instrument	$\mathcal{S}^{\text{stab}}$	$\mathcal{K}^{\text{stab}}$	$\tilde{q}_{LL}$
Uninformative	0.085	0.164	0.085
$\sigma_i = 0.1$	0.107	0.200	0.097
$\sigma_i = 0.25$	0.192	0.387	0.170
$\sigma_i = 0.5$	0.423	0.678	0.363
$\sigma_i = 1$	0.895	0.991	0.809

Notes: Monte Carlo rejection frequencies under the smooth-break alternative based on 5,000 replications. Panel A evaluates all statistics at the true steady-state parameter value β_0, while Panel B uses the plug-in estimator $\hat{\beta}$. Critical values for $\mathcal{S}^{\text{stab}}$ and $\mathcal{K}^{\text{stab}}$ are obtained from the Gaussian-process representation. The statistic $\tilde{q}_{LL}$ denotes the stability component of the generalized S test of Magnusson and Mavroeidis (2014).

Results for stability tests. Table 8 reports rejection frequencies for the integrated stability tests under the smooth-break alternative. Two main findings emerge. First, the integrated S-statistic $\mathcal{S}_T^{\text{stab}}$ and the $\tilde{q}_{LL}$ statistic of Magnusson and Mavroeidis (2014) exhibit similar power profiles across instrument strengths. At $\sigma_i = 1$, $\mathcal{S}_T^{\text{stab}}$ rejects with probability 0.862 when evaluated at β_0 and 0.895 at $\hat{\beta}$, compared to 0.802 and 0.809 for $\tilde{q}_{LL}$. At weaker instrument configurations, both statistics show comparable rejection frequencies: for $\sigma_i = 0.5$, $\mathcal{S}_T^{\text{stab}}$ achieves 30.1% (Panel A) and 42.3% (Panel B), while $\tilde{q}_{LL}$ achieves 29.7% and 36.3%, respectively.

Second, the integrated K-statistic $\mathcal{K}_T^{\text{stab}}$ delivers substantially higher power than both $\mathcal{S}_T^{\text{stab}}$ and $\tilde{q}_{LL}$. At $\sigma_i = 0.5$, $\mathcal{K}_T^{\text{stab}}$ rejects with probability 0.605 at β_0 and 0.678 at $\hat{\beta}$, roughly doubling the power of $\mathcal{S}_T^{\text{stab}}$. At $\sigma_i = 1$, $\mathcal{K}_T^{\text{stab}}$ achieves near-complete power (0.991) in both panels, compared to 0.862–0.895 for $\mathcal{S}_T^{\text{stab}}$. The power advantage of $\mathcal{K}_T^{\text{stab}}$ is most pronounced at intermediate instrument strengths ($\sigma_i \in \{0.25, 0.5\}$), where it approximately doubles the rejection frequency relative to the S-based alternatives. This reflects the efficiency gains from exploiting the Jacobian information in the K-statistic, which becomes increasingly valuable as identification strengthens.

Table 9: Finite-sample power of stability tests based on asymptotic normal distribution ($\alpha = 5\%$, $T = 500$)

Panel A: evaluation at β_0

Instrument	$\mathcal{S}^{\text{stab}}$	$\mathcal{K}^{\text{stab}}$
Uninformative	0.008	0.067
$\sigma_i = 0.1$	0.013	0.069
$\sigma_i = 0.25$	0.030	0.177
$\sigma_i = 0.5$	0.104	0.514
$\sigma_i = 1$	0.664	0.981

Panel B: evaluation at $\hat{\beta}$

Instrument	$\mathcal{S}^{\text{stab}}$	$\mathcal{K}^{\text{stab}}$
Uninformative	0.019	0.111
$\sigma_i = 0.1$	0.024	0.135
$\sigma_i = 0.25$	0.052	0.309
$\sigma_i = 0.5$	0.197	0.620
$\sigma_i = 1$	0.705	0.980

Notes: Monte Carlo rejection frequencies under the smooth-break alternative based on 5,000 replications. Panel A evaluates all statistics at the true steady-state parameter value β_0, while Panel B uses the plug-in estimator $\hat{\beta}$. Critical values are obtained from the asymptotic normal approximation.

Table 9 reports rejection frequencies when critical values are instead obtained from the asymptotic normal approximation. Comparing these results to Table 8 reveals that the normal approximation leads to a substantial loss of power, particularly for $\mathcal{S}_T^{\text{stab}}$. At $\sigma_i = 0.5$, $\mathcal{S}_T^{\text{stab}}$ rejects with probability 0.104 (Panel A) and 0.197 (Panel B) under normal critical values, compared to 0.301 and 0.423 under GP critical values—a reduction of roughly one-half to two-thirds. This power loss reflects the severe underrejection under the null documented in the size analysis: the normal approximation produces overly conservative critical values, which translates directly into reduced power under the alternative. The K-statistic is less affected because its size distortion under normal critical values is smaller, but it still exhibits meaningful power losses at intermediate instrument strengths.

5 Empirical Application

In this section, we apply our kernel-based inference framework to re-examine the stability of the New Keynesian Phillips Curve (NKPC). We revisit the analysis of Lewis and Mertens (2022), who utilize the "System Projection on Instrumental Variables" (SP-IV) estimator to identify the structural relationship between inflation and unemployment. While their analysis assumes constant structural parameters, the substantial changes in the U.S. macroeconomy over the last few decades—including the Great Moderation, the Great Recession, and the recent post-pandemic inflation surge—suggest that the parameters governing the inflation-unemployment trade-off may have evolved over time.

5.1 Model Specification and Stability Testing

Following Lewis and Mertens (2022), we consider a monthly specification of the Hybrid New Keynesian Phillips Curve. The full structural relationship is given by:

$$\pi_t^{1m} = \gamma_b \pi_{t-1}^{1y} + \gamma_f \pi_{t+12}^{1y} + \lambda U_t + u_t, \tag{46}$$

where π_t^{1m} is the annualized monthly percent change in Core CPI, π_t^{1y} is the percent change in Core CPI over the preceding year (a proxy for backward-looking expectations), π_{t+12}^{1y} is the realized future annual inflation (proxying for forward-looking expectations under rational expectations), and U_t is the headline unemployment rate (see Lewis and Mertens, 2022).

As is common in the literature, we impose the restriction that there is no long-run trade-off between inflation and unemployment, which implies that the coefficients on lagged and future inflation sum to unity:

$$\gamma_b + \gamma_f = 1. \tag{47}$$

Applying this restriction to equation (46) yields the two-parameter specification used for estimation:

$$\pi_t^{1m} - \pi_{t-1}^{1y} = \gamma_f(\pi_{t+12}^{1y} - \pi_{t-1}^{1y}) + \lambda U_t + u_t. \tag{48}$$

In this restricted form, the dependent variable is the change in inflation relative to the backward-looking trend, and the regressors are the "future inflation spread" and the unemployment rate. We estimate the model using U.S. monthly data. To address the endogeneity of unemployment and inflation expectations, we employ the Main Business Cycle (MBC) shock identified by Angeletos et al. (2020) as an instrument. As in Lewis and Mertens (2022), the estimation is based on the SP-IV framework, which identifies structural parameters by projecting the impulse response functions (IRFs) of the dependent variable onto the IRFs of the regressors. We utilize IRFs at horizons $h \in \{0, 3, 6, \ldots, 33\}$, yielding a total of $L = 12$ moment conditions for the $K = 2$ structural parameters (γ_f, λ).

We begin by formally testing the null hypothesis of parameter stability, $H_0 : \beta(\tau) = \beta_0$ for all $\tau \in (0, 1)$, against the alternative of smooth time variation. We employ the integrated kernel stability tests proposed in Section 2: the generalized stability S-statistic, $\mathcal{S}_T^{\text{stab}}$, and the stability K-statistic, $\mathcal{K}_T^{\text{stab}}$.

The tests are implemented using the Epanechnikov kernel. We consider two bandwidth specifications to ensure robustness: a smaller bandwidth of $h = 0.15$ and a data-driven plug-in bandwidth of $h = 0.23$ (corresponding to the rule-of-thumb $0.8 \cdot T^{-1/5}$). Critical values are obtained from the simulated distribution of the limit Gaussian bridge process under the null. In addition to the rule-of-thumb bandwidths, we select a data-driven smoothing parameter using a rolling-origin (expanding window) cross-validation procedure adapted to the time-series setting. Let $g_t(\beta) \in \mathbb{R}^L$ denote the vector of SP-

IV moment conditions at horizon grid $h \in \{0, 3, \ldots, 33\}$ evaluated at parameter value $\beta = (\gamma_f, \lambda)'$. For each candidate bandwidth $h \in \mathcal{H}$, we consider a collection of training endpoints T_i (expressed in fractions of the sample) and proceed as follows. First, we estimate $\widehat{\beta}_h(\tau_i)$ using only information available up to T_i via kernel weights centered at $\tau_i = T_i/T$ (left-sided weights, renormalized on the training sample). Second, we evaluate moment validity on a subsequent validation window $t \in \{T_i + 1, \ldots, T_i + V\}$. Let $\bar{g}_{h,i}^{\text{val}}$ denote the kernel-weighted average of $g_t(\widehat{\beta}_h(\tau_i))$ over the validation window, and let $\widehat{\Sigma}_{h,i}^{\text{HAC}}$ denote a (kernel-weighted) HAC covariance estimate of $g_t(\widehat{\beta}_h(\tau_i))$ computed on the training sample. The cross-validation objective is the average out-of-sample GMM fit,

$$\text{CV}(h) \;=\; \frac{1}{I} \sum_{i=1}^{I} \left(\bar{g}_{h,i}^{\text{val}}\right)' \left(\widehat{\Sigma}_{h,i}^{\text{HAC}}\right)^{-1} \left(\bar{g}_{h,i}^{\text{val}}\right), \tag{49}$$

and we select $\widehat{h} = \arg\min_{h \in \mathcal{H}} \text{CV}(h)$. In our application, this procedure selects a comparatively smooth bandwidth of $\widehat{h} = 0.70$. Specifically, the validation-window kernel weights $w_{t,T}(\tau_i)$ in $\bar{g}_{h,i}^{\text{val}}$ are centered at $\tau_i = T_i/T$ and renormalized to sum to one over the validation indices $t \in \{T_i + 1, \ldots, T_i + V\}$; the HAC estimator $\widehat{\Sigma}_{h,i}^{\text{HAC}}$ uses a Newey–West kernel applied to the training-sample scores with kernel weights centered at τ_i.

Table 10: Tests for Parameter Stability in the NKPC

	Bandwidth $h = 0.15$		**Bandwidth** $h = 0.23$		**Bandwidth** $h = 0.70$ (CV)	
Test Statistic	**Value**	**p-value**	**Value**	**p-value**	**Value**	**p-value**
$\mathcal{S}_T^{\text{stab}}$ (12 df)	14.86	0.0158	14.53	0.0180	7.17	0.0070
$\mathcal{K}_T^{\text{stab}}$ (2 df)	3.95	0.0136	4.38	0.0118	2.41	0.0098

Notes: The table reports p-values for the integrated kernel stability tests applied to the NKPC specification in equation (48). $\mathcal{S}_T^{\text{stab}}$ tests stability of the full moment vector ($L = 12$ degrees of freedom), while $\mathcal{K}_T^{\text{stab}}$ tests stability of the structural parameters projected onto the Jacobian ($K = 2$ degrees of freedom). P-values are based on 5,000 simulations of the limit Gaussian bridge process under the null. The column labeled "CV" uses the bandwidth $\widehat{h} = 0.70$ selected by rolling-origin cross-validation as described in equation (49).

Table 10 presents the results. We find strong evidence against the null hypothesis of constant parameters. Using the bandwidth $h = 0.15$, the integrated S-statistic yields a p-value of 0.0158, and the K-statistic yields a p-value of 0.0136. These results are robust to the choice of bandwidth; for the larger bandwidth $h = 0.23$, the p-values remain well below the 5% significance level (0.0180 and 0.0118, respectively). Finally, while the raw integrated statistics are smaller for the smoother bandwidth, their null distribution also changes with h; the corresponding Gaussian-bridge critical values imply rejections at conventional levels across all three bandwidth choices.

The rejection of the null by the $\mathcal{K}_T^{\text{stab}}$ statistic is particularly informative. Since this test focuses power specifically on the directions of the structural parameters (γ_f, λ) rather

than the full set of overidentifying restrictions, the result suggests that the rejection is driven by actual instability in the structural relationship rather than mere misspecification of the instrument at specific horizons. As discussed in Section 2.3, a rejection of the stability null can in principle reflect either parameter instability under maintained instrument exogeneity, or local violation of the exclusion restriction with stable parameters. For the MBC shock instrument of Angeletos et al. (2020), the exogeneity assumption rests on the structural VAR identification maintained throughout the sample. While we cannot rule out time-varying instrument validity, the consistency of the rejections across bandwidth choices and the systematic drift visible in the local estimates (Section 5.2) point toward genuine structural change rather than intermittent instrument failure. This substantial evidence of instability motivates the use of our local kernel-weighted estimator to track the evolution of the Phillips curve slope over time.

5.2 Local weak-identification robust inference

The stability tests in Section 5.1 provide consistent evidence against the null of constant NKPC parameters. This rejection motivates replacing full-sample inference with a local approach that allows the structural relationship to evolve smoothly over time. In this subsection we therefore implement kernel-weighted SP-IV estimation and weak-identification robust local confidence sets for the NKPC parameters.

We follow Lewis and Mertens (2022) in imposing VAR dynamics when constructing both the SP-IV point estimates and the corresponding weak-identification robust test statistic. Concretely, for each rescaled time index $\tau \in (0,1)$ we estimate the reduced-form VAR locally using kernel weights with bandwidth $\widehat{h} = 0.70$ (selected by rolling-origin cross-validation as described in equation (49)). Using the resulting local VAR parameters, we compute the impulse response functions (IRFs) of (i) the NKPC left-hand-side inflation transformation and (ii) the right-hand-side regressors, all with respect to the Angeletos et al. (2020) Main Business Cycle (MBC) shock instrument. Denoting by $\widehat{\theta}_y(\tau) \in \mathbb{R}^L$ the vector of local IRFs of the transformed dependent variable and by $\widehat{\Theta}_Y(\tau) \in \mathbb{R}^{L\times K}$ the matrix stacking the corresponding local IRFs of the regressors (evaluated at horizons $h \in \{0, 3, \ldots, 33\}$ so that $L = 12$), the local SP-IV estimator is obtained from the (local) projection

$$\widehat{\beta}(\tau) \;=\; \arg\min_{\beta\in\mathbb{R}^K} \left\|\widehat{\theta}_y(\tau) - \widehat{\Theta}_Y(\tau)\beta\right\|^2 \;=\; \left(\widehat{\Theta}_Y(\tau)'\widehat{\Theta}_Y(\tau)\right)^{-1}\widehat{\Theta}_Y(\tau)'\widehat{\theta}_y(\tau), \qquad (50)$$

where $\beta = (\gamma_f, \lambda)'$ and $K = 2$. Relative to the constant-parameter estimator, equation (50) traces the evolution of the NKPC mapping as τ varies, while maintaining the discipline that the horizon-specific objects entering SP-IV are generated by an internally consistent (local) VAR, as in Lewis and Mertens (2022).

Forecast-error horizons and localization. The SP-IV moments at horizon h use forecast errors extending h periods into the future. With our longest horizon of 33 months, the kernel average incorporates information from nearly three years ahead. This creates smoothing effects and may introduce censoring near the right boundary. The cross-validated bandwidth $\widehat{h} = 0.70$ is large enough that this horizon-induced smoothing is modest relative to the kernel window. The VAR dynamics ensure internal consistency of forecast errors. Interpretation of local estimates near the sample endpoints should account for this forward-looking data dependence.

Pointwise inference for $\beta(\tau)$ may be distorted when the instrument is weak for some horizons or subsamples. To address this issue, we construct local confidence sets by inverting the kernel-weighted K-statistic developed in Section 3. In our implementation, the local moments correspond to the SP-IV orthogonality conditions evaluated at the horizon grid, and the local Jacobian (required for the K-statistic) is computed under the same imposed VAR dynamics used for estimation. For each τ and candidate parameter value β, we compute the local K-statistic $K_T(\tau, \beta)$ based on kernel-weighted moments, using the effective sample size implied by the kernel weights. The $(1 - \alpha)$ pointwise confidence set is then

$$\mathcal{C}_{1-\alpha}(\tau) \;=\; \Big\{\beta \in \mathbb{R}^2 :\; K_T(\tau, \beta) \leq \chi^2_{K,\,1-\alpha}\Big\}, \tag{51}$$

with $K = 2$ degrees of freedom. By construction, these sets remain valid uniformly over local instrument strength in the sense characterized by the weak-identification asymptotics underlying the K-statistic, while preserving the dynamic restrictions embedded in the VAR/SP-IV specification of Lewis and Mertens (2022).

Figure 1 summarizes the local inference results at eight representative dates spanning the sample (April 1982, August 1986, December 1990, April 1995, July 1999, November 2003, March 2008, and July 2012). In each panel, we report the kernel-smoothed local SP-IV point estimate $\widehat{\beta}(\tau)$ together with the corresponding weak-identification robust confidence set obtained from inverting the local K-statistic in (51). For comparison, we also overlay the constant-parameter estimate and its associated confidence region.

Two systematic patterns emerge. First, the estimated slope parameter $\lambda(\tau)$ becomes markedly less negative over time, indicating a progressive flattening of the inflation–unemployment trade-off. Second, the forward-looking component $\gamma_f(\tau)$ increases, consistent with a NKPC that becomes more forward-looking in later subsamples. Importantly, these conclusions are supported by confidence sets that explicitly account for weak identification at the local level; while uncertainty remains material in some periods, the drift in the location of the admissible region over time is clearly visible. Taken together, the local estimates and robust confidence sets provide empirical content to the stability rejections in Table 10: the evidence against constancy is not merely statistical, but corresponds to economically meaningful changes in both the slope and the forward-looking weight of the

NKPC.

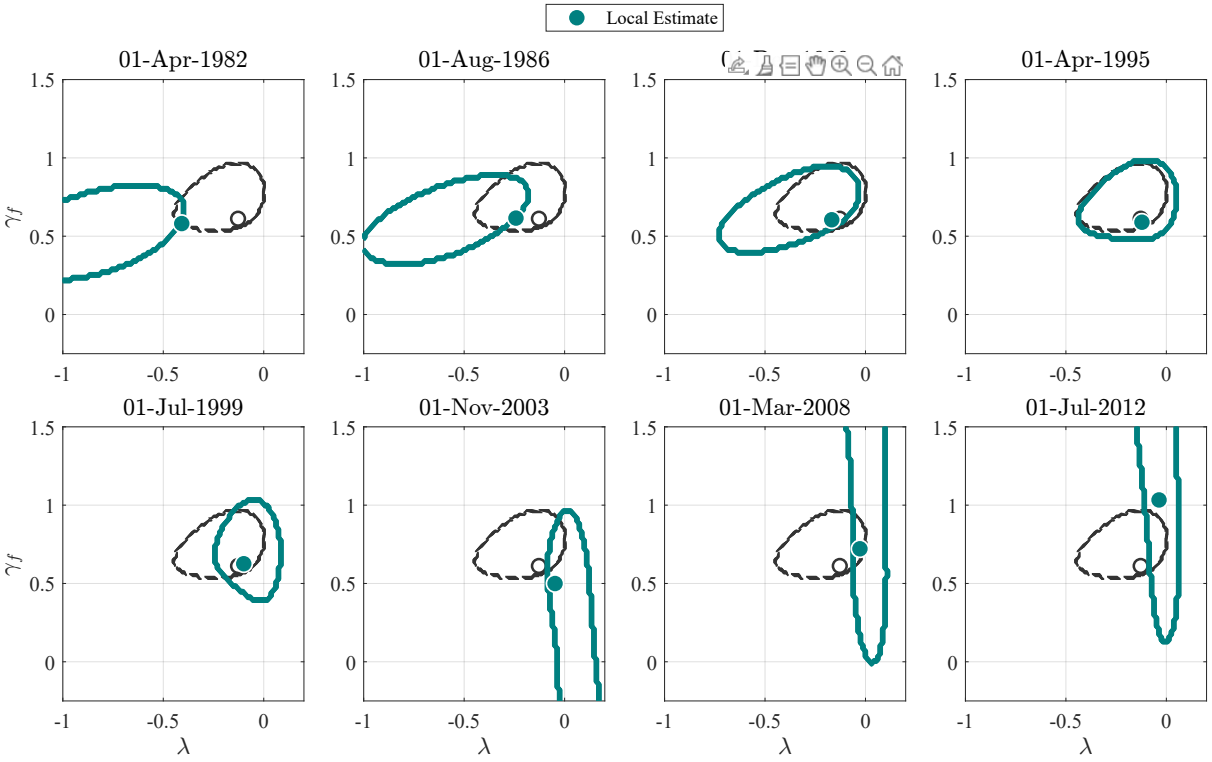


Figure 1: Local SP-IV estimates and weak-identification robust confidence sets for (γ_f, λ). Each panel reports the kernel-weighted local SP-IV point estimate $\widehat{\beta}(\tau)$ and the corresponding confidence set obtained by inverting the local K-statistic under imposed VAR dynamics, following Lewis and Mertens (2022). The bandwidth is $\widehat{h} = 0.70$ selected by rolling-origin cross-validation. The figure also overlays the constant-parameter estimate and its confidence region for comparison.

6 Conclusion

We develop a kernel-based GMM framework for testing parameter stability and conducting local-in-time inference in structural time-series models with potentially weak instruments. Building on the orthogonal decomposition of generalized S-statistics into full-sample and stability components established by Magnusson and Mavroeidis (2014) in the sample-splitting context, we show that an analogous decomposition holds for kernel-smoothed statistics, with a stability component whose pivotal limit directs power toward smooth or persistent departures from stability. The K-type version of the statistic projects onto the directions identified by the Jacobian, exploiting overidentifying information without requiring assumptions on the time-invariance of the identification subspace. We further show that the stability component remains pivotal when evaluated at a weakly identified estimator rather than at the hypothesized parameter value. When the stability tests reject, pointwise confidence sets for the time-varying parameter $\beta(\tau)$ are obtained by inverting the local kernel statistics and remain valid regardless of instrument strength.

The framework is well suited to modern macroeconomic applications that combine external instruments with overidentification, and provides a natural sequential workflow: test for stability first, then proceed to full-sample or local-in-time inference depending on the outcome.

References

Andrews, Donald W. K., "Tests for Parameter Instability and Structural Change With Unknown Change Point," *Econometrica*, 1993, *61* (4), 821–856.

Angeletos, George-Marios, Fabrice Collard, and Harris Dellas, "Business-Cycle Anatomy," *American Economic Review*, October 2020, *110* (10), 3030–70.

Antoine, Bertille and Otilia Boldea, "Efficient estimation with time-varying information and the New Keynesian Phillips Curve," *Journal of Econometrics*, 2018, *204* (2), 268–300.

Atkeson, Andrew, Lee E Ohanian et al., "Are Phillips curves useful for forecasting inflation?," *Federal Reserve bank of Minneapolis quarterly review*, 2001, *25* (1), 2–11.

Bai, Jushan and Pierre Perron, "Estimating and Testing Linear Models with Multiple Structural Changes," *Econometrica*, 1998, *66* (1), 47–78.

Ball, Laurence M and Sandeep Mazumder, "Inflation dynamics and the great recession," Technical Report, National Bureau of Economic Research 2011.

Barnichon, Regis and Geert Mesters, "Identifying modern macro equations with old shocks," *The Quarterly Journal of Economics*, 2020, *135* (4), 2255–2298.

Bergholt, Drago, Francesco Furlanetto, and Etienne Vaccaro-Grange, "Did monetary policy kill the Phillips curve? Some simple arithmetics," *Review of Economics and Statistics*, 2024, pp. 1–45.

Blanchard, Olivier, "The Phillips Curve: Back to the '60s?," *American Economic Review*, May 2016, *106* (5), 31–34.

Caner, Mehmet, "Boundedly pivotal structural change tests in continuous updating GMM with strong, weak identification and completely unidentified cases," *Journal of Econometrics*, 2007, *137* (1), 28–67.

_ , "Pivotal structural change tests in linear simultaneous equations with weak identification," *Econometric Theory*, 2011, *27* (2), 413–426.

Chen, Bin, "Modeling and Testing Smooth Structural Changes with Endogenous Regressors," *Journal of Econometrics*, 2015, *185* (1), 196–215.

_ **and Yongmiao Hong**, "Testing for smooth structural changes in time series models via nonparametric regression," *Econometrica*, 2012, *80* (3), 1157–1183.

Coibion, Olivier and Yuriy Gorodnichenko, "Is the Phillips Curve Alive and Well after All? Inflation Expectations and the Missing Disinflation," *American Economic Journal: Macroeconomics*, January 2015, *7* (1), 197–232.

Dahlhaus, Rainer, "Fitting time series models to nonstationary processes," *The annals of Statistics*, 1997, *25* (1), 1–37.

Elliott, Graham and Ulrich K Müller, "Efficient tests for general persistent time variation in regression coefficients," *The Review of Economic Studies*, 2006, *73* (4), 907–940.

Galí, Jordi and Luca Gambetti, "Has the US wage Phillips curve flattened? A semi-structural exploration," Technical Report, National Bureau of Economic Research 2019.

Galí, Jordi and Mark Gertler, "Inflation Dynamics: A Structural Econometric Analysis," *Journal of Monetary Economics*, 1999, *44* (2), 195–222.

_ **,** _ **, and J. David López-Salido**, "Robustness of the Estimates of the Hybrid New Keynesian Phillips Curve," *Journal of Monetary Economics*, 2005, *52* (6), 1107–1118.

Giraitis, Liudas, George Kapetanios, and Massimiliano Marcellino, "Time-varying instrumental variable estimation," *Journal of Econometrics*, 2021, *224* (2), 394–415.

Hall, Peter and Thomas E Wehrly, "A geometrical method for removing edge effects from kernel-type nonparametric regression estimators," *Journal of the American Statistical Association*, 1991, *86* (415), 665–672.

Hazell, Jonathon, Juan Herreno, Emi Nakamura, and Jón Steinsson, "The slope of the Phillips Curve: evidence from US states," *The Quarterly Journal of Economics*, 2022, *137* (3), 1299–1344.

Inoue, Atsushi, Barbara Rossi, and Yiru Wang, "Local projections in unstable environments," *Journal of Econometrics*, 2024, *244* (2), 105726.

_ **,** _ **, and** _ , "HAS THE PHILLIPS CURVE FLATTENED?," *Econometric Theory*, 2025, p. 1–46.

Kleibergen, Frank, "Testing parameters in GMM without assuming that they are identified," *Econometrica*, 2005, *73* (4), 1103–1123.

_ **and Sophocles Mavroeidis**, "Weak instrument robust tests in GMM and the new Keynesian Phillips curve," *Journal of Business & Economic Statistics*, 2009, *27* (3), 293–311.

Kulish, Mariano and Adrian Pagan, "Estimation and Solution of Models with Expectations and Structural Changes," *Journal of Applied Econometrics*, 2017, *32* (2), 255–274.

Lewis, Daniel J and Karel Mertens, "Dynamic Identification Using System Projections and Instrumental Variables," 2022.

Li, Haiqi, Jin Zhou, and Yongmiao Hong, "Estimating and testing for smooth structural changes in moment condition models," *Journal of Econometrics*, 2024, *246* (1), 105896.

Magnusson, Leandro M and Sophocles Mavroeidis, "Identification using stability restrictions," *Econometrica*, 2014, *82* (5), 1799–1851.

Mavroeidis, Sophocles, Mikkel Plagborg-Møller, and James H Stock, "Empirical evidence on inflation expectations in the New Keynesian Phillips Curve," *American Economic Journal: Journal of Economic Literature*, 2014, *52* (1), 124–188.

McLeay, Michael and Silvana Tenreyro, "Optimal inflation and the identification of the Phillips curve," *NBER macroeconomics annual*, 2020, *34* (1), 199–255.

Moreira, Marcelo J., "A Conditional Likelihood Ratio Test for Structural Models," *Econometrica*, 2003, *71* (4), 1027–1048.

Müller, Ulrich K and Philippe-Emmanuel Petalas, "Efficient estimation of the parameter path in unstable time series models," *The Review of Economic Studies*, 2010, *77* (4), 1508–1539.

Negro, Marco Del, Michele Lenza, Giorgio E Primiceri, and Andrea Tambalotti, "What's Up with the Phillips Curve?," *Brookings Papers on Economic Activity*, 2020, *2020* (1), 301–373.

Stock, James H and Jonathan H Wright, "GMM with weak identification," *Econometrica*, 2000, *68* (5), 1055–1096.

_ **and Mark W Watson**, "Why has US inflation become harder to forecast?," *Journal of Money, Credit and banking*, 2007, *39*, 3–33.

Stock, James H. and Mark W. Watson, "Slack and Cyclically Sensitive Inflation," *Journal of Money, Credit and Banking*, 2020, *52* (S2), 393–428.

Appendix

A Proofs

Throughout this appendix, all kernel-weighted averages are computed using Nadaraya–Watson weights $w_{t,T}(\tau)$ on the boundary-reflected (synthesized) sample described in Section 2.3. Boundary reflection ensures that the effective kernel window is symmetric at every $\tau \in (0,1)$, so that all points behave like interior points: the variance normalization is $R(K)$ uniformly over τ, and no boundary-specific constants appear. We therefore write $w_{t,T}(\tau)$, $\bar{g}_T(\tau)$, and $\tilde{g}_T(\tau)$ without further qualification; all weights are understood to be Nadaraya–Watson weights on the reflected sample.

A.1 Proof of Proposition 1

Fix $\tau \in (0,1)$ and write $\bar{g}_T(\tau) \equiv \bar{g}_T(\tau;\beta_0)$. Recall

$$S_T^{\mathrm{loc}}(\tau,\beta_0) = \frac{Th_T}{R(K)}\,\bar{g}_T(\tau)'\,\widehat{\Omega}_T^{-1}(\tau)\,\bar{g}_T(\tau).$$

Step 1: Local CLT. Let $\tilde{g}_t \equiv g_t(\beta_0) - E[g_t(\beta_0)]$. Under Assumptions 2–3(shrinking) and boundary reflection, which makes every $\tau \in (0,1)$ behave as an interior point, the Nadaraya–Watson mean estimator obeys a CLT with variance normalization $R(K)$:

$$\sqrt{\frac{Th_T}{R(K)}}\;\Omega(\tau)^{-1/2}\sum_{t=1}^{T} w_{t,T}(\tau)\,\tilde{g}_t \;\Rightarrow\; \mathcal{N}(0, I_L), \tag{A.1}$$

where $\Omega(\tau)$ is the local long-run covariance from Assumption 4(a). This is a direct vector-valued analogue of standard kernel-smoothing CLTs, where the local long-run covariance enters through the FCLT in Assumption 2 evaluated at rescaled time τ.

Step 2: Mean is zero under the null. Under $H_0^{\mathrm{loc}}(\tau) : \beta(\tau) = \beta_0$, the score satisfies $E[g_t(\beta_0)] = 0$, hence $E[\bar{g}_T(\tau)] = 0$. Moreover, while Assumption 1 allows for local-to-zero score drift under sequences, under the null the drift function satisfies $m(\cdot) \equiv 0$, so the centered and uncentered processes coincide asymptotically. Therefore, (A.1) yields

$$\sqrt{\frac{Th_T}{R(K)}}\;\Omega(\tau)^{-1/2}\bar{g}_T(\tau) \;\Rightarrow\; \mathcal{N}(0, I_L). \tag{A.2}$$

Step 3: Feasible weighting and quadratic form. By Assumption 4(a), $\widehat{\Omega}_T(\tau) \xrightarrow{p} \Omega(\tau)$, hence $\widehat{\Omega}_T(\tau)^{-1/2} \xrightarrow{p} \Omega(\tau)^{-1/2}$. Combining this with (A.2) and Slutsky's theorem gives

$$\sqrt{\frac{Th_T}{R(K)}}\,\widehat{\Omega}_T(\tau)^{-1/2}\bar{g}_T(\tau) \Rightarrow \mathcal{N}(0, I_L).$$

By the continuous mapping theorem,

$$S_T^{\mathrm{loc}}(\tau, \beta_0) = \left\| \sqrt{\frac{Th_T}{R(K)}}\,\widehat{\Omega}_T(\tau)^{-1/2}\bar{g}_T(\tau) \right\|^2 \Rightarrow \chi^2_L.$$

This proves Proposition 1. □

A.2 Proof of Proposition 2

Proof. Write the score at the true local value $\beta(\tau)$ as

$$g_t(\beta(\tau)) = z_t\big(y_t - x_t'\beta(\tau)\big) = z_t\Big(u_t + x_t'\big(\beta(t/T) - \beta(\tau)\big)\Big),$$

using the structural equation $y_t = x_t'\beta(t/T) + u_t$. Hence the kernel average satisfies

$$\bar{g}_T(\tau; \beta(\tau)) = \sum_{t=1}^{T} w_{t,T}(\tau)\, z_t u_t \;+\; \sum_{t=1}^{T} w_{t,T}(\tau)\, z_t x_t'\big(\beta(t/T) - \beta(\tau)\big) \equiv A_T(\tau) + B_T(\tau).$$

Step 1: Drift term is negligible. By twice differentiability of $\beta(\cdot)$, for t/T in the kernel window,

$$\beta(t/T) - \beta(\tau) = \beta'(\tau)(t/T - \tau) + \tfrac{1}{2}\beta''(\tau)(t/T - \tau)^2 + O(h_T^3).$$

Under kernel symmetry and $\int uK(u)\,du = 0$ (Assumption 3(i)), the normalized NW weights satisfy the standard moment property

$$\sum_{t=1}^{T} w_{t,T}(\tau)(t/T - \tau) = o(h_T) \qquad \text{and} \qquad \sum_{t=1}^{T} w_{t,T}(\tau)(t/T - \tau)^2 = O(h_T^2),$$

for fixed interior τ.[8] Under local stationarity, $\mathrm{E}[z_t x_t']$ is Lipschitz in rescaled time (i.e., $\|\mathrm{E}[z_t x_t'] - \mathrm{E}[z_s x_s']\| \le C|t/T - s/T|$ for some $C > 0$), so within the kernel window centered at τ it can be replaced by $\mathrm{E}[z_t(\tau)x_t(\tau)']$ up to $O(h_T)$ uniformly in t within the window. Combined with kernel symmetry ($\sum w_{t,T}(\tau)(t/T - \tau) = o(h_T)$) and the second-order

[8] Under boundary reflection, all $\tau \in (0, 1)$ behave as interior points, so these moment properties hold uniformly. Alternatively, one may replace NW weights with local-linear weights to achieve an exact zero first-moment condition; the argument below is unchanged.

expansion of $\beta(\cdot)$,

$$\mathrm{E}\big[B_T(\tau)\big] = \sum_{t=1}^{T} w_{t,T}(\tau)\,\mathrm{E}[z_t x_t']\big(\beta(t/T) - \beta(\tau)\big) = O(h_T^2),$$

and hence

$$\sqrt{Th_T}\,\mathrm{E}\big[B_T(\tau)\big] \;=\; O\big(\sqrt{Th_T}\,h_T^2\big) \;\to\; 0.$$

Moreover, the same local stationarity and moment bounds used for the FCLT imply $\sqrt{Th_T}\big(B_T(\tau) - \mathrm{E}[B_T(\tau)]\big) = o_p(1)$, so altogether

$$\sqrt{Th_T}\,B_T(\tau) = o_p(1).$$

Step 2: Asymptotic law of the local quadratic form. The leading term is $A_T(\tau) = \sum w_{t,T}(\tau) z_t u_t$. By exogeneity and the maintained FCLT/local stationarity assumptions (applied at τ),

$$\sqrt{Th_T}\,A_T(\tau) \Rightarrow \mathcal{N}\big(0,\ \Omega(\tau)/R(K)\big),$$

where $\Omega(\tau)$ is the local long-run covariance from Assumption 4(a). By Step 1, the same limit holds for $\sqrt{Th_T}\,\bar{g}_T(\tau;\beta(\tau))$. Studentization by $\widehat{\Omega}_T(\tau)$ and the normalization $Th_T/R(K)$ then yield

$$S_T^{\mathrm{loc}}(\tau,\beta(\tau)) \;\xrightarrow{d}\; \chi^2_L,$$

exactly as in Proposition 1, with β_0 replaced by $\beta(\tau)$.

Step 3: Coverage by inversion. By definition,

$$\Pr\big(\beta(\tau) \in CS_T(\tau)\big) = \Pr\Big(S_T^{\mathrm{loc}}(\tau,\beta(\tau)) \le \chi^2_{L,1-\alpha}\Big) \;\longrightarrow\; \Pr\big(\chi^2_L \le \chi^2_{L,1-\alpha}\big) = 1-\alpha,$$

which implies the stated lim inf coverage bound. □

A.3 Proof of Theorem 1

A.3.1 Proof of Theorem 1(i): Orthogonal decomposition

Throughout this proof, write $\bar{g}_T(\tau) \equiv \bar{g}_T(\tau;\beta_0)$, $\bar{g} \equiv \bar{g}(\beta_0) = T^{-1}\sum_{t=1}^{T} g_t(\beta_0)$, and $\tilde{g}_T(\tau) \equiv \bar{g}_T(\tau) - \bar{g}$. Recall that

$$\mathcal{S}_T^{\mathrm{gen}}(\beta_0) = \int_0^1 \frac{Th_T}{R(K)}\,\bar{g}_T(\tau)'\,\widehat{\Omega}_T^{-1}\,\bar{g}_T(\tau)\,d\tau, \qquad \mathcal{S}_T^{\mathrm{stab}}(\beta_0) = \int_0^1 \frac{Th_T}{R(K)}\,\tilde{g}_T(\tau)'\,\widehat{\Omega}_T^{-1}\,\tilde{g}_T(\tau)\,d\tau,$$

and $S_T^{\mathrm{full}}(\beta_0) = T\,\bar{g}'\widehat{\Omega}_T^{-1}\bar{g}$.

Lemma 1 (Integrated NW weights under boundary reflection)**.** *Suppose the NW weights $w_{t,T}(\tau)$ are computed on the boundary-reflected sample as described in Section 2.3. Then,*

uniformly in $t = 1, \ldots, T$,

$$\pi_{t,T} \equiv \int_0^1 w_{t,T}(\tau)\, d\tau = \frac{1}{T} + r_{t,T}, \qquad \max_{1\le t\le T} |r_{t,T}| = o\left(\frac{1}{T}\right).$$

Consequently,

$$\int_0^1 \bar{g}_T(\tau)\, d\tau = \sum_{t=1}^T \pi_{t,T}\, g_t(\beta_0) = \bar{g} + \sum_{t=1}^T r_{t,T}\, g_t(\beta_0), \qquad \textit{and} \qquad \int_0^1 \tilde{g}_T(\tau)\, d\tau = \sum_{t=1}^T r_{t,T}\, g_t(\beta_0).$$

Proof of Lemma 1. Under boundary reflection, each NW weight function $\tau \mapsto w_{t,T}(\tau)$ behaves as an interior NW weight centered at t/T over the full range $\tau \in (0,1)$, so that its total mass equals $1/T$ up to a uniformly negligible error. Formally, using the standard equivalent-kernel representation for NW weights on a reflected sample, one obtains $\int_0^1 w_{t,T}(\tau)\, d\tau = T^{-1} + r_{t,T}$ with $\max_t |r_{t,T}| = o(T^{-1})$; see the boundary-reflection arguments in Hall and Wehrly (1991) and their implementation in Chen and Hong (2012). The remaining identities follow by interchanging sum and integral in $\int_0^1 \bar{g}_T(\tau)\, d\tau = \sum_{t=1}^T g_t(\beta_0) \int_0^1 w_{t,T}(\tau)\, d\tau$. □

Proof of Theorem 1(i). Using $\bar{g}_T(\tau) = \tilde{g}_T(\tau) + \bar{g}$, expand the quadratic form:

$$\bar{g}_T(\tau)'\widehat{\Omega}_T^{-1}\bar{g}_T(\tau) = \tilde{g}_T(\tau)'\widehat{\Omega}_T^{-1}\tilde{g}_T(\tau) + 2\,\bar{g}'\widehat{\Omega}_T^{-1}\tilde{g}_T(\tau) + \bar{g}'\widehat{\Omega}_T^{-1}\bar{g}.$$

Integrating over $\tau \in (0,1)$ and multiplying by $Th_T/R(K)$ yields

$$\begin{aligned}\mathcal{S}_T^{\text{gen}}(\beta_0) &= \mathcal{S}_T^{\text{stab}}(\beta_0) + \frac{Th_T}{R(K)} \cdot 2\,\bar{g}'\widehat{\Omega}_T^{-1} \int_0^1 \tilde{g}_T(\tau)\, d\tau + \frac{Th_T}{R(K)} \cdot \bar{g}'\widehat{\Omega}_T^{-1}\bar{g} \\ &= \mathcal{S}_T^{\text{stab}}(\beta_0) + \frac{h_T}{R(K)}\, S_T^{\text{full}}(\beta_0) + \frac{2Th_T}{R(K)}\, \bar{g}'\widehat{\Omega}_T^{-1} \int_0^1 \tilde{g}_T(\tau)\, d\tau.\end{aligned}$$

It remains to show that the cross term is $o_p(1)$. By Lemma 1,

$$\int_0^1 \tilde{g}_T(\tau)\, d\tau = \sum_{t=1}^T r_{t,T}\, g_t(\beta_0).$$

Under H_0^{stab}, $\bar{g} = O_p(T^{-1/2})$ by the FCLT in Assumption 2, and $\widehat{\Omega}_T^{-1} = O_p(1)$ by Assumption 4(b). Moreover, $\max_t |r_{t,T}| = o(T^{-1})$ implies $\sum_{t=1}^T |r_{t,T}| = o(1)$, so the weighted sum $\sum_{t=1}^T r_{t,T} g_t(\beta_0) = o_p(T^{-1/2})$ under standard moment bounds implied by the FCLT. Hence,

$$\bar{g}'\widehat{\Omega}_T^{-1} \int_0^1 \tilde{g}_T(\tau)\, d\tau = O_p(T^{-1/2}) \cdot O_p(1) \cdot o_p(T^{-1/2}) = o_p(T^{-1}),$$

and therefore

$$\frac{2Th_T}{R(K)}\, \bar{g}'\widehat{\Omega}_T^{-1} \int_0^1 \tilde{g}_T(\tau)\, d\tau = h_T \cdot o_p(1) = o_p(1),$$

since h_T is bounded (under either Assumption 3(shrinking) or Assumption 3(fixed)). This establishes

$$\mathcal{S}_T^{\text{gen}}(\beta_0) = \frac{h_T}{R(K)}\, S_T^{\text{full}}(\beta_0) + \mathcal{S}_T^{\text{stab}}(\beta_0) + o_p(1),$$

which is the desired decomposition. □

A.3.2 Proof of Theorem 1(ii): Asymptotic independence

We first state the following lemma.

Lemma 2 (Brownian bridge orthogonality and independence)**.** *Let $u_t \equiv \Omega^{-1/2} g_t(\beta_0)$ and define*

$$U_T(s) \equiv \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor Ts \rfloor} u_t, \qquad s \in [0,1].$$

Under Assumption 2 and H_0^{stab}, $U_T(\cdot) \Rightarrow W(\cdot)$ in $D[0,1]$, where $W(\cdot)$ is L-dimensional standard Brownian motion. Let $B(s) \equiv W(s) - sW(1)$ denote the associated Brownian bridge. Then $W(1)$ and $B(\cdot)$ are independent.

Moreover, let $\{\kappa_{t,T}(\tau)\}_{t=1}^T$ be weights such that $\sum_{t=1}^T \kappa_{t,T}(\tau) = 0$ for each $\tau \in (0,1)$. If a process of the form $\{\sum_{t=1}^T \kappa_{t,T}(\tau) u_t : \tau \in (0,1)\}$ converges weakly to a (mean-zero) Gaussian process that is a continuous linear functional of $W(\cdot)$, then the weak limit is measurable with respect to $B(\cdot)$ and is therefore independent of $W(1)$.

Proof of Lemma 2. Since $(W(1), B(\cdot))$ is jointly Gaussian, it suffices to show zero covariance. For any $s \in [0,1]$,

$$\text{Cov}\big(B(s), W(1)\big) = \text{Cov}\big(W(s) - sW(1), W(1)\big) = sI_L - sI_L = 0,$$

so $B(\cdot) \perp W(1)$. For the second statement, write the orthogonal decomposition $W(s) = B(s) + sW(1)$ and note that the constraint $\sum_{t=1}^T \kappa_{t,T}(\tau) = 0$ removes the $W(1)$ component in the limit, so the limiting Gaussian process depends on $W(\cdot)$ only through $B(\cdot)$ and is independent of $W(1)$. □

Proceeding with Theorem 1(ii), by Assumption 4(b), replacing $\widehat{\Omega}_T$ with Ω does not affect joint limits (Slutsky). Let $u_t \equiv \Omega^{-1/2} g_t(\beta_0)$ and define $U_T(\cdot)$ as in Lemma 2. Then

$$\sqrt{T}\, \Omega^{-1/2} \bar{g} = \frac{1}{\sqrt{T}} \sum_{t=1}^T u_t = U_T(1) \Rightarrow W(1).$$

Next, for each $\tau \in (0,1)$,

$$\Omega^{-1/2} \tilde{g}_T(\tau) = \Omega^{-1/2}\big(\bar{g}_T(\tau) - \bar{g}\big) = \sum_{t=1}^T \Big(w_{t,T}(\tau) - \frac{1}{T}\Big) u_t \equiv \sum_{t=1}^T \kappa_{t,T}(\tau) u_t,$$

where $\kappa_{t,T}(\tau) \equiv w_{t,T}(\tau) - 1/T$. Since NW weights sum to one, $\sum_{t=1}^{T} w_{t,T}(\tau) = 1$, we have $\sum_{t=1}^{T} \kappa_{t,T}(\tau) = 0$ for each τ. Therefore, by Lemma 2, the weak limit of the centered local-mean process is measurable with respect to the Brownian bridge $B(\cdot)$ and is independent of $W(1)$.

Finally, $S_T^{\text{full}}(\beta_0) = \|\sqrt{T}\,\widehat{\Omega}_T^{-1/2}\bar{g}\|^2$ is a measurable functional of $\sqrt{T}\bar{g}$, and $\mathcal{S}_T^{\text{stab}}(\beta_0)$ is a measurable functional of the centered local-mean process $\{\tilde{g}_T(\tau) : \tau \in (0,1)\}$. Independence of the Gaussian limits and the continuous mapping theorem imply that the full-sample component and the stability component are asymptotically independent. □

A.3.3 Proof of Theorem 1(iii)(a): Full-sample component

Recall $S_T^{\text{full}}(\beta_0) = T\,\bar{g}(\beta_0)'\widehat{\Omega}_T^{-1}\bar{g}(\beta_0)$ with $\bar{g}(\beta_0) = T^{-1}\sum_{t=1}^{T} g_t(\beta_0)$. Note that

$$\sqrt{T}\,\bar{g}(\beta_0) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} g_t(\beta_0) = S_T(1),$$

where $S_T(\cdot)$ is the partial-sum process in Assumption 2. Under H_0^{stab} and $m(\cdot) \equiv 0$, Assumption 2 implies $S_T(1) \Rightarrow \Omega^{1/2}W(1) \sim \mathcal{N}(0,\Omega)$. Hence

$$T\,\bar{g}(\beta_0)'\Omega^{-1}\bar{g}(\beta_0) = \left\|\Omega^{-1/2}\sqrt{T}\,\bar{g}(\beta_0)\right\|^2 \Rightarrow \chi_L^2,$$

and in particular $T\,\bar{g}(\beta_0)'\Omega^{-1}\bar{g}(\beta_0) = O_p(1)$. By Assumption 4(b), $\widehat{\Omega}_T \xrightarrow{p} \Omega$, so Slutsky's theorem yields

$$S_T^{\text{full}}(\beta_0) = T\,\bar{g}(\beta_0)'\widehat{\Omega}_T^{-1}\bar{g}(\beta_0) \Rightarrow \chi_L^2, \qquad \text{and hence} \qquad S_T^{\text{full}}(\beta_0) = O_p(1).$$

Finally, under Assumption 3(shrinking) we have $h_T \to 0$, so

$$\frac{h_T}{R(K)}\,S_T^{\text{full}}(\beta_0) = o_p(1),$$

since $R(K) \in (0,\infty)$ is constant. □

A.3.4 Proof of Theorem 1(iii)(b): Stability component (fixed bandwidth)

Assume Assumption 3(fixed), so $h_T \to h \in (0,\infty)$. By Assumption 4(b), $\widehat{\Omega}_T \xrightarrow{p} \Omega$, hence replacing $\widehat{\Omega}_T^{-1}$ by Ω^{-1} does not affect the limit (Slutsky). Define the whitened, rescaled local-mean process

$$G_T(\tau) \;\equiv\; \sqrt{\frac{Th_T}{R(K)}}\,\Omega^{-1/2}\,\bar{g}_T(\tau;\beta_0), \qquad \tau \in (0,1),$$

computed on the boundary-reflected sample so that the normalization constant is $R(K)$ uniformly over $\tau \in (0,1)$.

Step 1: Gaussian-process limit for $G_T(\cdot)$. Under H_0^{stab}, $E[g_t(\beta_0)] = 0$ and Assumption 2 yields an FCLT for the partial sums of $g_t(\beta_0)$. Since, for fixed h, NW smoothing on the reflected sample is a continuous linear operator on the partial-sum path, a continuous mapping argument implies that $G_T(\cdot) \Rightarrow G(\cdot)$ in $C[0,1]$, where $G(\cdot)$ is an L-dimensional mean-zero Gaussian process. The (componentwise) covariance kernel is given by the overlap of the NW kernel windows:

$$\text{Cov}\big(G(\tau_1), G(\tau_2)\big) = \rho\left(\frac{\tau_2 - \tau_1}{h}\right) I_L, \qquad \tau_1, \tau_2 \in (0,1), \tag{A.3}$$

where

$$\rho(u) \equiv \frac{1}{R(K)} \int_{-1}^{1} K(v)\, K(v+u)\, dv, \qquad R(K) = \int_{-1}^{1} K(v)^2\, dv,$$

with the convention that $K(\cdot) = 0$ outside $[-1,1]$. This matches (22) in the main text.

Step 2: Centering and Gaussian bridge process. Define

$$\tilde{G}_T(\tau) \equiv G_T(\tau) - \int_0^1 G_T(s)\, ds, \qquad \tilde{G}(\tau) \equiv G(\tau) - \int_0^1 G(s)\, ds.$$

Since the map $x(\cdot) \mapsto x(\cdot) - \int_0^1 x(s)\, ds$ is continuous on $C[0,1]$, we have $\tilde{G}_T(\cdot) \Rightarrow \tilde{G}(\cdot)$ in $C[0,1]$.

Step 3: Rewrite $\mathcal{S}_T^{\text{stab}}(\beta_0)$ as an integral of $\|\tilde{G}_T(\tau)\|^2$. By definition,

$$\mathcal{S}_T^{\text{stab}}(\beta_0) = \int_0^1 \frac{Th_T}{R(K)}\, \tilde{g}_T(\tau;\beta_0)'\, \widehat{\Omega}_T^{-1}\, \tilde{g}_T(\tau;\beta_0)\, d\tau.$$

Replacing $\widehat{\Omega}_T^{-1}$ with Ω^{-1} incurs $o_p(1)$ uniformly in τ by Slutsky, and by construction

$$\frac{Th_T}{R(K)}\, \tilde{g}_T(\tau;\beta_0)'\Omega^{-1}\tilde{g}_T(\tau;\beta_0) = \Big\|\sqrt{\tfrac{Th_T}{R(K)}}\, \Omega^{-1/2}\tilde{g}_T(\tau;\beta_0)\Big\|^2 = \|\tilde{G}_T(\tau)\|^2 + o_p(1),$$

where the last step uses the fact that $\tilde{G}_T(\tau)$ is centered by $\int_0^1 G_T(s)\, ds$ while $\tilde{g}_T(\tau)$ is centered by $\bar{g}$; by Lemma 1, these two centering operations differ by $o_p((Th_T)^{-1/2})$. Hence

$$\mathcal{S}_T^{\text{stab}}(\beta_0) = \int_0^1 \|\tilde{G}_T(\tau)\|^2\, d\tau + o_p(1).$$

Step 4: Continuous mapping. The functional $x(\cdot) \mapsto \int_0^1 \|x(\tau)\|^2 d\tau$ is continuous on $C[0,1]$. Therefore,

$$\int_0^1 \|\tilde{G}_T(\tau)\|^2\, d\tau \Rightarrow \int_0^1 \|\tilde{G}(\tau)\|^2\, d\tau,$$

which implies

$$\mathcal{S}_T^{\text{stab}}(\beta_0) \Rightarrow \int_0^1 \|\tilde{G}(\tau)\|^2 \, d\tau.$$

This proves Theorem 1(iii)(b). □

A.3.5 Proof of Theorem 1(iii)(c): Stability component (shrinking bandwidth)

This proof follows the structure of Chen and Hong (2012, Appendix).

Let $g_t \equiv g_t(\beta_0) \in \mathrm{R}^L$ be the L-vector of moment functions under the global stability null $H_0^{\text{stab}} : \beta(\tau) \equiv \beta_0$ for all $\tau \in (0,1)$. Let $\widehat{\Omega}_T \Omega$, with Ω symmetric positive definite (Assumption 4(b)). Define the NW local average

$$\bar{g}_T(\tau) \equiv \bar{g}_T(\tau; \beta_0) = \sum_{t=1}^{T} w_{t,T}(\tau) \, g_t, \qquad \tau \in (0,1), \tag{A.4}$$

and its centered version $\tilde{g}_T(\tau) = \bar{g}_T(\tau) - \bar{g}$, where $\bar{g} = T^{-1} \sum_{t=1}^{T} g_t$.

For shrinking bandwidth $h_T \to 0$ and $Th_T \to \infty$ (Assumption 3(shrinking)), consider

$$\mathcal{S}_T^{\text{stab}}(\beta_0) = \int_0^1 \frac{Th_T}{R(K)} \, \tilde{g}_T(\tau)' \, \widehat{\Omega}_T^{-1} \, \tilde{g}_T(\tau) \, d\tau. \tag{A.5}$$

Claim. Under Assumptions 1–4(b),

$$\frac{\mathcal{S}_T^{\text{stab}}(\beta_0) - L}{\sqrt{h_T \cdot 4L \, C_B \, R(K)^{-2}}} \;\; N(0,1), \qquad C_B = \int_0^2 \left[\int_{-1}^{1} K(u) K(u+v) \, du \right]^2 dv. \tag{A.6}$$

This is Theorem 1(iii)(c).

Preliminaries

Whitening. Write the whitened score $\xi_t = \Omega^{-1/2} g_t \in \mathrm{R}^L$. Then under H_0^{stab}, the FCLT (Assumption 2) implies

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor T\tau \rfloor} \xi_t \Rightarrow W(\tau) \qquad \text{in } D[0,1], \tag{A.7}$$

where $W(\cdot)$ is L-dimensional standard Brownian motion. By Slutsky, it suffices to work with Ω known and replace $\widehat{\Omega}_T^{-1}$ by Ω^{-1} throughout; all remainder terms introduced by this substitution are $o_p(1)$ because $\int_0^1 \tilde{g}_T(\tau)^2 \, d\tau = O_p((Th_T)^{-1})$ (which follows from $\mathcal{S}_T^{\text{stab}}(\beta_0) = O_p(1)$) and $\widehat{\Omega}_T^{-1} - \Omega^{-1} = o_p(1)$.

Accordingly define

$$\bar{\xi}_T(\tau) = \sum_{t=1}^{T} w_{t,T}(\tau)\xi_t, \qquad \tilde{\xi}_T(\tau) = \bar{\xi}_T(\tau) - \bar{\xi}, \qquad \bar{\xi} = T^{-1}\sum_{t=1}^{T}\xi_t. \tag{A.8}$$

Then $\mathcal{S}_T^{\text{stab}}(\beta_0)$ equals

$$\mathcal{S}_T^{\text{stab}}(\beta_0) = \int_0^1 \frac{Th_T}{R(K)}\,\tilde{\xi}_T(\tau)^2\,d\tau + o_p(1). \tag{A.9}$$

Riemann approximation. Let $\tau_t = t/T$. Under the kernel's compact support, $\tilde{\xi}_T(\tau)$ is piecewise constant between grid points up to $o_p((Th_T)^{-1/2})$, so

$$\int_0^1 \tilde{\xi}_T(\tau)^2\,d\tau = \frac{1}{T}\sum_{t=1}^{T}\tilde{\xi}_T(\tau_t)^2 + o_p\!\left(\frac{1}{Th_T}\right). \tag{A.10}$$

Plugging (A.10) into (A.9) yields the discrete representation

$$\mathcal{S}_T^{\text{stab}}(\beta_0) = \frac{h_T}{R(K)}\sum_{t=1}^{T}\tilde{\xi}_T(\tau_t)^2 + o_p(1). \tag{A.11}$$

Quadratic-form decomposition

Rewriting as a quadratic form in the sample scores. Write the centered NW weights as

$$\tilde{w}_{s,T}(\tau) \equiv w_{s,T}(\tau) - \frac{1}{T}, \qquad \text{so that} \qquad \tilde{\xi}_T(\tau) = \sum_{s=1}^{T}\tilde{w}_{s,T}(\tau)\,\xi_s. \tag{A.12}$$

Then

$$\begin{aligned}\mathcal{S}_T^{\text{stab}}(\beta_0) &= \frac{h_T}{R(K)}\sum_{t=1}^{T}\left(\sum_{r=1}^{T}\tilde{w}_{r,T}(\tau_t)\xi_r\right)'\left(\sum_{s=1}^{T}\tilde{w}_{s,T}(\tau_t)\xi_s\right) + o_p(1)\\ &= \frac{h_T}{R(K)}\sum_{r=1}^{T}\sum_{s=1}^{T}\left[\sum_{t=1}^{T}\tilde{w}_{r,T}(\tau_t)\tilde{w}_{s,T}(\tau_t)\right]\xi_r'\xi_s + o_p(1).\end{aligned} \tag{A.13}$$

Define the weight matrix

$$A_{rs} \equiv A_{rs,T} \equiv h_T\sum_{t=1}^{T}\tilde{w}_{r,T}(\tau_t)\tilde{w}_{s,T}(\tau_t). \tag{A.14}$$

Then (A.13) is

$$\mathcal{S}_T^{\text{stab}}(\beta_0) = \frac{1}{R(K)} \sum_{r=1}^{T} \sum_{s=1}^{T} A_{rs}\, \xi_r' \xi_s + o_p(1). \tag{A.15}$$

Deterministic approximation to A_{rs}. For NW weights with boundary reflection, the weights are asymptotically equivalent to interior NW weights based on K. In particular, uniformly in r, s with $|r - s| \leq Th_T$,

$$A_{rs} = \frac{1}{T} \int_{-1}^{1} K(u) K\left(u + \frac{r-s}{Th_T}\right) du + o\left(\frac{1}{T}\right), \tag{A.16}$$

and $A_{rs} = o(T^{-1})$ if $|r - s| > Th_T$. Define the deterministic kernel overlap weights

$$w_{rs} \equiv w_{rs,T} \equiv \int_{-1}^{1} K(u) K\left(u + \frac{r-s}{Th_T}\right) du, \tag{A.17}$$

so that $A_{rs} = T^{-1} w_{rs} + o(T^{-1})$.

Substituting into (A.15) yields

$$\mathcal{S}_T^{\text{stab}}(\beta_0) = \frac{1}{TR(K)} \sum_{r=1}^{T} \sum_{s=1}^{T} w_{rs}\, \xi_r' \xi_s + o_p(1). \tag{A.18}$$

Diagonal vs. off-diagonal decomposition. Write

$$\frac{1}{T} \sum_{r=1}^{T} \sum_{s=1}^{T} w_{rs}\, \xi_r' \xi_s = \underbrace{\frac{1}{T} \sum_{s=1}^{T} w_{ss}\, \xi_s' \xi_s}_{\mathcal{D}_T} + \underbrace{\frac{2}{T} \sum_{s=2}^{T} \sum_{r=1}^{s-1} w_{rs}\, \xi_r' \xi_s}_{2\mathcal{U}_T}. \tag{A.19}$$

Note that $w_{ss} = \int_{-1}^{1} K(u)^2\, du = R(K)$ by definition. Hence,

$$\mathcal{D}_T = R(K) \cdot \frac{1}{T} \sum_{s=1}^{T} \xi_s' \xi_s. \tag{A.20}$$

Under H_0^{stab}, $\mathrm{E}[\xi_s \xi_s'] = I_L$ and a law of large numbers (implied by the same dependence restrictions that yield the FCLT) gives

$$\frac{1}{T} \sum_{s=1}^{T} \xi_s' \xi_s \mathrm{E}[\xi_1' \xi_1] = L, \qquad \text{so} \qquad \mathcal{D}_T = R(K)\, L + o_p(1). \tag{A.21}$$

Combining (A.18), (A.19), and (A.21) yields the key reduction

$$\mathcal{S}_T^{\text{stab}}(\beta_0) = L + \frac{2}{\sqrt{h_T}} \underbrace{\left(\frac{h_T^{1/2}}{TR(K)} \sum_{s=2}^{T} \sum_{r=1}^{s-1} w_{rs}\, \xi_r' \xi_s \right)}_{\widetilde{U}_T} + o_p(1). \tag{A.22}$$

Thus, it suffices to prove a CLT for the scalar term $\widetilde{U}_T$.

Martingale CLT for the off-diagonal term

Martingale array representation. Let $\mathcal{F}_s = \sigma(\xi_1, \ldots, \xi_s)$ and define

$$R_s = \frac{1}{T} h_T^{-1/2} \sum_{r=1}^{s-1} w_{rs}\, \xi_r' \xi_s, \qquad s = 2, \ldots, T. \tag{A.23}$$

Then $\widetilde{U}_T = \frac{1}{R(K)} \sum_{s=2}^{T} R_s$ and $\{R_s, \mathcal{F}_s\}$ is a martingale difference array because $\mathrm{E}[\xi_s \mid \mathcal{F}_{s-1}] = 0$ under H_0^{stab} and w_{rs} is $\mathcal{F}_{s-1}$-measurable. Therefore, we may apply a martingale CLT (Brown, 1971) in the same way as in Chen and Hong (2012): it suffices to verify a Lindeberg condition and conditional variance convergence.

Variance computation and the C_B constant. First compute the unconditional variance. Using $\mathrm{E}[\xi_s \xi_s'] = I_L$ and (weak) dependence/mixing to control higher-order terms, we obtain

$$\begin{aligned} \mathrm{Var}\left(\sum_{s=2}^{T} R_s \right) &= \sum_{s=2}^{T} \mathrm{E}[R_s^2] + o(1) \\ &= T^{-2} h_T^{-1} \sum_{s=2}^{T} \sum_{r=1}^{s-1} w_{rs}^2\, \mathrm{E}\big[(\xi_r' \xi_s)^2\big] + o(1) \\ &= T^{-2} h_T^{-1} \sum_{s=2}^{T} \sum_{r=1}^{s-1} w_{rs}^2\, \mathrm{tr}(I_L) + o(1) \\ &= L\, T^{-2} h_T^{-1} \sum_{s=2}^{T} \sum_{r=1}^{s-1} w_{rs}^2 + o(1). \end{aligned} \tag{A.24}$$

Because $w_{rs} = 0$ when $|r - s| > T h_T$ (compact support), the double sum is effectively over lags $j = s - r$ with $1 \le j \le T h_T$:

$$T^{-2} h_T^{-1} \sum_{s=2}^{T} \sum_{r=1}^{s-1} w_{rs}^2 = T^{-1} h_T^{-1} \sum_{j=1}^{T-1} \left(1 - \frac{j}{T} \right) \left[\int_{-1}^{1} K(u) K\left(u + \frac{j}{T h_T} \right) du \right]^2. \tag{A.25}$$

A Riemann approximation (change of variables $v = j/(Th_T)$ and using $Th_T \to \infty$) yields

$$T^{-2}h_T^{-1}\sum_{s=2}^{T}\sum_{r=1}^{s-1} w_{rs}^2 \to C_B, \qquad C_B = \int_0^2 \left[\int_{-1}^{1} K(u)K(u+v)\,du\right]^2 dv. \tag{A.26}$$

Combining (A.24)–(A.26) gives

$$\mathrm{Var}\left(\sum_{s=2}^{T} R_s\right) \to L\,C_B. \tag{A.27}$$

Therefore,

$$\mathrm{Var}(\widetilde{U}_T) = \frac{1}{R(K)^2}\mathrm{Var}\left(\sum_{s=2}^{T} R_s\right) \to \frac{L\,C_B}{R(K)^2}. \tag{A.28}$$

Lindeberg and conditional variance conditions. To apply Brown's martingale CLT, we verify:

(i) Lindeberg condition. For any $\eta > 0$,

$$\frac{1}{\mathrm{Var}(\sum_{s=2}^{T} R_s)}\sum_{s=2}^{T}\mathrm{E}\left[R_s^2\mathbf{1}\left\{|R_s| > \eta\sqrt{\mathrm{Var}(\sum_{s=2}^{T} R_s)}\right\}\right] \to 0. \tag{A.29}$$

Under the same moment and weak-dependence conditions used by Chen and Hong (2012) (bounded $2+\delta$ moments and strong mixing), one has $\sum_{s=2}^{T}\mathrm{E}[R_s^4] = o(1)$ because R_s is a sum of $O(Th_T)$ terms, each of order $T^{-1}h_T^{-1/2}$, and the kernel overlap weights satisfy $\max_{r,s}|w_{rs}| \le R(K)$ and $\sum_{r<s} w_{rs}^2 = O(Th_T)$. Then (A.29) follows by Cauchy–Schwarz and Markov.

(ii) Conditional variance convergence. We require

$$\frac{1}{\mathrm{Var}(\sum_{s=2}^{T} R_s)}\sum_{s=2}^{T}\mathrm{E}\big[R_s^2 \mid \mathcal{F}_{s-1}\big]\,1. \tag{A.30}$$

Write

$$\mathrm{E}[R_s^2 \mid \mathcal{F}_{s-1}] = T^{-2}h_T^{-1}\sum_{r=1}^{s-1}\sum_{q=1}^{s-1} w_{rs}w_{qs}\,\xi_r'\mathrm{E}[\xi_s\xi_s' \mid \mathcal{F}_{s-1}]\,\xi_q. \tag{A.31}$$

Under H_0^{stab}, $\mathrm{E}[\xi_s\xi_s' \mid \mathcal{F}_{s-1}] = I_L + \Delta_s$ where Δ_s is a mean-zero term. Using mixing to bound the contribution of Δ_s and the off-diagonal $r \neq q$ terms, one shows (as in Chen and Hong (2012), their verification of Brown's condition (A6)) that

$$\sum_{s=2}^{T}\big(\mathrm{E}[R_s^2 \mid \mathcal{F}_{s-1}] - \mathrm{E}[R_s^2]\big) = o_p(1), \tag{A.32}$$

which implies (A.30) given (A.27).

CLT for $\widetilde{U}_T$. By Brown's martingale CLT, (A.29) and (A.30) imply

$$\frac{\sum_{s=2}^{T} R_s}{\sqrt{\mathrm{Var}(\sum_{s=2}^{T} R_s)}} N(0,1). \tag{A.33}$$

Using (A.27)–(A.28) and Slutsky,

$$\frac{\widetilde{U}_T}{\sqrt{LC_B}\, R(K)^{-1}} N(0,1). \tag{A.34}$$

Finishing the proof. From (A.22),

$$\mathcal{S}_T^{\mathrm{stab}}(\beta_0) - L = 2\widetilde{U}_T + o_p(\sqrt{h_T}). \tag{A.35}$$

Multiplying and dividing by the correct scaling and using (A.34) gives

$$\frac{\mathcal{S}_T^{\mathrm{stab}}(\beta_0) - L}{\sqrt{h_T \cdot 4LC_B R(K)^{-2}}} = \frac{2\widetilde{U}_T}{2\sqrt{LC_B} R(K)^{-1}} + o_p(1)$$
$$N(0,1), \tag{A.36}$$

which is exactly (A.6). □

A.3.6 Proof of Proposition 3: Weak-identification robustness

This proof exploits a Caner-type orthogonalization idea: centering by the full-sample mean removes τ-invariant components, analogous to projecting partial-sample objects off the full-sample span in Caner (2011), which yields Brownian-bridge limits under the null.

For the linear IV/GMM score $g_t(\beta) = z_t(y_t - x_t'\beta)$, define the Jacobian process

$$J_t \equiv z_t x_t' \in \mathbb{R}^{L\times K}, \qquad \bar{J} \equiv \frac{1}{T}\sum_{t=1}^{T} J_t, \qquad \bar{J}_T(\tau) \equiv \sum_{t=1}^{T} w_{t,T}(\tau)\, J_t.$$

We assume:

Assumption A.1 (Integrated Jacobian stability)**.** *Under H_0^{stab},*

$$\int_0^1 \left\| \bar{J}_T(\tau) - \bar{J} \right\|_F^2 d\tau \;=\; o_p\!\left(\frac{1}{T}\right), \tag{A.37}$$

where $\|\cdot\|_F$ denotes the Frobenius norm.

Scope of Assumption A.1. Assumption A.1 is satisfied when the Jacobian has slowly varying conditional mean. In the SP-IV implementation, the Jacobian equals the matrix of impulse response functions, which is estimated from VAR coefficients and inherits their smoothness. Under fixed bandwidth, the condition holds when Jacobian variation is negligible relative to score variation.

Recall $\tilde{g}_T(\tau;\beta) \equiv \bar{g}_T(\tau;\beta) - \bar{g}(\beta)$. Let $\Delta_T \equiv \hat{\beta} - \beta_0$.

Lemma 3 (Centering removes τ-invariant components)**.** *For every $\tau \in (0,1)$,*

$$\tilde{g}_T(\tau;\hat{\beta}) = \tilde{g}_T(\tau;\beta_0) - \big(\bar{J}_T(\tau) - \bar{J}\big)\,\Delta_T. \tag{A.38}$$

Proof of Lemma 3. Using $g_t(\hat{\beta}) = g_t(\beta_0) - J_t(\hat{\beta} - \beta_0)$, NW smoothing gives $\bar{g}_T(\tau;\hat{\beta}) = \bar{g}_T(\tau;\beta_0) - \bar{J}_T(\tau)\Delta_T$. Likewise $\bar{g}(\hat{\beta}) = \bar{g}(\beta_0) - \bar{J}\,\Delta_T$. Subtracting yields (A.38). □

Step 1: Control the difference in stability statistics. Write, for brevity,

$$a_T(\tau) \equiv \tilde{g}_T(\tau;\beta_0), \qquad b_T(\tau) \equiv \tilde{g}_T(\tau;\hat{\beta}) - \tilde{g}_T(\tau;\beta_0) = -\big(\bar{J}_T(\tau) - \bar{J}\big)\Delta_T,$$

where the last equality is Lemma 3. By Slutsky, we may replace $\widehat{\Omega}_T^{-1}$ with Ω^{-1} throughout this proof (Assumption 4(b)).

Using $\|x+y\|_{\Omega^{-1}}^2 - \|x\|_{\Omega^{-1}}^2 = 2x'\Omega^{-1}y + y'\Omega^{-1}y$ and Cauchy–Schwarz,

$$\begin{aligned}\big|\mathcal{S}_T^{\mathrm{stab}}(\hat{\beta}) - \mathcal{S}_T^{\mathrm{stab}}(\beta_0)\big| &\le \frac{Th_T}{R(K)}\int_0^1 \Big(2\,\|a_T(\tau)\|_{\Omega^{-1}}\,\|b_T(\tau)\|_{\Omega^{-1}} + \|b_T(\tau)\|_{\Omega^{-1}}^2\Big)\,d\tau \\ &\le \frac{Th_T}{R(K)}\left(2\Big[\int_0^1 \|a_T(\tau)\|_{\Omega^{-1}}^2\,d\tau\Big]^{1/2}\Big[\int_0^1 \|b_T(\tau)\|_{\Omega^{-1}}^2\,d\tau\Big]^{1/2} + \int_0^1 \|b_T(\tau)\|_{\Omega^{-1}}^2\,d\tau\right).\end{aligned} \tag{A.39}$$

We now bound the two L^2-terms.

Bound for a_T. By definition,

$$\int_0^1 \|a_T(\tau)\|_{\Omega^{-1}}^2\,d\tau = \frac{R(K)}{Th_T}\,\mathcal{S}_T^{\mathrm{stab}}(\beta_0) + o_p\!\left(\frac{1}{Th_T}\right),$$

so under either Assumption 3(fixed) or Assumption 3(shrinking), $\mathcal{S}_T^{\mathrm{stab}}(\beta_0) = O_p(1)$ (Theorem 1(iii)(b) and (iii)(c) both imply $O_p(1)$), hence

$$\int_0^1 \|a_T(\tau)\|_{\Omega^{-1}}^2\,d\tau = O_p\!\left(\frac{1}{Th_T}\right). \tag{A.40}$$

Bound for b_T. Since $\Delta_T = O_p(1)$ by assumption and $\|\Omega^{-1/2}A\|_F^2 \le \|\Omega^{-1}\|\,\|A\|_F^2$, As-

sumption A.1 implies

$$\int_0^1 \|b_T(\tau)\|^2_{\Omega^{-1}}\,d\tau = \int_0^1 \Delta_T'\big(\bar{J}_T(\tau)-\bar{J}\big)'\Omega^{-1}\big(\bar{J}_T(\tau)-\bar{J}\big)\Delta_T\,d\tau$$
$$\leq \|\Delta_T\|^2\,\|\Omega^{-1}\|\int_0^1 \|\bar{J}_T(\tau)-\bar{J}\|_F^2\,d\tau = o_p\!\left(\frac{1}{T}\right). \tag{A.41}$$

Combine. Plugging (A.40)–(A.41) into (A.39) yields

$$\mathcal{S}_T^{\text{stab}}(\hat{\beta}) - \mathcal{S}_T^{\text{stab}}(\beta_0) = o_p(1) \qquad \text{under Assumption 3(fixed),} \tag{A.42}$$

and, moreover, under Assumption 3(shrinking) (where $h_T \to 0$),

$$\mathcal{S}_T^{\text{stab}}(\hat{\beta}) - \mathcal{S}_T^{\text{stab}}(\beta_0) = o_p(\sqrt{h_T}). \tag{A.43}$$

Indeed, the leading term in (A.39) is of order $(Th_T)\cdot O_p((Th_T)^{-1/2})\cdot o_p(T^{-1/2}) = o_p(\sqrt{h_T})$, and the b_T^2 term is $(Th_T)\cdot o_p(T^{-1}) = o_p(h_T) = o_p(\sqrt{h_T})$.

Step 2: Conclude part (i) (fixed bandwidth). Under Assumption 3(fixed), Theorem 1(iii)(b) gives $\mathcal{S}_T^{\text{stab}}(\beta_0) \Rightarrow \int_0^1 \|\tilde{G}(\tau)\|^2\,d\tau$. By (A.42) and Slutsky,

$$\mathcal{S}_T^{\text{stab}}(\hat{\beta}) \Rightarrow \int_0^1 \|\tilde{G}(\tau)\|^2\,d\tau.$$

Step 3: Conclude part (ii) (shrinking bandwidth). Under Assumption 3(shrinking), Theorem 1(iii)(c) yields

$$\frac{\mathcal{S}_T^{\text{stab}}(\beta_0) - L}{\sqrt{h_T\cdot 4L\,C_B\,R(K)^{-2}}} \Rightarrow N(0,1),$$

(where the additional quadratic-form regularity conditions required for (iii)(c) are understood). By (A.43),

$$\frac{\mathcal{S}_T^{\text{stab}}(\hat{\beta}) - \mathcal{S}_T^{\text{stab}}(\beta_0)}{\sqrt{h_T}} \xrightarrow{p} 0,$$

and Slutsky implies

$$\frac{\mathcal{S}_T^{\text{stab}}(\hat{\beta}) - L}{\sqrt{h_T\cdot 4L\,C_B\,R(K)^{-2}}} \Rightarrow N(0,1).$$

This proves Proposition 3. □

B Proofs for Section 3 (K-statistic)

Throughout this subsection, we maintain the notation of Section 3 and Appendix A. All NW weights are computed on the boundary-reflected sample as described in the preamble to Appendix A.

B.0.1 Proof of Proposition 4: Local K-statistic

Fix $\tau \in (0,1)$. The proof has two steps: (i) establish a joint CLT for the local score and orthogonalized Jacobian, and (ii) show that the projected quadratic form is χ^2_K.

Step 1: Joint local CLT. By the joint FCLT (Assumption 6) and the same kernel-smoothing CLT argument used in the proof of Proposition 1 (Appendix A.1, Step 1), applied to the $(L + LK)$-dimensional stacked process $\phi_t(\beta_0)$, we obtain

$$\sqrt{\frac{Th_T}{R(K)}} \begin{pmatrix} \bar{g}_T(\tau;\beta_0) \\ \mathrm{vec}\big(\bar{H}_T(\tau;\beta_0) - J(\beta_0)\big) \end{pmatrix} \Rightarrow \mathcal{N}(0,\, V(\tau)), \tag{B.1}$$

where $V(\tau)$ is the local long-run covariance from Assumption 7(a), with partition

$$V(\tau) = \begin{pmatrix} \Omega(\tau) & V_{gh}(\tau) \\ V_{hg}(\tau) & V_{hh}(\tau) \end{pmatrix}.$$

Under $H_0^{\mathrm{loc}}(\tau)$, the mean of $\bar{g}_T(\tau;\beta_0)$ is zero by the same centering argument as in Appendix A.1, Step 2.

Step 2: Orthogonalization and projection. The orthogonalized local Jacobian is

$$\mathrm{vec}(\bar{D}_T(\tau)) = \mathrm{vec}(\bar{H}_T(\tau)) - \widehat{V}_{hg,T}(\tau)\,\widehat{\Omega}_T(\tau)^{-1}\,\bar{g}_T(\tau).$$

By (B.1) and $\widehat{V}_{hg,T}(\tau) \xrightarrow{p} V_{hg}(\tau)$, $\widehat{\Omega}_T(\tau) \xrightarrow{p} \Omega(\tau)$ (Assumption 7(a)), Slutsky's theorem gives

$$\sqrt{\frac{Th_T}{R(K)}}\,\mathrm{vec}\big(\bar{D}_T(\tau) - J(\beta_0)\big) \Rightarrow \mathcal{N}\big(0,\, V_{hh}(\tau) - V_{hg}(\tau)\,\Omega(\tau)^{-1}\,V_{gh}(\tau)\big). \tag{B.2}$$

The key property of the orthogonalization is that, jointly,

$$\sqrt{\frac{Th_T}{R(K)}} \begin{pmatrix} \bar{g}_T(\tau;\beta_0) \\ \mathrm{vec}\big(\bar{D}_T(\tau) - J(\beta_0)\big) \end{pmatrix} \Rightarrow \mathcal{N}\left(0,\, \begin{pmatrix} \Omega(\tau) & 0 \\ 0 & V_{hh\cdot g}(\tau) \end{pmatrix}\right), \tag{B.3}$$

where $V_{hh\cdot g}(\tau) = V_{hh}(\tau) - V_{hg}(\tau)\,\Omega(\tau)^{-1}\,V_{gh}(\tau)$ is the Schur complement. The off-diagonal block vanishes because the orthogonalization is constructed precisely to remove the pre-

dictable component of $\bar{H}_T(\tau)$ given $\bar{g}_T(\tau)$; this is the kernel analogue of the argument in Kleibergen (2005, Theorem 1).

Since $\bar{D}_T(\tau) \xrightarrow{p} J(\beta_0)$ and the randomness in $\bar{D}_T(\tau)$ is asymptotically independent of $\bar{g}_T(\tau)$, the projection $P_{\bar{D}_T(\tau)}$ converges in probability to the deterministic projection $P_{J_0} = \Omega(\tau)^{-1/2} J(\beta_0)(J(\beta_0)'\Omega(\tau)^{-1}J(\beta_0))^{-1}J(\beta_0)'\Omega(\tau)^{-1/2}$, independently of the limiting distribution of $\bar{g}_T(\tau)$.

Now write

$$K_T^{\text{loc}}(\tau;\beta_0) = \left\| \sqrt{\frac{Th_T}{R(K)}}\, \widehat{\Omega}_T(\tau)^{-1/2}\, \bar{g}_T(\tau) \right\|^2_{P_{\bar{D}_T(\tau)}},$$

where $\|x\|_P^2 = x'Px$. By the asymptotic independence established in (B.3), $P_{\bar{D}_T(\tau)}$ and $\sqrt{Th_T/R(K)}\,\widehat{\Omega}_T(\tau)^{-1/2}\bar{g}_T(\tau)$ are asymptotically independent. The latter converges to $\mathcal{N}(0, I_L)$ by the proof of Proposition 1. Projecting a standard normal L-vector onto a fixed K-dimensional subspace yields a χ^2_K variate. By Slutsky's theorem and the continuous mapping theorem,

$$K_T^{\text{loc}}(\tau;\beta_0) \xrightarrow{d} \chi^2_K.$$

The limit is uniform over the local-to-zero sequences in Assumption 1 because neither the CLT in (B.1) nor the orthogonalization depend on the identification strength: $J(\beta_0)$ may be zero or near-zero, and $P_{\bar{D}_T(\tau)}$ remains a valid rank-K projection (or has rank less than K, in which case the statistic is stochastically dominated by χ^2_K). □

B.0.2 Proof of Corollary 1: Stability K-statistic

The proof reduces the K-statistic results to the corresponding S-statistic results (Theorem 1) via a fixed-rank projection.

Step 1: Decomposition. By Assumption 4(b), replace $\widehat{\Omega}_T$ with Ω throughout (Slutsky). The full-sample orthogonalized Jacobian $\bar{D}$ defined in (31) satisfies $\bar{D} \xrightarrow{p} J(\beta_0)$, so $P_{\bar{D}} \xrightarrow{p} P_{J_0} = \Omega^{-1/2}J(\beta_0)(J(\beta_0)'\Omega^{-1}J(\beta_0))^{-1}J(\beta_0)'\Omega^{-1/2}$, a deterministic rank-$K$ projection.

The decomposition $\mathcal{K}_T^{\text{int}} = (h_T/R(K))K_T^{\text{full}} + \mathcal{K}_T^{\text{stab}} + o_p(1)$ follows by exactly the same algebra as Theorem 1(i) (Appendix A.3.1), with $\widehat{\Omega}_T^{-1}$ replaced by $\widehat{\Omega}_T^{-1/2}P_{\bar{D}}\widehat{\Omega}_T^{-1/2}$ throughout. The cross-term bound uses Lemma 1 identically, since the integrated-weights property is a statement about $w_{t,T}(\tau)$ and does not depend on the weighting matrix.

Step 2: Null distribution via projection of the Gaussian bridge. Define the whitened, centered process

$$\tilde{G}_T(\tau) = \sqrt{\frac{Th_T}{R(K)}}\, \Omega^{-1/2}\tilde{g}_T(\tau;\beta_0),$$

which is the same object as in the proof of Theorem 1(iii) (Appendices A.3.4–A.3.5). Under H_0^{stab}, the S-statistic proofs establish:

(a) Fixed bandwidth: $\tilde{G}_T(\cdot) \Rightarrow \tilde{G}(\cdot)$ in $C[0,1]$, where $\tilde{G}$ is an L-dimensional centered Gaussian process with covariance $\text{Cov}(\tilde{G}(\tau_1), \tilde{G}(\tau_2))$ determined by $\rho((\tau_2 - \tau_1)/h) I_L$ (Appendix A.3.4).

(b) Shrinking bandwidth: the integrated quadratic form $\int_0^1 \|\tilde{G}_T(\tau)\|^2 \, d\tau$ satisfies a CLT with mean L and variance $4Lh_T C_B R(K)^{-2}$ (Appendix A.3.5).

Now, $P_{\bar{D}} \xrightarrow{p} P_{J_0}$, so by Slutsky the stability K-statistic satisfies

$$\mathcal{K}_T^{\text{stab}}(\beta_0) = \int_0^1 \tilde{G}_T(\tau)' \, P_{\bar{D}} \, \tilde{G}_T(\tau) \, d\tau = \int_0^1 \tilde{G}_T(\tau)' \, P_{J_0} \, \tilde{G}_T(\tau) \, d\tau + o_p(1). \tag{B.4}$$

Define $\tilde{\xi}_T(\tau) = P_{J_0}^{1/2} \tilde{G}_T(\tau) \in \mathbb{R}^K$, where $P_{J_0}^{1/2}$ is any $K \times L$ matrix satisfying $P_{J_0}^{1/2\prime} P_{J_0}^{1/2} = P_{J_0}$—for instance, $P_{J_0}^{1/2} = (J(\beta_0)'\Omega^{-1} J(\beta_0))^{-1/2} J(\beta_0)'\Omega^{-1/2}$. Then

$$\mathcal{K}_T^{\text{stab}}(\beta_0) = \int_0^1 \|\tilde{\xi}_T(\tau)\|^2 \, d\tau + o_p(1).$$

Since $P_{J_0}^{1/2}$ is a fixed $K \times L$ matrix and the map $x(\cdot) \mapsto P_{J_0}^{1/2} x(\cdot)$ is continuous on $C[0,1]$, the continuous mapping theorem gives:

Part (i): Fixed bandwidth. $\tilde{\xi}_T(\cdot) \Rightarrow \tilde{\xi}(\cdot)$ in $C[0,1]$, where $\tilde{\xi}(\tau) = P_{J_0}^{1/2} \tilde{G}(\tau)$ is a K-dimensional centered Gaussian process with covariance

$$\text{Cov}(\tilde{\xi}(\tau_1), \tilde{\xi}(\tau_2)) = P_{J_0}^{1/2} \, \text{Cov}(\tilde{G}(\tau_1), \tilde{G}(\tau_2)) \, P_{J_0}^{1/2\prime} = \rho\Big(\frac{\tau_2 - \tau_1}{h}\Big) P_{J_0}^{1/2} P_{J_0}^{1/2\prime} = \rho\Big(\frac{\tau_2 - \tau_1}{h}\Big) I_K.$$

The last equality uses $P_{J_0}^{1/2} P_{J_0}^{1/2\prime} = I_K$, which holds by the definition of $P_{J_0}^{1/2}$ and the idempotency of P_{J_0}. The continuous mapping theorem then gives $\int_0^1 \|\tilde{\xi}_T(\tau)\|^2 \, d\tau \Rightarrow \int_0^1 \|\tilde{\xi}(\tau)\|^2 \, d\tau$, which is (37).

Part (ii): Shrinking bandwidth. The proof of Theorem 1(iii)(c) (Appendix A.3.5) expresses the S-statistic as a quadratic form in whitened scores $\xi_t = \Omega^{-1/2} g_t \in \mathbb{R}^L$, decomposes into diagonal and off-diagonal terms, and applies a martingale CLT to the off-diagonal part. The same argument applies verbatim to the K-dimensional projected scores $\eta_t = P_{J_0}^{1/2} \xi_t \in \mathbb{R}^K$. Since $P_{J_0}^{1/2}$ is a fixed matrix, $\{\eta_t\}$ inherits the FCLT and moment conditions from $\{\xi_t\}$. The diagonal term converges to $R(K) \cdot K$ (replacing L by K because $\text{E}[\eta_t'\eta_t] = \text{tr}(P_{J_0}^{1/2} P_{J_0}^{1/2\prime}) = K$), and the off-diagonal martingale CLT yields variance $4KC_B R(K)^{-2} h_T$. This gives (38). □

B.0.3 Proof of Corollary 2: Plug-in robustness

Since $g_t(\beta) = z_t(y_t - x_t'\beta)$ is affine in β, the centered local average satisfies $\tilde{g}_T(\tau; \hat{\beta}) = \tilde{g}_T(\tau; \beta_0) - (\bar{J}_T(\tau) - \bar{J})\Delta_T$ with $\Delta_T = \hat{\beta} - \beta_0$, exactly as in Lemma 3. The stability

K-statistic at $\hat{\beta}$ is

$$\mathcal{K}_T^{\text{stab}}(\hat{\beta}) = \int_0^1 \frac{Th_T}{R(K)}\, \tilde{g}_T(\tau;\hat{\beta})'\, \widehat{\Omega}_T^{-1/2}\, P_{\bar{D}}\, \widehat{\Omega}_T^{-1/2}\, \tilde{g}_T(\tau;\hat{\beta})\, d\tau.$$

Since $P_{\bar{D}}$ is a contraction ($\|P_{\bar{D}}x\| \leq \|x\|$), the bound in Proposition 3 (Appendix A.3.6, Step 1) carries over:

$$\left|\mathcal{K}_T^{\text{stab}}(\hat{\beta}) - \mathcal{K}_T^{\text{stab}}(\beta_0)\right| \leq \left|\mathcal{S}_T^{\text{stab}}(\hat{\beta}) - \mathcal{S}_T^{\text{stab}}(\beta_0)\right|,$$

where the inequality follows from $x'P_{\bar{D}}x \leq x'x$ for any x, applied to the integrand difference. More precisely, writing $a_T(\tau) = \widehat{\Omega}_T^{-1/2}\tilde{g}_T(\tau;\beta_0)$ and $b_T(\tau) = -\widehat{\Omega}_T^{-1/2}(\bar{J}_T(\tau) - \bar{J})\Delta_T$ as in Appendix A.3.6, we have

$$\begin{aligned}\left|\|a_T + b_T\|_{P_{\bar{D}}}^2 - \|a_T\|_{P_{\bar{D}}}^2\right| &\leq 2\|P_{\bar{D}}a_T\|\,\|P_{\bar{D}}b_T\| + \|P_{\bar{D}}b_T\|^2 \\ &\leq 2\|a_T\|\,\|b_T\| + \|b_T\|^2,\end{aligned}$$

so the integrated difference is bounded by the corresponding S-statistic difference. The rates $o_p(1)$ (fixed bandwidth) and $o_p(\sqrt{h_T})$ (shrinking bandwidth) established in (A.42)–(A.43) therefore apply, and the result follows by Slutsky. □

C SP-IV Implementation Details

This appendix provides a self-contained description of the System Projections on Instrumental Variables (SP-IV) estimator of Lewis and Mertens (2022), as implemented in the Monte Carlo experiments (Section 4) and the empirical application (Section 5).

C.1 General SP-IV framework

Consider a structural equation $y_t = Y_t'\beta + u_t$ with K endogenous regressors Y_t and N_z instruments z_t. Let $w_t = (y_t, Y_t', q_t')'$ be a vector of variables that follows a VAR(p):

$$w_t = \Phi_1 w_{t-1} + \cdots + \Phi_p w_{t-p} + c + \eta_t, \tag{C.1}$$

where c is an intercept and η_t is the reduced-form innovation. Let n denote the dimension of w_t and let $\hat{A}$ collect the OLS coefficient estimates. From the companion-form representation, compute the h-step impulse response matrices $\Phi_s^{(h)} = JA_{\text{comp}}^h J'$ for $h = 0, 1, \ldots, H$, where A_{comp} is the companion matrix and $J = [I_n, 0_{n\times n(p-1)}]$.

Forecast errors. Using the moving-average representation implied by the estimated VAR, the h-step-ahead forecast error of variable i at time t is

$$u_{i,t}^{\perp}(h) = \sum_{j=0}^{h} \Phi_s^{(j)} \hat{\eta}_{t+h-j}, \tag{C.2}$$

where $\hat{\eta}_t = w_t - \hat{\Phi}_1 w_{t-1} - \cdots - \hat{\Phi}_p w_{t-p} - \hat{c}$ are the VAR residuals. By construction, $u_{i,t}^{\perp}(h)$ is orthogonal to the information set $\{w_{t-1}, w_{t-2}, \ldots\}$ spanned by the VAR conditioning variables. Stacking across horizons yields the $H \times T_g$ matrices $y_H^{\perp}$ and $Y_H^{\perp}$ (the latter of dimension $HK \times T_g$), where $T_g = T - p - H - 1$ is the effective sample size after accounting for lags and leads.[9]

SP-IV estimator. Following Lewis and Mertens (2022), the SP-IV estimator uses the baseline weighting matrix $W_{\text{SP}} = I_H \otimes Q^{-1}$, where $Q = T_g^{-1}\sum_t z_t^{\perp} z_t^{\perp\prime}$. Define the instrument projection matrix $P_Z = Z^{\perp\prime}(Z^{\perp}Z^{\perp\prime})^{-1}Z^{\perp}$ and the restriction matrix $R = I_K \otimes \text{vec}(I_H)$. The estimator is

$$\hat{\beta} = \left[R'(Y_H^{\perp} P_Z Y_H^{\perp\prime} \otimes I_H)R\right]^{-1} R' \,\text{vec}(y_H^{\perp} P_Z Y_H^{\perp\prime}). \tag{C.3}$$

When VAR dynamics are imposed on the impulse responses—our preferred implementation following Lewis and Mertens (2022)—one replaces $y_H^{\perp} P_Z Y_H^{\perp\prime}$ with $\hat{\Theta}_y(\hat{\Theta}_Y)'$

[9] In the implementation, we drop the last $H+1$ observations from the forecast-error matrices so that all horizons are available for every retained observation. The instrument is aligned as $z_t^{\perp}$ for $t = 2, \ldots, T_g+1$ (shifted by one period relative to the VAR residuals) to ensure proper lead–lag alignment.

and $Y_H^{\perp} P_Z Y_H^{\perp\prime}$ with $\hat{\Theta}_Y(\hat{\Theta}_Y)'$, where $\hat{\Theta}_y \in \mathbb{R}^{H\times N_z}$ and $\hat{\Theta}_Y \in \mathbb{R}^{HK\times N_z}$ are the VAR-based impulse response functions of y_t and Y_t to the instrument.

Covariance of the moment conditions. The SP-IV moment conditions are $g_t(\beta) = z_t^{\perp} \otimes u_{H,t}^{\perp}(\beta)$, where $u_{H,t}^{\perp}(\beta)$ is the H-vector of residual forecast errors at parameter value β. Under the maintained assumption that $z_t^{\perp}$ is serially uncorrelated—which holds when the VAR conditioning set X_{t-1} adequately controls for predictable variation in z_t—the product $z_t^{\perp}\otimes u_{H,t}^{\perp}(\beta)$ is also serially uncorrelated (Lewis and Mertens, 2022, Proposition 2). The covariance of the moment conditions therefore takes the Kronecker form

$$\Omega = \Sigma_{u_H^{\perp}} \otimes Q, \tag{C.4}$$

where $\Sigma_{u_H^{\perp}} = \mathrm{E}[u_{H,t}^{\perp}(\beta)u_{H,t}^{\perp}(\beta)']$ is the $H \times H$ covariance of the forecast-error residuals. This structure eliminates the need for HAC estimation of Ω.[10] We estimate $\Sigma_{u_H^{\perp}}$ by $\hat{\Sigma}_{u_H^{\perp}} = \hat{u}_H^{\perp}\hat{u}_H^{\perp\prime}/(T_g - K)$ as in Lewis and Mertens (2022, eq. (19)).

C.2 Implementation in the Monte Carlo experiments

The DGP in Section 4 generates data from a bivariate system $w_t = (y_t, x_t)'$ with $n = 2$. The instrument $z_t = \varepsilon_t^i$ (or $z_t \sim \mathcal{N}(0,1)$ in the uninformative case) is a scalar, $N_z = 1$. We estimate a VAR(4) on w_t and construct forecast errors via (C.2) up to horizon H. The forecast errors are thinned by a factor of $n_{\text{thin}} = 2$ (retaining every second horizon), yielding $H_h = H/n_{\text{thin}}$ effective horizons.

In the forward-looking NKPC specification (39), the structural equation involves three endogenous regressors: $Y_t = (x_t, y_{t-1}, \mathrm{E}_t y_{t+1})'$, so $K = 3$. The SP-IV stacking assigns the forecast-error matrices as follows:

$$\begin{aligned}
&y_H^{\perp} : \text{forecast errors of } y_t \text{ at horizons } h \in \{2, 4, \ldots, H\},\\
&Y_H^{\perp} : \text{vertically stacking forecast errors of } x_t \text{ at } h \in \{2, 4, \ldots, H\},\\
&\qquad y_t \text{ at } h \in \{1, 3, \ldots, H-1\} \text{ (lagged)},\\
&\qquad y_t \text{ at } h \in \{3, 5, \ldots, H+1\} \text{ (forward)}.
\end{aligned}$$

This yields $L = H_h \times N_z$ moment conditions for $K = 3$ parameters. With $H = 20$ and $n_{\text{thin}} = 2$, we obtain $L = 10$ moments.

The SP-IV estimator is computed from (C.3) with $R = I_3 \otimes \mathrm{vec}(I_{H_h})$ and the weighting matrix $I_{H_h} \otimes \hat{Q}^{-1}$. For the stability and local-inference statistics, the score vector at

[10] When the instrument is not rendered serially uncorrelated by the VAR conditioning set—for instance, because the VAR lag length is insufficient—one may instead use a HAC estimator of Ω. In our simulations and empirical application, the VAR(4) specification ensures adequate whitening of the instrument, and we use the plug-in estimator $\hat{\Omega} = \hat{\Sigma}_{u_H^{\perp}} \otimes \hat{Q}$ throughout.

observation t is

$$g_t(\beta) = \big(z_t \cdot e_{1,t}(\beta),\ \ldots,\ z_t \cdot e_{H_h,t}(\beta)\big)' \in \mathbb{R}^L, \tag{C.5}$$

where $e_{j,t}(\beta)$ is the j-th horizon residual forecast error evaluated at β, and the covariance is estimated as $\hat{\Omega} = \hat{\Sigma}_{u_H^\perp} \otimes \hat{Q}$.

C.3 Implementation in the empirical application

The empirical application in Section 5 estimates a monthly hybrid NKPC with the long-run restriction $\gamma_b + \gamma_f = 1$ imposed, reducing the parameter vector to $\beta = (\gamma_f, \lambda)'$ with $K = 2$. The VAR is estimated on a system of n macroeconomic variables including monthly Core CPI inflation and the unemployment rate, with $p = 4$ lags. The instrument is the Main Business Cycle (MBC) shock of Angeletos et al. (2020), identified as the unit-length vector r that maximizes the forecast-error-variance share of the target variable at business-cycle frequencies (**?**).

Annualization matrices. Because the NKPC (48) involves annual inflation rates constructed from monthly data, the SP-IV forecast errors must be transformed accordingly. Define the $H \times H$ annualization matrices

$$C_1(i,j) = \frac{1}{12}\,\mathbf{1}(1 \le j - i \le 12), \qquad C_2(i,j) = \frac{1}{12}\,\mathbf{1}(1 \le i - j \le 12), \tag{C.6}$$

where H is the total number of forecast-error horizons. The matrix C_2 converts horizon-h forecast errors into the backward-looking annual average (capturing π^{1y}_{t-1}), while C_1 converts them into the forward-looking annual average (capturing π^{1y}_{t+12}). Using the $H \times T_g$ matrix $U^\perp_{\pi,H}$ of inflation forecast errors and $U^\perp_{U,H}$ of unemployment forecast errors, the transformed SP-IV objects are:

$$y^\perp_H = U^\perp_{\pi,H} - C_2\, U^\perp_{\pi,H}, \tag{C.7}$$

$$Y^\perp_H = \begin{pmatrix} C_1\, U^\perp_{\pi,H} - C_2\, U^\perp_{\pi,H} \\ U^\perp_{U,H} \end{pmatrix}, \tag{C.8}$$

which correspond to the forecast errors of the restricted NKPC left-hand side ($\pi^{1m}_t - \pi^{1y}_{t-1}$) and right-hand side ($(\pi^{1y}_{t+12} - \pi^{1y}_{t-1}),\ U_t)'$, respectively.

We select horizons $h \in \{0, 3, 6, \ldots, 33\}$, yielding $H_h = 12$ rows from each transformed matrix, and hence $L = 12$ moment conditions for $K = 2$ parameters. The SP-IV estimator and its covariance are computed as in Section C.1, with VAR dynamics imposed on the impulse responses. For the stability tests, the score vector $g_t(\beta)$ is constructed from (C.5) using the transformed forecast-error residuals, and $\hat{\Omega}$ is estimated from the sample covariance of $\{g_t(\hat{\beta})\}_{t=1}^{T_g}$.

Local (kernel-weighted) SP-IV. For local-in-time inference at rescaled time τ, all reduced-form quantities—VAR coefficients, forecast errors, the MBC shock identification, and the instrument projection—are estimated using Nadaraya–Watson kernel weights $w_{t,T}(\tau)$. Operationally, this is implemented by premultiplying each observation by $\sqrt{w_{t,T}(\tau)}$ and applying the standard SP-IV formulas to the reweighted sample, so that all inner products become kernel-weighted averages. The local SP-IV estimator $\hat{\beta}(\tau)$ is then obtained from (C.3) evaluated at the kernel-weighted IRFs, as described in equation (50). The local K-statistic uses the same kernel-weighted moment conditions and Jacobian (which equals $-\hat{\Theta}_Y(\tau)$ under the imposed VAR dynamics).

D Critical values for the stability statistic

Under fixed-bandwidth asymptotics (Assumption 3(fixed)), the stability component converges to $\mathcal{S}_T^{\text{stab}}(\beta_0) \Rightarrow Q \equiv \int_0^1 \|\tilde{G}(\tau)\|^2 \, d\tau$, where $\tilde{G}(\tau) = G(\tau) - \int_0^1 G(s)\, ds$ is the centered version of an L-dimensional Gaussian process G with covariance kernel $\text{Cov}(G(\tau_1), G(\tau_2)) = \rho\big((\tau_2 - \tau_1)/h\big) I_L$ and $\rho(u) = R(K)^{-1} \int_{-1}^{1} K(v)K(v+u)\, dv$. The distribution of Q depends on L, h, and the kernel K, but not on the data or on identification strength, and can be simulated to arbitrary precision.

We describe two equivalent simulation strategies. The first—smoothing synthetic white noise—is faster and easier to implement. The second—direct Cholesky sampling from the GP covariance—is a useful cross-check.

Algorithm 1: Smoothing white noise. The key observation is that NW-smoothing i.i.d. standard normal draws and rescaling reproduces the finite-dimensional distributions of G on any grid. Algorithm 1 summarizes the procedure.

Algorithm 1 GP critical values via smoothed white noise

Require: Dimension L; bandwidth $h \in (0,1)$; kernel K; grid size T; number of replications B; significance levels $\alpha_1, \ldots, \alpha_J$

Ensure: Critical values $c_{\alpha_1}, \ldots, c_{\alpha_J}$

1: Compute kernel roughness $R(K) = \int_{-1}^{1} K(u)^2\, du$
2: Set window radius $m = \lfloor Th \rfloor$; grid spacing $\Delta = 1/T$
3: Precompute kernel vector: $k_j = K(j/(Th))$ for $j = -m, \ldots, m$
4: **for** $b = 1, \ldots, B$ **do**
5: Draw $\varepsilon_t^{(\ell)} \overset{iid}{\sim} N(0,1)$ for $t = 1, \ldots, T$, $\ell = 1, \ldots, L$
6: *(Optional: boundary reflection)* Prepend ε_{m+1-j} for $j = 1, \ldots, m$ and append ε_{T+1-j} for $j = 1, \ldots, m$
7: **for** each component $\ell = 1, \ldots, L$ **do**
8: Smooth via NW weights: $\bar{\varepsilon}_t^{(\ell)} = \sum_{j=-m}^{m} k_j\, \varepsilon_{t+j}^{(\ell)} \Big/ \sum_{j=-m}^{m} k_j$ (use FFT for speed)
9: **end for**
10: If reflection was used, extract the middle T values
11: Scale: $G_t = \sqrt{Th/R(K)}\; \bar{\varepsilon}_t \in \mathbb{R}^L$
12: Bridge: $\tilde{G}_t = G_t - T^{-1} \sum_{s=1}^{T} G_s$
13: Integrate: $Q_b = \Delta \sum_{t=1}^{T} \|\tilde{G}_t\|^2$
14: **end for**
15: Return $c_{\alpha_j} = (1-\alpha_j)$-quantile of $\{Q_1, \ldots, Q_B\}$

Step 6 implements boundary reflection: the extended sample is symmetric at both endpoints, so the NW denominator is (approximately) constant across $\tau \in (0,1)$ and the covariance of the smoothed process is stationary with kernel $\rho(u/h)$.[11,12]

In our implementations, we use $T = 6{,}000$, $B = 20{,}000$, and the Epanechnikov kernel throughout. The convolution in Step 8 is computed via FFT, so each replication costs $O(TL \log T)$ operations.

Under shrinking-bandwidth asymptotics (Assumption 3(shrinking)), the stability statistic is asymptotically normal (Theorem 1(iii)(c)) and critical values are read directly from the standard normal table:

$$\text{reject } H_0^{\text{stab}} \text{ if } \quad \frac{\mathcal{S}_T^{\text{stab}}(\hat{\beta}) - L}{\sqrt{h_T \cdot 4L\, C_B\, R(K)^{-2}}} > z_{1-\alpha},$$

where $C_B = \int_0^2 [\int_{-1}^{1} K(u)K(u+v)\, du]^2\, dv$ and $z_{1-\alpha}$ is the standard normal quantile. No simulation is required in this case.

[11] Without reflection, the NW denominator varies near $\tau \in \{0,1\}$ and the GP covariance becomes boundary-nonstationary. The algorithm remains valid if reflection is omitted: one simply smooths the original T-vector with the actual (boundary-varying) NW weights, and the resulting draws are from the correct boundary-nonstationary GP. In our experiments, the two approaches yield critical values that agree to within Monte Carlo error for $B \geq 10{,}000$.

[12] An alternative to Algorithm 1 is direct Cholesky sampling. With boundary reflection, the covariance matrix $\Sigma_{ij} = \rho\big((\tau_i - \tau_j)/h\big)$ is Toeplitz on a uniform grid $\tau_1, \ldots, \tau_N$. One computes $L_{\text{chol}} = \text{chol}(\Sigma)$, draws $Z \sim N(0, I_N)$, sets $G = L_{\text{chol}} Z$, bridges, and integrates as in Steps 12–14. This approach is numerically exact on the chosen grid but requires $O(N^2)$ storage and $O(N^3)$ factorization, making it slower than the FFT-based smoothing approach for fine grids.