Supplemental Material for "A Smoothed P-Value Test When There is a Nuisance Parameter under the Alternative"

Jonathan B. Hill*
University of North Carolina – Chapel Hill

April 11, 2022

Contents

*Dept. of Economics, University of North Carolina, Chapel Hill; www.unc.edu/∼jbhill; jbhill@email.unc.edu

A Outline

Appendix B shows how the PVOT estimates weighted average power of the underlying test when framed in the setting of Andrews and Ploberger (1994).

In Appendix C we tackle local power for a PVOT test of omitted nonlinearity. We prove Theorem 3.3 which provides sufficient conditions for a required weak convergence property.

Finally, Appendix D contains omitted figures from the main paper.

Recall how the PVOT is constructed. We assume $\mathcal{T}_n(\lambda) \geq 0$, and that large values are indicative of H_1. Let $p_n(\lambda)$ be a p-value or asymptotic p-value based on $\mathcal{T}_n(\lambda)$: $p_n(\lambda)$ may be based on a known limit distribution, or if the limit distribution is non-standard then a bootstrap or simulation method is assumed available for computing an asymptotically valid approximation to $p_n(\lambda)$. Assume that $p_n(\lambda)$ leads to an asymptotically correctly sized test, uniformly on Λ:

$$\sup_{\lambda \in \Lambda} |P(p_n(\lambda) < \alpha | H_0) - \alpha| \to 0 \text{ for any } \alpha \in (0,1). \tag{1}$$

The p-value [PV] test with nominal level α for a chosen value of λ is (1):

$$\textbf{PV Test: } \text{reject } H_0 \text{ if } p_n(\lambda) < \alpha, \text{ otherwise fail to reject } H_0. \tag{2}$$

Now assume Λ has unit Lebesgue measure $\int_\Lambda d\lambda = 1$, and compute the *p-value occupation time* [PVOT] of $p_n(\lambda)$ below the nominal level $\alpha \in (0,1)$:

$$\textbf{PVOT: } \mathcal{P}_n^*(\alpha) \equiv \int_\Lambda I(p_n(\lambda) < \alpha) \, d\lambda, \tag{3}$$

where $I(\cdot)$ is the indicator function. If $\int_\Lambda d\lambda \neq 1$ then we use $\mathcal{P}_n^*(\alpha) \equiv \int_\Lambda I(p_n(\lambda) < \alpha) d\lambda / \int_\Lambda d\lambda$.

Recall $\Rightarrow^*$ denotes weak convergence on l_∞, the space of bounded functions with sup-norm topology (Dudley, 1978; Pollard, 1984; Hoffman-Jϕrgensen, 1991).

Assumption 1 (weak convergence)**.** *Let H_0 be true.*

a. $\{\mathcal{T}_n(\lambda)\} \Rightarrow^* \{\mathcal{T}(\lambda)\}$, a process with a version that has almost surely *uniformly continuous sample paths (with respect to some norm $||\cdot||$). $\mathcal{T}(\lambda) \geq 0$ a.s., $\sup_{\lambda \in \Lambda} \mathcal{T}(\lambda) < \infty$ a.s., and $\mathcal{T}(\lambda)$ has an absolutely continuous distribution function $F_0(c) \equiv P(\mathcal{T}(\lambda) \leq c)$ that is not a function of λ.*

b. Under H_1^L weak convergence (11) is valid with $c(\lambda) = E[w_t^2(\lambda)]/(E[\epsilon_t^2 w_t^2(\lambda)])^{1/2} > 0$ where $w_t(\lambda) \equiv F_t(\lambda)$ $-$ $E[F_t(\lambda) g_t(\zeta_0)'] \times (E[g_t(\zeta_0) g_t(\zeta_0)'])^{-1} g_t(\zeta_0)$.

We use the following notation. $[z]$ rounds z to the nearest integer. $I(\cdot)$ is the indicator function: $I(A)$ $= 1$ if A is true, otherwise $I(A) = 0$. $a_n/b_n \sim c$ implies $a_n/b_n \to c$ as $n \to \infty$. $|\cdot|$ is the l_1-matrix norm.

$|| \cdot ||_p$ is the L_p-norm. $K > 0$ is a finite constant whose value may change from place to place. $\iota > 0$ is a tiny number whose value may be different in different places. $0_{a \times b}$ is a $a \times b$ dimensional matrix of zeros.

B PVOT as a Measure of Power and Test Optimality

We work in Andrews and Ploberger's 1994 likelihood framework with a nuisance parameter under the alternative. The null hypothesis is therefore composite, in which case it is convenient to work with weighted average power. We show how the PVOT relates to weighted average power.

Let $\mathcal{Y}_n \equiv [y_1, ..., y_n]'$ be an observed sample of variables $y_t \in \mathbb{R}^k$, with joint probability density $f(y, \theta_0, \lambda)$, $y \in \mathbb{R}^{nk}$ and $\theta_0 = [\beta_0', \delta_0']' \in \mathbb{R}^s$ where $\beta_0 \in \mathbb{R}^r, 0 < r \leq s$. If $\beta_0 = 0$ then the distribution $f(y, \theta_0)$ does not depend on λ. Thus, in this section λ is assumed to be part of the data generating process. Assume $f(y, \theta, \lambda) > 0$ *almost everywhere* on $\mathbb{S} \times \Theta \times \Lambda$, for some subset $\mathbb{S} \subseteq \mathbb{R}^{nk}$, Θ is a compact subset of $\mathbb{R}^s$ containing θ_0, and $\int_\Lambda d\lambda = 1$ by convention.

We want to test $H_0 : \beta_0 = 0$ against $H_1 : \beta_0 \neq 0$, in which case λ is part of the data generating process only under H_1. Consider a sequence of local alternatives H_1^L of the form $f(y, \theta_0 + \mathcal{N}_n^{-1}b, \lambda)$ where $\mathcal{N}_n = [\mathcal{N}_{i,j,n}]_{i,j=1}^s$ is a diagonal matrix, $b \in \mathbb{R}^s$, and $\mathcal{N}_{i,i,n} \to \infty$. Under regular asymptotics $\mathcal{N}_n = \sqrt{n} I_s$, but $\mathcal{N}_n$ may differ from $\sqrt{n} I_s$ if some variables are trending, or negligible trimming is used for possibly heavy tailed data (e.g. Hill and Aguilar, 2013).

Let $\xi(\mathcal{Y}_n, b, \lambda) \in \{0, 1\}$ be any asymptotic level α test of H_0 for some imputed (b, λ), and as in Andrews and Ploberger (1994) let Q_λ be for each λ be an absolutely continuous probability measure and $\mathbb{R}^s$, and let J be an integrable weight function on Λ. For example, the LR statistic is $\xi(\mathcal{Y}_n, b, \lambda) = I(f(\mathcal{Y}_n, \theta_0 + \mathcal{N}_n^{-1}b, \lambda)/f(\mathcal{Y}_n, \theta_0) > c_{n,\alpha}(b, \lambda))$ where $c_{n,\alpha}(b, \lambda)$ is the asymptotic level α critical value, hence $E[\xi(\mathcal{Y}_n, b, \lambda)] \to \alpha$ under H_0. Andrews and Ploberger (1994) require Q_λ to be a Gaussian density that depends on λ in order to show that their exp-LM statistic is optimal. We allow Q_λ to depend on λ merely for generality, but it is not imperative for showing how the PVOT relates to weighted average power.[1]

A test of H_0 against the sequence of simple alternatives $\{f(y, \theta_0 + \mathcal{N}_n^{-1}b, \lambda) : (b, \lambda) \in \mathbb{R}^s \times \Lambda\}_{n \geq 1}$ has weighted average local power (cf. Andrews and Ploberger, 1994)

$$\int_\Lambda \int_{\mathbb{R}^s} \left[\int_{\mathbb{R}^{nk}} \xi(y, b, \lambda) f(y, \theta_0 + \mathcal{N}_n^{-1}b, \lambda) dy \right] dQ_\lambda(b) dJ(\lambda).$$

Now let $g(y)$ be any joint probability measure that is positive on $\mathbb{R}^{nk}$ *a.e.*, define the expectations operator

[1] Andrews and Ploberger (1994) fix $Q_\lambda(b) = N(0, c\Sigma_\lambda)$ for some constant $c > 0$ that guides weight toward certain alternatives, and a covariance matrix Σ_λ that depends on λ. They also use Lebesgue measure J for the weight on Λ in their simulations as a default tactic when information about the true λ under H_1 is not available.

$E_g[m(\mathcal{Z})] \equiv \int_{\mathbb{R}^{nk}} m(z)g(z)dz$ for an arbitrary scalar mapping m, and define:

$$d\omega(y, \theta_0 + \mathcal{N}_n^{-1}b, \lambda) \equiv \frac{f(y, \theta_0 + \mathcal{N}_n^{-1}b, \lambda)}{g(y)} dQ_\lambda(b)dJ(\lambda).$$

We do not require $d\omega(y, \theta_0 + \mathcal{N}_n^{-1}b, \lambda)$ to be a probability measure, although it will be for an obvious choice of $g(y)$ discussed below. By Fubini's Theorem, and the definitions of $d\omega$ and E_g:

$$\begin{aligned}
&\int_\Lambda \int_{\mathbb{R}^s} \left[\int_{\mathbb{R}^{nk}} \xi(y, b, \lambda) f(y, \theta_0 + \mathcal{N}_n^{-1}b, \lambda)dy\right] dQ_\lambda(b)dJ(\lambda) \\
&\quad = \int_{\mathbb{R}^{nk}} \left[\int_\Lambda \int_{\mathbb{R}^s} \xi(y, b, \lambda) \frac{f(y, \theta_0 + \mathcal{N}_n^{-1}b, \lambda)}{g(y)} dQ_\lambda(b)dJ(\lambda)\right] g(y)dy \\
&\quad = E_g\left[\int_\Lambda \int_{\mathbb{R}^s} \xi(y, b, \lambda)d\omega(y, \theta_0 + \mathcal{N}_n^{-1}b, \lambda)\right].
\end{aligned}$$

We will call the above integral under expectations,

$$\mathcal{P}_\xi^{(\omega)}(\mathcal{Y}_n) \equiv \int_\Lambda \int_{\mathbb{R}^s} \xi(\mathcal{Y}_n, b, \lambda)d\omega(\mathcal{Y}_n, \theta_0 + \mathcal{N}_n^{-1}b, \lambda), \tag{4}$$

the ω-PVOT since it gives the ω measure of the subset of local drift b and nuisance parameter λ on which a test based on $\xi(\mathcal{Y}_n, b, \lambda)$ rejects H_0 in favor of $f(y, \theta_0 + \mathcal{N}_n^{-1}b, \lambda)$. Thus, $\mathcal{P}_\xi^{(\omega)}(\mathcal{Y}_n)$ is a generalized version of the PVOT. We say *generalized* because it smooths over both the nuisance parameter λ and local drift b, and $d\omega$ need not be Lebesgue measure.[2] Weighted average local power is therefore a mean ω-PVOT:

$$\int_\Lambda \int_{\mathbb{R}^s} \left[\int_{\mathbb{R}^{nk}} \xi(y, b, \lambda) f(y, \theta_0 + \mathcal{N}_n^{-1}b, \lambda)dy\right] dQ_\lambda(b)dJ(\lambda) = E_g\left[\mathcal{P}_\xi^{(\omega)}(\mathcal{Y}_n)\right]. \tag{5}$$

The ω-PVOT provides a natural way to rank tests: a test is optimal, in the sense of having the highest weighted average local power for given probability measures (J, Q_λ), *if and only if* it has the highest mean ω-PVOT. This seems natural since relative to all other tests an optimal test should spend more time in the rejection region, over the nuisance parameter λ and local drift b.

As a special case, the probability measure

$$g(y) = \int_\Lambda \int_{\mathbb{R}^s} f(y, \theta_0 + \mathcal{N}_n^{-1}\tilde{b}, \tilde{\lambda})dQ_\lambda(\tilde{b})dJ(\tilde{\lambda}) \text{ on } \mathbb{R}^{nk} \tag{6}$$

yields a probability measure $d\omega$ on $\mathbb{R}^s \times \Lambda$ for each y:

$$d\omega(y, \theta_0 + \mathcal{N}_n^{-1}b, \lambda) = \frac{f(y, \theta_0 + \mathcal{N}_n^{-1}b, \lambda)dQ_\lambda(b)dJ(\lambda)}{\int_\Lambda \int_{\mathbb{R}^s} f(y, \theta_0 + \mathcal{N}_n^{-1}\tilde{b}, \tilde{\lambda})dQ_\lambda(\tilde{b})dJ(\tilde{\lambda})}. \tag{7}$$

[2] Recall the PVOT $\mathcal{P}_n^*(\alpha) \equiv \int_\Lambda I(p_n(\lambda) < \alpha)d\lambda$ uses Lebesgue measure.

If we define an expectations operator $E_{f_n(b,\lambda)}[\mathcal{Z}] \equiv \int_{\mathbb{R}^{nk}} z f(z, \theta_0 + \mathcal{N}_n^{-1}b, \lambda)dz$, then (7) and Fubini's Theorem yield:

$$\begin{aligned} E_g\left[\mathcal{P}_\xi^{(\omega)}(\mathcal{Y}_n)\right] &= \int_{\mathbb{R}^{nk}} \mathcal{P}_\xi^{(\omega)}(y)\left[\int_\Lambda \int_{\mathbb{R}^s} f(y, \theta_0 + \mathcal{N}_n^{-1}b, \lambda)dQ_\lambda(b)dJ(\lambda)\right]dy \\ &= \int_\Lambda \int_{\mathbb{R}^s}\left[\int_{\mathbb{R}^{nk}} \mathcal{P}_\xi^{(\omega)}(y)f(y, \theta_0 + \mathcal{N}_n^{-1}b, \lambda)dy\right]dQ_\lambda(b)dJ(\lambda) \\ &= \int_\Lambda \int_{\mathbb{R}^s} E_{f_n(b,\lambda)}\left[\mathcal{P}_\xi^{(\omega)}(\mathcal{Y}_n)\right]dQ_\lambda(b)dJ(\lambda). \end{aligned} \tag{8}$$

Combine (5)-(8) to deduce weighted average local power can be represented as a weighted average mean ω-PVOT, where the mean is with respect to the alternative density $f(z, \theta_0 + \mathcal{N}_n^{-1}b, \lambda)$. The above conclusions are summarized as follows.

Proposition B.1. *Weighted average local power of a test of $H_0 : \beta_0 = 0$ against $\{f(y, \theta_0 + \mathcal{N}_n^{-1}b, \lambda) : (b, \lambda) \in \mathbb{R}^s \times \Lambda\}_{n\geq 1}$ is a mean ω-PVOT. Under probability measure (7) weighted average local power is a weighted average mean ω-PVOT (8), where the mean is with respect to the alternative density $f(z, \theta_0 + \mathcal{N}_n^{-1}b, \lambda)$.*

Remark 1. By the Neyman-Pearson Lemma and Proposition B.1, the LR test has the highest weighted average mean ω-PVOT amongst asymptotic level α tests of H_0 against the sequence of simple alternatives $\{f(y, \theta_0 + \mathcal{N}_n^{-1}b, \lambda) : (b, \lambda) \in \mathbb{R}^s \times \Lambda\}_{n\geq 1}$. The result carries over to Wald and LM tests by asymptotic equivalence with the LR test.

Remark 2. The LR test must be of the form $I(f(y, \theta_0 + \mathcal{N}_n^{-1}b, \lambda)/f(y, \theta_0) > c_{n,\alpha}(b, \lambda))$ in order to rewrite weighted average power in terms of the ω-PVOT, hence we are restricted to testing H_0 against the sequence of alternatives $\{f(y, \theta_0 + \mathcal{N}_n^{-1}b, \lambda) : (b, \lambda) \in \mathbb{R}^s \times \Lambda\}_{n\geq 1}$. Evidently there does not exist a comparable result showing PVOT optimality of Andrews and Ploberger's (1994) smoothed LR test $\xi(\mathcal{Y}_n) \equiv I(\int_\Lambda \int_{\mathbb{R}^s} f(y, \theta_0 + \mathcal{N}_n^{-1}b, \lambda)dQ_\lambda(b)dJ(\lambda)/f(y, \theta_0) > c_{n,\alpha})$ of H_0 against the sequence of local alternatives $\{\int_\Lambda \int_{\mathbb{R}^s} f(y, \theta_0 + \mathcal{N}_n^{-1}b, \lambda)/f(y, \theta_0)dQ_\lambda(b)dJ(\lambda)\}_{n\geq 1}$. Logically, we cannot obtain a PVOT on Λ for a smoothed test statistic like $\xi(\mathcal{Y}_n)$, as well as average and supremum statistics: the PVOT is a fundamental entity for measuring the power of test statistics that are not smoothed on Λ, precisely by measuring how often the non-smoothed PV test rejects on Λ. By comparison, Andrews and Ploberger (1994) only treat test statistics like $\xi(\mathcal{Y}_n) \in \{0, 1\}$ which involve presmoothing over the nuisance parameter λ and drift b.

The PVOT (3) used as a test statistic obviously does not average over local alternatives, so consider a level α test $\xi(\mathcal{Y}_n, \lambda) \in \{0, 1\}$ of $H_0 : \beta_0 = 0$ against global alternatives $\{f(y, \theta_1, \lambda) : \lambda \in \Lambda\}$. The LR statistic, for example, is $\xi(\mathcal{Y}_n, \lambda) = I(f(\mathcal{Y}_n, \theta_1, \lambda)/f(\mathcal{Y}_n, \theta_0) > c_{n,\alpha}(\lambda))$. Weighted average power is simply $\int_\Lambda[\int_{\mathbb{R}^{nk}} \xi(y, \lambda)f(y, \theta_1, \lambda)dy]dJ(\lambda)$.

Define the operator $E_{f(\theta_1,\lambda)}[\mathcal{Z}] \equiv \int_{\mathbb{R}^{nk}} zf(z,\theta_1,\lambda)dz$, and define the corresponding ω-PVOT $\mathcal{P}_{\xi}^{(\omega)}(\mathcal{Y}_n) \equiv \int_{\Lambda} \xi(\mathcal{Y}_n,\lambda)d\omega(\mathcal{Y}_n,\theta_1,\lambda)$. If the probability measure in (6) is now $g(y) = \int_{\Lambda} f(y,\theta_1,\lambda)dJ(\lambda)$, then $d\omega(y,\theta_1,\lambda)$ $= f(y,\theta_1,\lambda)dJ(\lambda)/(\int_{\Lambda} f(y,\theta_1,\tilde{\lambda})dJ(\tilde{\lambda}))$ and we obtain the following result.

Corollary B.2. *Weighted average power of a test of H_0 against the simple alternative $f(y,\theta_1,\lambda)$ is identically $\int_{\Lambda} E_{f(\theta_1,\lambda)}[\mathcal{P}_{\xi}^{(\omega)}(\mathcal{Y}_n)]dJ(\lambda)$, the weighted average mean ω-PVOT, where the mean is evaluated under H_1.*

Now use Lebesgue measure $J(\lambda)$ on Λ, as in Andrews and Ploberger (1994, pp. 1384, 1395, 1398), and evaluate the joint density $f(y,\theta_1,\lambda)$ under the null $\theta_1 = \theta_0$ (hence $f(y,\theta_1,\lambda) = f(y,\theta_0)$) to yield $d\omega(y,\theta_0,\lambda)$ $= dJ(\lambda) = d\lambda$. The ω-PVOT now reduces to $\mathcal{P}_{\xi}^{(\omega)}(\mathcal{Y}_n) = \int_{\Lambda} \xi(\mathcal{Y}_n,\lambda)d\lambda$, which is simply PVOT (3). Power under the null, of course, is trivial: by construction $\int_{\mathbb{R}^{nk}} \xi(y,\lambda)f(y,\theta_0)dy = P(\xi(\mathcal{Y}_n,\lambda) = 1) = P(p_n(\lambda) < \alpha) \rightarrow \alpha$ for each λ, hence by Fubini's Theorem and bounded convergence $\int_{\Lambda} E_{f(\theta_1,\lambda)}[\mathcal{P}_{\xi}^{(\omega)}(\mathcal{Y}_n)]dJ(\lambda) \rightarrow \alpha$.

This reveals that the PVOT $\int_{\Lambda} \xi(\mathcal{Y}_n,\lambda)d\lambda$ as in (3) is just the power relevant ω-PVOT evaluated under the null with Lebesgue measure. Thus, PVOT $\int_{\Lambda} \xi(\mathcal{Y}_n,\lambda)d\lambda$ is simply a point estimate of the p-value test weighted average probability of rejection, identically $\int_{\Lambda} E_{f(\theta_1,\lambda)}[\mathcal{P}_{\xi}^{(\omega)}(\mathcal{Y}_n)]dJ(\lambda)$, evaluated under H_0. This probability is no larger than α when H_0 is in fact true, hence if the PVOT $\int_{\Lambda} \xi(\mathcal{Y}_n,\lambda)d\lambda \leq \alpha$ then we have evidence that either H_0 is correct, or global power is trivial. Conversely, $\int_{\Lambda} \xi(\mathcal{Y}_n,\lambda)d\lambda > \alpha$ for a given sample provides evidence in favor of H_1 *and* suggests global power of the PV test is non-trivial. Finally, we show below that the PVOT test is consistent if the PV test is consistent on a subset of Λ with measure greater than α, in which case, $\int_{\Lambda} \xi(\mathcal{Y}_n,\lambda)d\lambda \leq \alpha$ suggests unambiguously that the null is true.

C Theorem 3.3: Local Power and Test of Omitted Nonlinearity

C.1 Main Results

The proposed model to be tested is

$$y_t = f(x_t, \zeta_0) + e_t,$$

where ζ_0 lies in the interior of $\mathfrak{Z}$, a compact subset of $\mathbb{R}^q$, $x_t \in \mathbb{R}^k$ contains a constant term and may contain lags of y_t, and $f : \mathbb{R}^k \times \mathfrak{Z} \rightarrow \mathbb{R}$ is a known response function. Assume $\{e_t, x_t, y_t\}$ are stationary for simplicity. Let Ψ be a 1-1 bounded mapping from $\mathbb{R}^k$ to $\mathbb{R}^k$, let $\mathcal{F} : \mathbb{R} \rightarrow \mathbb{R}$ be analytic and non-polynomial (e.g. exponential or logistic), and assume $\lambda \in \Lambda$, a compact subset of $\mathbb{R}^k$. Misspecification $\sup_{\zeta \in \mathbb{R}^q} P(E[y_t|x_t] = f(x_t,\zeta)) < 1$ implies $E[e_t\mathcal{F}(\lambda'\Psi(x_t))] \neq 0 \ \forall \lambda \in \Lambda/\mathcal{S}$, where $\mathcal{S}$ has Lebesgue measure zero. See Bierens (1990), Bierens and Ploberger (1997) and Stinchcombe and White (1998) for seminal results for iid data. The test statistic for a test of the hypothesis $H_0 : E[y_t|x_t] = f(x_t,\zeta_0)$ *a.s.* is

$$\mathcal{T}_n(\lambda) = \left(\frac{1}{\hat{v}_n(\lambda)} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} e_t(\hat{\zeta}_n)\mathcal{F}\left(\lambda'\Psi(x_t)\right)\right)^2 \text{ where } e_t(\zeta) \equiv y_t - f(x_t,\zeta).$$

The estimator $\hat{\zeta}_n$ is $\sqrt{n}$-consistent of a strongly identified ζ_0, and $\hat{v}_n^2(\lambda)$ is a consistent estimator of $E[\{1/\sqrt{n}\sum_{t=1}^n e_t(\hat{\zeta}_n)\mathcal{F}(\lambda'\Psi(x_t))\}^2]$. By application of Theorem 3.3, below, under regularity conditions detailed below the asymptotic p-value is $p_n(\lambda) \equiv 1 - \bar{F}_0(\mathcal{T}_n(\lambda))$ where $\bar{F}_0$ is the $\chi^2(1)$ distribution function.

The test is asymptotically equivalent to a score test of $H_0 : \beta_0 = 0$ in the model

$$y_t = f(x_t, \zeta_0) + \beta_0 \mathcal{F}\left(\lambda'\Psi(x_t)\right) + \epsilon_t. \tag{9}$$

In view of $\sqrt{n}$-asymptotes, a sequence of local-to-null alternatives is

$$H_1^L : \beta_0 = b/n^{1/2} \text{ for } b \in \mathbb{R}. \tag{10}$$

The following regularity conditions suffice for Assumption 1. They also suffice to prove that under H_1^L for some sequence of positive finite non-random numbers $\{c(\lambda)\}$:

$$\{\mathcal{T}_n(\lambda) : \lambda \in \Lambda\} \Rightarrow^* \{(\mathcal{Z}(\lambda) + bc(\lambda))^2 : \lambda \in \Lambda\} \tag{11}$$

where $\{\mathcal{Z}(\lambda) + c(\lambda)b\}$ is a Gaussian process with mean $\{c(\lambda)b\}$, and *almost surely* uniformly continuous sample paths. See Theorem 3.3 below.

Assumption 2 (nonlinear regression and functional form test)**.**

a. Memory and Moments*: All random variables lie on the same complete measure space.* $\{y_t, x_t, \epsilon_t\}$ *are stationary;* $E|y_t|^{4+\iota} < \infty$ *and* $E|\epsilon_t|^{4+\iota}$ *for tiny* $\iota > 0$*;* $E[\epsilon_t|x_t] = 0$ *a.s. under* H_1^L*;* $E[\inf_{\lambda\in\Lambda} w_t^2(\lambda)] > 0$*,* $E[\epsilon_t^2 \inf_{\lambda\in\Lambda} w_t^2(\lambda)] > 0$*, and* $\inf_{\lambda\in\Lambda} ||(\partial/\partial\lambda)E[\epsilon_t^2 F(\lambda'\Psi(x_t))^2]|| > 0$*;* $\{x_t, \epsilon_t\}$ *are* β*-mixing with mixing coefficients* $\beta_h = O(h^{-4-\delta})$ *for tiny* $\delta > 0$*.*

b. Response Function*:* $f : \mathbb{R}^k \times \mathfrak{Z} \to \mathbb{R}$*;* $f(\cdot, \zeta)$ *is twice continuously differentiable;* $(\partial/\partial\zeta)^i f(x, \zeta)$ *are Borel measurable for each* $\zeta \in \mathfrak{Z}$ *and* $i = 0, 1, 2$*; write* $h_t(\zeta) = (\partial/\partial\zeta)^i f(x_t, \cdot)$ *for* $i = 0, 1, 2$*:* $E[\sup_{\zeta\in\mathfrak{Z}} |h_t(\zeta)|^{4+\delta}]$ $< \infty$ *for tiny* $\delta > 0$*;* $(\partial/\partial\zeta)f(x_t, \zeta_0)$ *has full column rank.*

c. Test Weight*:* $F(\cdot)$ *is analytic, nonpolynomial, and* $(\partial/\partial c)^i F(c)$ *is bounded for* $i = 0, 1, 2$ *uniformly on any compact subset;* Ψ *is one-to-one and bounded.*

d. Variance Estimator*:*

$$\hat{v}_n^2(\lambda) \equiv \frac{1}{n}\sum_{s,t=1}^n \mathcal{K}((s-t)/\gamma_n)e_s(\hat{\zeta}_n)e_t(\hat{\zeta}_n)\hat{w}_{n,s}(\lambda, \hat{\zeta}_n)\hat{w}_{n,t}(\lambda, \hat{\zeta}_n)$$

with kernel $\mathcal{K}$ *and bandwidth* $\gamma_n \to \infty$ *and* $\gamma_n = o(\sqrt{n})$*.* $\mathcal{K}$ *is continuous at* 0 *and all but a finite number of points,* $\mathcal{K} : \mathbb{R} \to [-1, 1]$*,* $\mathcal{K}(0) = 1$*,* $\mathcal{K}(x) = \mathcal{K}(-x)$ $\forall x \in \mathbb{R}$*,* $\int_{-\infty}^{\infty} |\mathcal{K}(x)|dx < \infty$*; and there exists* $\{\delta_n\}$*,* $\delta_n > 0$*,* $\delta_n/\sqrt{n} \to \infty$*, such that* $\int_{\delta_n}^{\infty}\{|\mathcal{K}(x)| + |\mathcal{K}(-x)|\}dx = o(1/\sqrt{n})$*.*

e. Plug-In*: ζ_0 is an interior point of $\mathfrak{Z}$, and $\hat{\zeta}_n \equiv \arg\min_{\zeta\in\mathfrak{Z}}\{1/n\sum_{t=1}^n(y_t - f(x,\zeta))^2\}$.*

Remark 3. The kernel variance $\hat{v}_n^2(\lambda)$ form follows from a standard expansion of $1/\sqrt{n}\sum_{t=1}^n e_t(\hat{\zeta}_n)\mathcal{F}(\lambda'\Psi(x_t))$ around ζ_0 under H_0. We exploit a kernel estimator in order to prove uniform convergence of $\hat{v}_n^2(\lambda)$ without the assumption that H_0 is true, a generality that may be of separate interest. See Lemma C.1, below.

Remark 4. Property (d), other than the requirement that $\mathcal{I}_n \equiv \int_{\delta_n}^{\infty}\{|\mathcal{K}(x)| + |\mathcal{K}(-x)|\}dx = o(1/\sqrt{n})$ for $\delta_n/\sqrt{n} \to \infty$, is similar to properties in Andrews (1991) and elsewhere, covering Bartlett, Parzen, Tukey-Hanning and Quadratic-Spectral kernels. We use $\mathcal{I}_n = o(1/\sqrt{n})$ with $\delta_n/\sqrt{n} \to \infty$ to prove uniform convergence $\sup_{\lambda\in\Lambda}|\hat{v}_n^2(\lambda) - v^2(\lambda)| \stackrel{p}{\to} 0$ under model (9) and (10). The bound $\mathcal{I}_n = o(1/\sqrt{n})$ is trivially satisfied for any $\delta_n \geq K$ and some finite $K > 0$ for Bartlett, Parzen, and Tukey-Hanning kernels, while the Quadratic-Spectral kernel obtains $\mathcal{I}_n \leq K\int_{\delta_n}^{\infty} x^{-2}dx = K\delta_n^{-3}$ hence $\mathcal{I}_n = o(1/\sqrt{n})$ for any $\delta_n/n^{1/6} \to \infty$.

Theorem 3.3.

a. Assumption 2 implies Assumption 1. In particular, under H_0 we have $\{\mathcal{T}_n(\lambda) : \lambda \in \Lambda\} \Rightarrow^* \{\mathcal{Z}(\lambda)^2 : \lambda \in \Lambda\}$ where $\{\mathcal{Z}(\lambda) : \lambda \in \Lambda\}$ is a zero mean Gaussian process with a version that has almost surely *uniformly continuous sample paths, and covariance kernel*

$$E\left[\tilde{\mathcal{Z}}_n(\lambda)\tilde{\mathcal{Z}}_n(\tilde{\lambda})\right] = \frac{E\left[\epsilon_t^2 w_t(\lambda)w_t(\tilde{\lambda})\right]}{\left(E\left[\epsilon_t^2 w_t^2(\lambda)\right] E\left[\epsilon_t^2 w_t^2(\tilde{\lambda})\right]\right)^{1/2}}. \tag{12}$$

b. Under H_1^L weak convergence (11) is valid with $c(\lambda) = E[w_t^2(\lambda)]/(E[\epsilon_t^2 w_t^2(\lambda)])^{1/2} > 0$.

C.2 Proofs

We need three preliminary results in order to prove Theorem 3.3. Define:

$$\begin{aligned} &g_t(\zeta) \equiv \frac{\partial}{\partial\zeta}f(x_t,\zeta) \quad \text{and} \quad F_t(\lambda) \equiv F(\lambda'\Psi(x_t)) \\ &\varphi_t(\lambda) \equiv F_t(\lambda)^2 - E[F_t(\lambda)g_t'] \times (E[g_t g_t'])^{-1} \times E[F_t(\lambda)g_t] \end{aligned} \tag{13}$$

$$w_t(\lambda,\zeta) \equiv F_t(\lambda) - E\left[F_t(\lambda)g_t(\zeta)'\right] \times \left(E\left[g_t(\zeta)g_t(\zeta)'\right]\right)^{-1} g_t(\zeta) \quad \text{and} \quad w_t(\lambda) = w_t(\lambda,\zeta_0)$$

$$\hat{w}_{n,t}(\lambda,\hat{\zeta}_n) \equiv F_t(\lambda) - \frac{1}{n}\sum_{s=1}^n F_s(\lambda)g_s(\hat{\zeta}_n) \times \left(\frac{1}{n}\sum_{s=1}^n g_s(\hat{\zeta}_n)g_s(\hat{\zeta}_n)'\right)^{-1} g_t(\hat{\zeta}_n)$$

$$m_t(\lambda,\zeta) \equiv \epsilon_t(\zeta)w_t(\lambda) \quad \text{and} \quad m_t(\lambda) \equiv \epsilon_t w_t(\lambda)$$

$$\mathcal{M}_t(h,\lambda) \equiv m_t(\lambda)m_{t-h}(\lambda) - E\left[m_t(\lambda)m_{t-h}(\lambda)\right]$$

$$\tilde{\mathcal{Z}}_n(\lambda) \equiv \frac{1}{v(\lambda)}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\epsilon_t w_t(\lambda) + b\frac{1}{v(\lambda)}\frac{1}{n}\sum_{t=1}^{n}\varphi_t(\lambda) \text{ and } v^2(\lambda) \equiv \lim_{n\to\infty}\frac{1}{n}E\left(\sum_{t=1}^{n}\epsilon_t w_t(\lambda)\right)^2.$$

In the first two lemmas we only exploit a subset of Assumption 2. In particular, we do not require ϵ_t in model (9) to satisfy $E[\epsilon_t|x_t] = 0$ *a.s.* We only require ϵ_t to be stationary, mixing, $E[\epsilon_t] = 0$ and $E[\epsilon_t g_t] = 0$. First, the kernel estimator $\hat{v}_n^2(\lambda)$ is uniformly consistent for $v^2(\lambda)$.

Lemma C.1 (uniform kernel variance consistency)**.** *Let Assumption 2.b-e hold. Let $\{y_t, x_t, \epsilon_t\}$ be stationary; $E|\epsilon_t|^{4+\iota} < \infty$ for tiny $\iota > 0$; $E[\epsilon_t] = 0$ and $E[\epsilon_t g_t] = 0$; $\{x_t, \epsilon_t\}$ are β-mixing with mixing coefficients $\beta_h = O(1/(h^2 \ln(h))$. Let H_1^L hold. Then $v^2(\lambda) < \infty$ and $\sup_{\lambda\in\Lambda}|\hat{v}_n^2(\lambda) - v^2(\lambda)| \xrightarrow{p} 0$.*

Remark 5. Since $E[\epsilon_t|x_t] = 0$ *a.s.* is not assumed here, it is possible that $v^2(\lambda) \neq E[\epsilon_t^2 w_t^2(\lambda)]$.

The NLLS $\hat{\zeta}_n$ is asymptotically linear, hence $1/\sqrt{n}\sum_{t=1}^{n}\epsilon_t(\hat{\zeta}_n)F_t(\lambda)$ has a simple expansion.

Lemma C.2 (**NLLS residual expansion**)**.** *Under the conditions of Lemma C.1, and $E[\epsilon_t(\partial/\partial\zeta)g_t] = 0$:*

$$\sup_{\lambda\in\Lambda}\left|\frac{1}{v(\lambda)}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\{\epsilon_t(\hat{\zeta}_n)F_t(\lambda) - \epsilon_t w_t(\lambda) - b\frac{1}{\sqrt{n}}\varphi_t(\lambda)\right\}\right| \xrightarrow{p} 0.$$

Finally, $\tilde{\mathcal{Z}}_n(\lambda)$ satisfies a functional CLT under Assumption 2.

Lemma C.3 (Functional CLT)**.** *Under Assumption 2 $\{\tilde{\mathcal{Z}}_n(\lambda)\} \Rightarrow^* \{\mathcal{Z}(\lambda) + bE[\varphi_t(\lambda)]/v(\lambda)\}$, where $\{\mathcal{Z}(\lambda)\}$ is a zero mean Gaussian process with a version that has* almost surely *uniformly continuous sample paths, and covariance kernel (12).*

Proof of Theorem 3.3. Assume $H_1^L : \beta_0 = b/n^{1/2}$ for $b \in \mathbb{R}$ is true. Assumption 2.a ensures $v^2(\lambda) = E[\epsilon_t^2 w_t^2(\lambda)] \in (0, \infty)$, and by Lemma C.1 $\sup_{\lambda\in\Lambda}|\hat{v}_n^2(\lambda) - v^2(\lambda)| \xrightarrow{p} 0$. Hence, by Lemmas C.2 and C.3:

$$\left\{\frac{1}{\hat{v}_n(\lambda)n^{1/2}}\sum_{t=1}^{n}\epsilon_t(\hat{\zeta}_n)F\left(\lambda'\Psi(x_t)\right) : \lambda \in \Lambda\right\} \Rightarrow^* \left\{\mathcal{Z}(\lambda) + \frac{b}{v(\lambda)}E\left[\varphi_t(\lambda)\right]\right\},$$

a mean $bE[\varphi_t(\lambda)]/v(\lambda)$ Gaussian process with almost surely uniformly continuous sample paths. Finally, it is easily verified that $E[\varphi_t(\lambda)] = E[w_t^2(\lambda)]$, and from above $v^2(\lambda) = E[\epsilon_t^2 w_t^2(\lambda)] \in (0, \infty)$, hence $E[\varphi_t(\lambda)]/v(\lambda) = E[w_t^2(\lambda)]/(E[\epsilon_t^2 w_t^2(\lambda)])^{1/2}$. $\mathcal{QED}$.

In order to prove kernel variance estimator uniform consistency Lemma C.1, we require a uniform central limit theorem for partial sums of $e_t(\zeta)w_t(\lambda)e_{t-h}(\zeta)w_{t-h}(\lambda)$.

Lemma C.4 (functional covariance CLT). *Let Assumption 2.b-e hold. Let $\{y_t, x_t, \epsilon_t\}$ be stationary; $E|\epsilon_t|^{4+\iota} < \infty$ for tiny $\iota > 0$; $E[\epsilon_t] = 0$; $\{x_t, \epsilon_t\}$ are β-mixing with mixing coefficients $\beta_i = O(1/(i^2 \ln(i))$. Then*

$$\left\{ \frac{1}{\sqrt{n}} \sum_{t=h+1}^{n} \mathcal{M}_t(h, \lambda) : h, \lambda \in \mathbb{N}\times\Lambda \right\} \Rightarrow^* \{\mathcal{M}(h, \lambda) : h, \lambda \in \mathbb{N}\times\Lambda\},$$

where $\{\mathcal{M}(h, \lambda)\}$ is a zero mean Gaussian process with covariance function $E[\mathcal{M}(h, \lambda)\mathcal{M}(h, \tilde{\lambda})]$, $E[\mathcal{M}(h, \lambda)^2] < \infty$, and a version that has almost surely *uniformly continuous sample paths.*

Proof.

Step 1. Consider the sub-space $L_{2,\beta}(\mathcal{P})$ of $L_2(\mathcal{P})$ endowed with the norm $||\cdot||_{2,\beta}$: see (2.5) in Doukhan, Massart, and Rio (1995). We need only prove convergence in finite dimensional distributions and tightness on $\mathcal{H} \times \Lambda$ for any compact subsets $\mathcal{H} \subset \mathbb{N}$ and $\Lambda \subset \mathbb{R}^k$ (cf. Pollard, 1990, Theorem 10.2). Convergence in finite dimensional distributions follows easily from the β-mixing property, measurability of $\mathcal{M}_t(h, \lambda)$, boundedness of F, $E[\sup_{\zeta\in\mathfrak{Z}} |g_t(\zeta)|^{4+\iota}] < \infty$ and $E|\epsilon_t|^{4+\iota} < \infty$, by exploiting a well known Cramér-Wold theorem argument and a β-mixing central limit theorem, e.g. Theorem 2.1 in Ibragimov (1975) combined with Theorem 1.a in Bradley (1993). In Step 2 we show $\{\mathcal{M}_t(h, \lambda) : h, \lambda \in \mathcal{H} \times \Lambda\}$ satisfies the metric entropy with L_2-bracketing bound $\int_0^1 | \ln(N_{[\,]}(\varepsilon, \mathcal{H}\times\Lambda, ||\cdot||_2)|^{1/2} d\varepsilon < \infty$ with L_2-bracketing numbers $N_{[\,]}(\varepsilon, \mathcal{H}\times\Lambda, ||\cdot||_{2,\beta})$. See Pollard (1984, 1990) and van der Vaart and Wellner (1996) for textbook treatments on bracketing numbers. The claim now follows from Doukhan, Massart, and Rio (1995, Theorem 1, eq. (2.17), Application 4).

Step 2. It suffices to show (see, e.g. Pollard, 1984):Doukhan, Massart, and Rio (1995, Theorem 1, eq. (2.17), Application 4)

$$\left(E\left[\left(\mathcal{M}_t(h, \lambda) - \mathcal{M}_t(\tilde{h}, \tilde{\lambda}) \right)^2 \right] \right)^{1/2} \leq K \left| h - \tilde{h} \right| + K \left\| \lambda - \tilde{\lambda} \right\|_2 \;\; \forall h, \tilde{h} \in \mathcal{H} \text{ and } \lambda, \tilde{\lambda} \in \Lambda. \tag{14}$$

By Minkowski's inequality:

$$\begin{aligned} &\left(E\left[\left(\mathcal{M}_t(h, \lambda) - \mathcal{M}_t(\tilde{h}, \tilde{\lambda}) \right)^2 \right] \right)^{1/2} \\ &\quad \leq \left(E\left[\left(m_t(\lambda) m_{t-h}(\lambda) - m_t(\tilde{\lambda}) m_{t-h}(\tilde{\lambda}) \right)^2 \right] \right)^{1/2} + \left(E\left[m_t^2(\tilde{\lambda}) \left(m_{t-h}(\tilde{\lambda}) - m_{t-\tilde{h}}(\tilde{\lambda}) \right)^2 \right] \right)^{1/2} \\ &\qquad + \left| E\left[m_t(\lambda) m_{t-h}(\lambda) \right] - E\left[m_t(\tilde{\lambda}) m_{t-h}(\tilde{\lambda}) \right] \right| + \left| E\left[m_t(\tilde{\lambda}) m_{t-h}(\tilde{\lambda}) \right] - E\left[m_t(\tilde{\lambda}) m_{t-\tilde{h}}(\tilde{\lambda}) \right] \right|. \end{aligned}$$

The mean-value-theorem, boundedness of $(\partial/\partial u)F(u)$, $E|\epsilon_t|^{4+\iota} < \infty$, and the Cauchy-Schwartz inequality imply $\forall \lambda, \tilde{\lambda} \in \Lambda$:

$$\left(E\left[\left(m_t(\lambda)m_{t-h}(\lambda) - m_t(\tilde{\lambda})m_{t-h}(\tilde{\lambda})\right)^2\right]\right)^{1/2} \leq K\left\|\lambda - \tilde{\lambda}\right\|_2 \tag{15}$$

$$\left|E\left[m_t(\lambda)m_{t-h}(\lambda)\right] - E\left[m_t(\tilde{\lambda})m_{t-h}(\tilde{\lambda})\right]\right| \leq K\left\|\lambda - \tilde{\lambda}\right\|_2.$$

Furthermore, since $\mathcal{H}$ is compact, by boundedness of F and $E|\epsilon_t|^{4+\iota} < \infty$:

$$\begin{aligned}\sup_{\tilde{h}\in\mathcal{H}}\left(E\left[m_t^2(\tilde{\lambda})\left(m_{t-h}(\tilde{\lambda}) - m_{t-\tilde{h}}(\tilde{\lambda})\right)^2\right]\right)^{1/2} &\leq \sum_{\tilde{h}\in\mathcal{H}}\left(E\left[m_t^2(\tilde{\lambda})\left(m_{t-h}(\tilde{\lambda}) - m_{t-\tilde{h}}(\tilde{\lambda})\right)^2\right]\right)^{1/2} \\ &\leq K.\end{aligned}$$

Therefore, since $\mathcal{H}$ is countable, we can always find a large enough K that satisfies

$$\left(E\left[m_t^2(\tilde{\lambda})\left(m_{t-h}(\tilde{\lambda}) - m_{t-\tilde{h}}(\tilde{\lambda})\right)^2\right]\right)^{1/2} \leq K \sup_{h\neq\tilde{h}:h,\tilde{h}\in\mathcal{H}}\left|h - \tilde{h}\right|.$$

Since trivially $E[m_t^2(\tilde{\lambda})(m_{t-h}(\tilde{\lambda}) - m_{t-\tilde{h}}(\tilde{\lambda}))^2] = 0$ for $h = \tilde{h}$, we have shown

$$\left(E\left[m_t^2(\tilde{\lambda})\left\{m_{t-h}(\tilde{\lambda}) - m_{t-\tilde{h}}(\tilde{\lambda})\right\}^2\right]\right)^{1/2} \leq K\left|h - \tilde{h}\right|.$$

Similarly, it follows

$$\left|E\left[m_t(\tilde{\lambda})m_{t-h}(\tilde{\lambda})\right] - E\left[m_t(\tilde{\lambda})m_{t-\tilde{h}}(\tilde{\lambda})\right]\right| \leq K\left|h - \tilde{h}\right|.$$

This, combined with (15), proves (14). $\mathcal{QED}$.

Proof of Lemma C.1. In the following we exploit the fact that β-mixing implies mixing in the ergodic sense, hence ergodicity (see, e.g., Petersen, 1983). Write

$$\hat{\mathcal{M}}_{n,t,h}(\lambda,\zeta) \equiv e_t(\zeta)\hat{w}_{n,t}(\lambda,\zeta)e_{t-h}(\zeta)\hat{w}_{n,t-h}(\lambda,\zeta) \text{ and } \widehat{\hat{\mathcal{M}}}_{n,h}(\lambda,\zeta) \equiv \frac{1}{n}\sum_{t=|h|+1}^{n}\hat{\mathcal{M}}_{n,t,h}(\lambda,\zeta)$$

$$\tilde{\mathcal{M}}_{t,h}(\lambda,\zeta) \equiv e_t(\zeta)w_t(\lambda,\zeta)e_{t-h}(\zeta)w_{t-h}(\lambda,\zeta) \text{ and } \widehat{\tilde{\mathcal{M}}}_h(\lambda,\zeta) \equiv \frac{1}{n}\sum_{t=|h|+1}^{n}\tilde{\mathcal{M}}_{t,h}(\lambda,\zeta)$$

$$\mathcal{M}_{t,h}(\lambda) \equiv \epsilon_t w_t(\lambda)\epsilon_{t-h}w_{t-h}(\lambda) \text{ and } \hat{\mathcal{M}}_h(\lambda) \equiv \frac{1}{n}\sum_{t=|h|+1}^{n}\mathcal{M}_{t,h}(\lambda).$$

We do not show that $\tilde{\mathcal{M}}_{t,h}(\lambda,\zeta)$ depends on n through e_t under H_1^L. By construction $\hat{v}_n^2(\lambda) =$

$\sum_{h=1-n}^{n-1} \mathcal{K}(h/\gamma_n)\widehat{\hat{\mathcal{M}}}_{n,h}(\lambda, \hat{\zeta}_n)$ and $v^2(\lambda) = \sum_{h=-\infty}^{\infty} E\left[\mathcal{M}_{t,h}(\lambda)\right]$. We have:

$$
\begin{aligned}
\hat{v}_n^2(\lambda) - v^2(\lambda) &= \sum_{h=-\infty}^{\infty} E\left[\mathcal{M}_{t,h}(\lambda)\right] - v^2(\lambda) \\
&\quad + \left\{ \sum_{h=1-n}^{n-1} \mathcal{K}(h/\gamma_n) E\left[\hat{\mathcal{M}}_h(\lambda)\right] - \sum_{h=-\infty}^{\infty} E\left[\mathcal{M}_{t,h}(\lambda)\right] \right\} \\
&\quad + \sum_{h=1-n}^{n-1} \mathcal{K}(h/\gamma_n) \left\{ \hat{\mathcal{M}}_h(\lambda) - E\left[\hat{\mathcal{M}}_h(\lambda)\right] \right\} \\
&\quad + \sum_{h=1-n}^{n-1} \mathcal{K}(h/\gamma_n) \left\{ \widehat{\tilde{\mathcal{M}}}_h(\lambda) - \hat{\mathcal{M}}_h(\lambda) \right\} \\
&\quad + \sum_{h=1-n}^{n-1} \mathcal{K}(h/\gamma_n) \left\{ \widehat{\tilde{\mathcal{M}}}_h(\lambda, \hat{\zeta}_n) - \widehat{\tilde{\mathcal{M}}}_h(\lambda) \right\} \\
&\quad + \sum_{h=1-n}^{n-1} \mathcal{K}(h/\gamma_n) \left\{ \widehat{\hat{\mathcal{M}}}_{n,h}(\lambda, \hat{\zeta}_n) - \widehat{\tilde{\mathcal{M}}}_h(\lambda, \hat{\zeta}_n) \right\} \\
&= \mathcal{A}_n(\lambda) + \mathcal{B}_n(\lambda) + \mathcal{C}_n(\lambda) + \mathcal{D}_n(\lambda) + \mathcal{E}_n(\lambda) + \mathcal{F}_n(\lambda).
\end{aligned}
$$

By construction $\mathcal{A}_n(\lambda) = 0$. We show that each remaining term converges to zero in probability, uniformly on Λ. In view of $E|\epsilon_t|^{4+\iota} < \infty$ and Assumption 2.b,c, it follows that $E(\sup_{\lambda \in \Lambda} |w_t(\lambda)|^{4+\delta}) < \infty$ for tiny $\delta > 0$, hence $\sup_{\lambda \in \Lambda} E|m_t(\lambda)|^{2+\delta} < \infty$ for some $\delta > 2$.

Step 1 ($\mathcal{B}_n(\lambda)$): Drop ζ_0. We have:

$$
\begin{aligned}
\sup_{\lambda \in \Lambda} |\mathcal{B}_n(\lambda)| &= \sup_{\lambda \in \Lambda} \left| \sum_{h=-\infty}^{\infty} E\left[\mathcal{M}_{t,h}(\lambda)\right] - \sum_{h=1-n}^{n-1} \mathcal{K}(h/\gamma_n) \frac{n-h}{n} E\left[\mathcal{M}_{t,h}(\lambda)\right] \right| \\
&\leq \sup_{\lambda \in \Lambda} \left| \sum_{h=-\infty}^{\infty} E\left[\mathcal{M}_{t,h}(\lambda)\right] - \sum_{h=1-n}^{n-1} \mathcal{K}(h/\gamma_n) E\left[\mathcal{M}_{t,h}(\lambda)\right] \right| \\
&\quad + \frac{\gamma_n}{n} \sup_{\lambda \in \Lambda} \left| \sum_{h=1-n}^{n-1} \frac{h}{\gamma_n} \mathcal{K}(h/\gamma_n) E\left[\mathcal{M}_{t,h}(\lambda)\right] \right| \\
&\leq \sup_{\lambda \in \Lambda} \left| \sum_{h=1-n}^{n-1} \left\{ E\left[\mathcal{M}_{t,h}(\lambda)\right] - \mathcal{K}(h/\gamma_n) E\left[\mathcal{M}_{t,h}(\lambda)\right] \right\} \right| + \sup_{\lambda \in \Lambda} \left| \sum_{|h| \geq n} E\left[\mathcal{M}_{t,h}(\lambda)\right] \right| \\
&\quad + \frac{\gamma_n}{n} \sup_{\lambda \in \Lambda} \left| \sum_{h=1-n}^{n-1} \frac{h}{\gamma_n} \mathcal{K}(h/\gamma_n) E\left[\mathcal{M}_{t,h}(\lambda)\right] \right|
\end{aligned}
$$

$$
\begin{aligned}
&\leq \sup_{\lambda\in\Lambda}\left|\sum_{h=1-n}^{n-1}\left\{1-\mathcal{K}(h/\gamma_n)\right\}E\left[\mathcal{M}_{t,h}(\lambda)\right]\right| + \sup_{\lambda\in\Lambda}\left|\sum_{|h|\geq n}E\left[\mathcal{M}_{t,h}(\lambda)\right]\right| \\
&\quad + \frac{\gamma_n}{n}\sup_{\lambda\in\Lambda}\left|\sum_{h=1-n}^{n-1}\frac{h}{\gamma_n}\mathcal{K}(h/\gamma_n)E\left[\mathcal{M}_{t,h}(\lambda)\right]\right| \\
&\leq \sum_{h=1-n}^{n-1}\left|1-\mathcal{K}(h/\gamma_n)\right|\sup_{\lambda\in\Lambda}\left|E\left[\mathcal{M}_{t,h}(\lambda)\right]\right| + \sum_{|h|\geq n}\sup_{\lambda\in\Lambda}\left|E\left[\mathcal{M}_{t,h}(\lambda)\right]\right| \\
&\quad + \frac{\gamma_n}{n}\sup_{\lambda\in\Lambda}\left|\sum_{h=1-n}^{n-1}\frac{h}{\gamma_n}\mathcal{K}(h/\gamma_n)E\left[\mathcal{M}_{t,h}(\lambda)\right]\right| \\
&\leq K\sup_{\lambda\in\Lambda}\left(E\left|m_t(\lambda)\right|^{2+\delta}\right)^2\sum_{h=1}^{n-1}\left|1-\mathcal{K}(h/\gamma_n)\right|\alpha_h^{\delta/(2+\delta)} \\
&\quad + K\sup_{\lambda\in\Lambda}\left(E\left|m_t(\lambda)\right|^{2+\delta}\right)^2\sum_{|h|\geq n}\alpha_h^{\delta/(2+\delta)} \\
&\quad + K\sup_{\lambda\in\Lambda}\left(E\left|m_t(\lambda)\right|^{2+\delta}\right)^2\sum_{h=1}^{n-1}\frac{h}{n}\alpha_h^{\delta/(2+\delta)},
\end{aligned}
$$

where the last inequality follows from Lemma 1.3 in Ibragimov (1962) for some $\delta > 0$ and large $K > 0$. Recall $\sup_{\lambda\in\Lambda}E|m_t(\lambda)|^{2+\delta} < \infty$ for some $\delta > 2$. Now use $\alpha_h = O(1/(h^2\ln(h))$ to deduce

$$
\sum_{h=1}^{n-1}\left|1-\mathcal{K}(h/\gamma_n)\right|\alpha_h^{\delta/(2+\delta)} \leq K\sum_{h=1}^{n-1}\left|1-\mathcal{K}(h/\gamma_n)\right|\frac{1}{h^{2\delta/(2+\delta)}\left(\ln(h)\right)^{\delta/(2+\delta)}}.
$$

Hence, $\sum_{h=1}^{n-1}|1-\mathcal{K}(h/\gamma_n)|\alpha_h^{2\delta/(2+\delta)} \to 0$ in view of $2\delta/(2+\delta) > 1$ for $\delta > 2$, and Assumption 2.d. Similarly $\sum_{|h|\geq n}\alpha_h^{\delta/(2+\delta)} \to 0$, and for tiny $\vartheta > 0$:

$$
\sum_{h=1}^{n-1}(h/n)\alpha_h^{\delta/(2+\delta)} \leq \frac{1}{n}\sum_{h=1}^{n-1}h^{1-2\delta/(2+\delta)} \leq K\frac{1}{n}\sum_{h=1}^{n-1}h^{-1-\vartheta} \to 0.
$$

Therefore $\sup_{\lambda\in\Lambda}|\mathcal{B}_n(\lambda)| \to 0$.

Step 2 ($\mathcal{C}_n(\lambda)$): Drop ζ_0. Let $\{\delta_n\}$ be a sequence of numbers, $\delta_n \in \{1,...,n-1\}$, $\delta_n = o(n)$. Then:

$$
\begin{aligned}
\sup_{\lambda\in\Lambda}|\mathcal{C}_n(\lambda)| &= \sup_{\lambda\in\Lambda}\left|\sum_{h=1-n}^{n-1}\mathcal{K}(h/\gamma_n)\left\{\hat{\mathcal{M}}_h(\lambda) - E\left[\hat{\mathcal{M}}_h(\lambda)\right]\right\}\right| \\
&= \sup_{\lambda\in\Lambda}\left|\sum_{h=1-n}^{n-1}\mathcal{K}(h/\gamma_n)\frac{1}{n}\sum_{t=|h|+1}^{n}\left\{\mathcal{M}_{t,h}(\lambda) - E\left[\mathcal{M}_{t,h}(\lambda)\right]\right\}\right|
\end{aligned}
$$

$$
\begin{aligned}
&\le \sum_{h=1-n}^{n-1} |\mathcal{K}(h/\gamma_n)| \sup_{\lambda\in\Lambda} \left| \frac{1}{n} \sum_{t=|h|+1}^{n} \{\mathcal{M}_{t,h}(\lambda) - E[\mathcal{M}_{t,h}(\lambda)]\} \right| \\
&= \sum_{|h|\le\delta_n} |\mathcal{K}(h/\gamma_n)| \sup_{\lambda\in\Lambda} \left| \frac{1}{n} \sum_{t=|h|+1}^{n} \{\mathcal{M}_{t,h}(\lambda) - E[\mathcal{M}_{t,h}(\lambda)]\} \right| \\
&\quad + \sum_{|h|=\delta_n+1}^{n-1} |\mathcal{K}(h/\gamma_n)| \sup_{\lambda\in\Lambda} \left| \frac{1}{n} \sum_{t=|h|+1}^{n} \{\mathcal{M}_{t,h}(\lambda) - E[\mathcal{M}_{t,h}(\lambda)]\} \right| \\
&= a_n + b_n,
\end{aligned}
$$

say. Consider a_n, and define

$$
\mathcal{W}_n(h/\gamma_n) \equiv \sup_{\lambda\in\Lambda} \left| \frac{1}{\sqrt{n}} \sum_{t=|h|+1}^{n} \{\mathcal{M}_{t,h}(\lambda) - E[\mathcal{M}_{t,h}(\lambda)]\} \right|.
$$

The limit process in Lemma C.4 has a version with continuous paths, hence the continuous mapping theorem applies (Dudley, 1967, 1978). We may therefore combine Lemma C.4 with the continuous mapping theorem to deduce that there exists a stochastic process $\{\mathcal{W}(x) : -\infty < x < \infty\}$ that satisfies

$$
\left\{ \sup_{h\in\{1,\ldots,\delta_n\}} \mathcal{W}_n(h/\gamma_n) \right\} \Rightarrow^* \left\{ \sup_{x\in\mathbb{R}} \mathcal{W}(x) \right\} = O_p(1).
$$

Now use $\delta_n = o(n)$, $\gamma_n = o(\sqrt{n})$, and Assumption 2.d to deduce

$$
\begin{aligned}
a_n &= \frac{1}{\gamma_n} \sum_{|h|\le\delta_n} |\mathcal{K}(h/\gamma_n)| \frac{\gamma_n}{\sqrt{n}} \sup_{\lambda\in\Lambda} \left| \frac{1}{\sqrt{n}} \sum_{t=|h|+1}^{n} \{\mathcal{M}_{t,h}(\lambda) - E[\mathcal{M}_{t,h}(\lambda)]\} \right| \\
&\le \frac{1}{\gamma_n} \sum_{|h|\le\delta_n} |\mathcal{K}(h/\gamma_n)| \frac{\gamma_n}{\sqrt{n}} \sup_{h\in\{1,\ldots,\delta_n\}} \sup_{\lambda\in\Lambda} \left| \frac{1}{\sqrt{n}} \sum_{t=|h|+1}^{n} \{\mathcal{M}_{t,h}(\lambda) - E[\mathcal{M}_{t,h}(\lambda)]\} \right| \\
&= o_p\left(\int_{-\infty}^{\infty} |\mathcal{K}(x)|\, dx \right) = o_p(1).
\end{aligned}
$$

Next, b_n. Observe that

$$
\frac{1}{\gamma_n} \sum_{|h|=\delta_n+1}^{n-1} |\mathcal{K}(h/\gamma_n)| = \frac{1}{\gamma_n} \sum_{|x_n|=\delta_n/\gamma_n+1}^{(n-1)/\gamma_n} |\mathcal{K}(x_n)|
$$

$$= \int_{\delta_n/\gamma_n}^{(n-1)/\gamma_n} |\mathcal{K}(x_n)|\, dx_n \leq K \int_{\delta_n/\gamma_n}^{\infty} |\mathcal{K}(x)|\, dx + o(1).$$

Since $\gamma_n = o(\sqrt{n})$ we can always choose $\delta_n \to \infty$ such that $\delta_n = o(n)$ and $(\delta_n/\gamma_n)/\sqrt{n} \to \infty$. Therefore, $\varepsilon_n \equiv \int_{\delta_n/\gamma_n}^{\infty} |\mathcal{K}(x)|dx = o(1/\sqrt{n})$ under Assumption 2.d. Now use boundedness of F, stationarity and the β-mixing property to deduce

$$\sup_{\lambda \in \Lambda} \left| \frac{1}{n} \sum_{t=|h|+1}^{n} \{\mathcal{M}_{t,h}(\lambda) - E[\mathcal{M}_{t,h}(\lambda)]\} \right| \leq K \left(\frac{1}{n} \sum_{t=1}^{n} |\epsilon_t \epsilon_{t-h}| + E |\epsilon_t \epsilon_{t-h}| \right) = KE |\epsilon_t \epsilon_{t-h}| + o_p(1).$$

Therefore $b_n \leq K\varepsilon_n \gamma_n = o(1)$. This proves $\sup_{\lambda \in \Lambda} |\mathcal{C}_n(\lambda)| \xrightarrow{p} 0$.

Step 3 ($\mathcal{D}_n(\lambda)$): Use $e_t = bn^{-1/2}F_t(\lambda) + \epsilon_t$ under H_1^L to deduce

$$\begin{aligned} \left| \frac{1}{\sqrt{n}} \sum_{t=|h|+1}^{n} \{e_t e_{t-h} - \epsilon_t \epsilon_{t-h}\} w_{t-h}(\lambda) w_t(\lambda) \right| &\leq |b| \frac{1}{n} \sum_{t=|h|+1}^{n} |\{\epsilon_t F_{t-h}(\lambda) + \epsilon_{t-h} F_t(\lambda)\} w_{t-h}(\lambda) w_t(\lambda)| \\ &\quad + b^2 \frac{1}{n^{3/2}} \sum_{t=|h|+1}^{n} |F_t(\lambda) F_{t-h}(\lambda)| \times |w_{t-h}(\lambda) w_t(\lambda)| \\ &\leq |b| \frac{1}{n} \sum_{t=1}^{n} |\{\epsilon_t F_{t-h}(\lambda) + \epsilon_{t-h} F_t(\lambda)\} w_{t-h}(\lambda) w_t(\lambda)| \\ &\quad + b^2 \frac{1}{n^{3/2}} \sum_{t=1}^{n} |F_t(\lambda) F_{t-h}(\lambda)| \times |w_{t-h}(\lambda) w_t(\lambda)| \\ &\equiv \mathcal{X}_n(\lambda, h), \end{aligned}$$

say. Since $|\epsilon_t F_{t-h}(\lambda) + \epsilon_{t-h} F_t(\lambda)| \times |w_{t-h}(\lambda) w_t(\lambda)|$ and $|F_t(\lambda) F_{t-h}(\lambda)| \times |w_{t-h}(\lambda) w_t(\lambda)|$ are stationary, ergodic and integrable, it follows that

$$\mathcal{X}_n(\lambda, h) \xrightarrow{p} E |\{\epsilon_t F_{t-h}(\lambda) + \epsilon_{t-h} F_t(\lambda)\} w_{t-h}(\lambda) w_t(\lambda)|.$$

Let

$$\mathcal{H}_N \equiv \{h_1, ..., h_N : h_i \in \{1, ..., n-1\}, h_i \neq h_j \forall i \neq j\} \tag{16}$$

for arbitrary $N \in \mathbb{N}$. Moreover, the tightness argument in the proof of Lemma C.4 can be easily adapted here to show $\{\mathcal{X}_n(\lambda, h) : \lambda, h \in \Lambda \times \mathcal{H}_N\}$ is tight, hence (cf. Newey, 1991, Corollary 3.1)

$$\sup_{\lambda \in \Lambda} \max_{1 \leq h \leq n-1} |\mathcal{X}_n(\lambda, h)| = \sup_{\lambda \in \Lambda} \max_{1 \leq h \leq \mathcal{H}} E |\{\epsilon_t F_{t-h}(\lambda) + \epsilon_{t-h} F_t(\lambda)\} w_{t-h}(\lambda) w_t(\lambda)| < \infty.$$

Therefore

$$|\mathcal{D}_n(\lambda)| = O_p\left(\frac{1}{\gamma_n}\sum_{h=1-n}^{n-1}|\mathcal{K}(h/\gamma_n)|\frac{\gamma_n}{\sqrt{n}}\right) = o_p\left(\frac{1}{\gamma_n}\sum_{h=1-n}^{n-1}|\mathcal{K}(h/\gamma_n)|\right) = o_p\left(\int_{-\infty}^{\infty}|\mathcal{K}(x)|\,dx\right) = o_p(1).$$

Step 4 ($\mathcal{E}_n(\lambda)$): By the mean value theorem, for some ζ_n^*, $||\zeta_n^* - \zeta_0|| \leq ||\hat{\zeta}_n - \zeta_0||$:

$$\begin{aligned}
\sup_{\lambda\in\Lambda}|\mathcal{E}_n(\lambda)| &\leq \frac{1}{\gamma_n}\sum_{h=1-n}^{n-1}|\mathcal{K}(h/\gamma_n)|\frac{\gamma_n}{\sqrt{n}}\sup_{\lambda\in\Lambda}\left|\sqrt{n}\frac{1}{n}\sum_{t=|h|+1}^{n}\left\{\tilde{\mathcal{M}}_{t,h}(\lambda,\hat{\zeta}_n) - \tilde{\mathcal{M}}_{t,h}(\lambda,\zeta_0)\right\}\right| \\
&\leq \frac{1}{\gamma_n}\sum_{h=1-n}^{n-1}|\mathcal{K}(h/\gamma_n)|\frac{\gamma_n}{\sqrt{n}}\sup_{\lambda\in\Lambda}\left\{\left\|\frac{1}{n}\sum_{t=|h|+1}^{n}\frac{\partial}{\partial\zeta}\tilde{\mathcal{M}}_{t,h}(\lambda,\zeta_n^*)\right\|\sqrt{n}\left\|\hat{\zeta}_n - \zeta_0\right\|\right\} \\
&= a_n,
\end{aligned}$$

say. By Lemma C.2.b, for some $o_p(1)$ that is not a function of λ:

$$\sqrt{n}\left(\hat{\zeta}_n - \zeta_0\right)\times(1 + o_p(1)) = \left(E\left[g_t g_t'\right]\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\epsilon_t g_t + b\times E\left[F_t(\lambda)\,g_t\right]\right) + o_p(1).$$

The β-mixing decay rate $\beta_i = O(1/(i^2\ln(i))$ combined with $E[\epsilon_t g_t] = 0$ imply $E[(1/\sqrt{n}\sum_{t=1}^{n}\epsilon_t g_t)^2] < \infty$ by application of Lemma 1.3 in Ibragimov (1962). This implies $1/\sqrt{n}\sum_{t=1}^{n}\epsilon_t g_t = O_p(1)$. Since F is bounded and g_t is integrable, is therefore follows that

$$\sup_{\lambda\in\Lambda}\sqrt{n}\left\|\hat{\zeta}_n - \zeta_0\right\| = O_p(1).$$

Further, under the maintained assumptions we can write:

$$\begin{aligned}
\sup_{\lambda\in\Lambda}\left\|\frac{1}{n}\sum_{t=|h|+1}^{n}\frac{\partial}{\partial\zeta}\mathcal{M}_{t,h}(\lambda,\zeta_n^*)\right\| &\leq K\frac{1}{n}\sum_{t=h+1}^{n+h}|e_{t-h}|\left(K + K\sup_{\zeta\in\mathfrak{Z}}\left\|\frac{\partial}{\partial\zeta}f(x_{t,},\zeta)\right\|\right) \\
&\quad + K\frac{1}{n}\sum_{t=h+1}^{n+h}|e_t|\left(K + K\sup_{\zeta\in\mathfrak{Z}}\left\|\frac{\partial}{\partial\zeta}f(x_{t-h,},\zeta)\right\|\right).
\end{aligned}$$

In view of completeness of the measure space, and measurability of $(\partial/\partial\zeta)f(x_{t,},\zeta)$ under Assumption 2.b, $\sup_{\zeta\in\mathfrak{Z}}||(\partial/\partial\zeta)f(x_{t,},\zeta)||$ is measurable with respect to $\sigma(x_{t,})$. Further, $|e_{t-h}|\sup_{\zeta\in\mathfrak{Z}}||(\partial/\partial\zeta)f(x_{t,},\zeta)||$ is integrable under Assumption 2.a,b. Therefore

$$\frac{1}{n}\sum_{t=h+1}^{n+h}|e_{t-h}|\left(K + K\sup_{\zeta\in\mathfrak{Z}}\left\|\frac{\partial}{\partial\zeta}f(x_{t,},\zeta)\right\|\right) + K\frac{1}{n}\sum_{t=h+1}^{n+h}|e_t|\left(K + K\sup_{\zeta\in\mathfrak{Z}}\left\|\frac{\partial}{\partial\zeta}f(x_{t-h,},\zeta)\right\|\right) = O_p(1).$$

Now use $\gamma_n = o(\sqrt{n})$ and Assumption 2.d to conclude

$$\begin{aligned} a_n &= \frac{1}{\gamma_n} \sum_{|h| \le \delta_n} |\mathcal{K}(h/\gamma_n)| \frac{\gamma_n}{\sqrt{n}} \sup_{\lambda \in \Lambda} \left\| \frac{1}{n} \sum_{t=|h|+1}^{n} \frac{\partial}{\partial \zeta} \mathcal{M}_{t,h}(\lambda, \zeta_n^*) \right\| \sqrt{n} \left\| \hat{\zeta}_n - \zeta_0 \right\| \\ &= o_p \left(\frac{1}{\gamma_n} \sum_{|h| \le \delta_n} |\mathcal{K}(h/\gamma_n)| \right) = o_p(1). \end{aligned}$$

Step 5 ($\mathcal{F}_n(\lambda)$): By the same arguments used in Step 2 of the proof of Lemma C.4, the sequence of distributions governing $\{1/\sqrt{n} \sum_{t=1}^{n} (F_t(\lambda) g_t(\zeta_0) - E[F_t(\lambda) g_t(\zeta_0)]) : \lambda \in \Lambda\}$ can be shown to be tight. Therefore, by the arguments in the proof of Lemma C.4 that lead to the uniform CLT there, it similarly follows here that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left\{ g_t(\zeta_0) g_t(\zeta_0)' - E\left[g_t(\zeta_0) g_t(\zeta_0)' \right] \right\} = O_p(1) \tag{17}$$

$$\sup_{\lambda \in \Lambda} \left\| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} F_t(\lambda) g_t(\zeta_0) - E\left[F_t(\lambda) g_t(\zeta_0) \right] \right\| = O_p(1).$$

We have:

$$\begin{aligned} &\sup_{\lambda \in \Lambda} |\mathcal{F}_n(\lambda)| \\ &\quad \le \frac{1}{\gamma_n} \sum_{h=1-n}^{n-1} |\mathcal{K}(h/\gamma_n)| \, \gamma_n \left\{ \frac{1}{n} \sum_{t=1}^{n} \left| e_t(\hat{\zeta}_n) e_{t-h}(\hat{\zeta}_n) \right| \right. \\ &\qquad\qquad \left. \times \sup_{\lambda \in \Lambda} \left| \hat{w}_{n,t}(\lambda, \hat{\zeta}_n) \hat{w}_{n,t-h}(\lambda, \hat{\zeta}_n) - w_t(\lambda, \hat{\zeta}_n) w_{t-h}(\lambda, \hat{\zeta}_n) \right| \right\}. \end{aligned} \tag{18}$$

Define

$$\mathcal{F}(\lambda, \zeta) = E\left[F_t(\lambda) g_t(\zeta)' \right] \times \left(E\left[g_t(\zeta) g_t(\zeta)' \right] \right)^{-1}$$

$$\hat{\mathcal{F}}_n(\lambda, \zeta) = \frac{1}{n} \sum_{t=1}^{n} F_t(\lambda) g_t(\zeta) \times \left(\frac{1}{n} \sum_{t=1}^{n} g_t(\zeta) g_t(\zeta)' \right)^{-1}.$$

Re-arrange terms in (18) to deduce:

$$\sup_{\lambda \in \Lambda} \left| \hat{w}_{n,t}(\lambda, \hat{\zeta}_n) \hat{w}_{n,t-h}(\lambda, \hat{\zeta}_n) - w_t(\lambda, \hat{\zeta}_n) w_{t-h}(\lambda, \hat{\zeta}_n) \right|$$

$$\leq \sup_{\lambda\in\Lambda}\left|\hat{\mathcal{F}}_n(\lambda,\zeta)-\mathcal{F}(\lambda,\zeta)\right|\times\left(\left|g_t(\hat{\zeta}_n)\right|+\left|g_{t-h}(\hat{\zeta}_n)\right|\right)$$
$$+\sup_{\lambda\in\Lambda}\left|\hat{\mathcal{F}}_n(\lambda,)\hat{\mathcal{F}}_n(\lambda,\zeta)-\mathcal{F}(\zeta)\mathcal{F}(\zeta)\right|\left|g_t(\hat{\zeta}_n)g_{t-h}(\hat{\zeta}_n)\right|.$$

Now combine (17) and (18):

$$\sup_{\lambda\in\Lambda}|\mathcal{F}_n(\lambda)| = O_p\left(\frac{1}{\gamma_n}\sum_{h=1-n}^{n-1}\left\{|\mathcal{K}(h/\gamma_n)|\frac{\gamma_n}{\sqrt{n}}\right.\right. \tag{19}$$
$$\left.\left.\times\frac{1}{n}\sum_{t=1}^{n}\left|e_t(\hat{\zeta}_n)e_{t-h}(\hat{\zeta}_n)\right|\left\{\left|g_t(\hat{\zeta}_n)\right|+\left|g_{t-h}(\hat{\zeta}_n)\right|+\left|g_t(\hat{\zeta}_n)g_{t-h}(\hat{\zeta}_n)\right|\right\}\right\}\right).$$

Let $\mathcal{H}_N$ be the set in (16) for arbitrary $N\in\mathbb{N}$. The argument used to prove Step 4, combined with moment and memory properties under Assumption 2, and the proof of Lemma C.2, can be adapted to show for some process $\{\mathcal{X}_n(\lambda,h):\lambda,h\in\Lambda\times\mathcal{H}_N\}$, $\sup_{\lambda,h\in\Lambda\times\mathcal{H}_N}||\mathcal{X}_n(\lambda,h)||=O_p(1)$:

$$\sup_{\lambda,h\in\Lambda\times\mathcal{H}_N}\left\|\frac{1}{n}\sum_{t=1}^{n}\left|e_t(\hat{\zeta}_n)e_{t-h}(\hat{\zeta}_n)\right|\left\{\left|g_t(\hat{\zeta}_n)\right|+\left|g_{t-h}(\hat{\zeta}_n)\right|+\left|g_t(\hat{\zeta}_n)g_{t-h}(\hat{\zeta}_n)\right|\right\}-\mathcal{X}_n(\lambda,h)\right\|\xrightarrow{p}0,$$

hence

$$\sup_{\lambda,h\in\Lambda\times\mathcal{H}_N}\frac{1}{n}\sum_{t=1}^{n}\left|e_t(\hat{\zeta}_n)e_{t-h}(\hat{\zeta}_n)\right|\left\{\left|g_t(\hat{\zeta}_n)\right|+\left|g_{t-h}(\hat{\zeta}_n)\right|+\left|g_t(\hat{\zeta}_n)g_{t-h}(\hat{\zeta}_n)\right|\right\}=O_p(1). \tag{20}$$

In view of integrability $\int_{-\infty}^{\infty}|\mathcal{K}(x)|\,dx<\infty$, it now follows by dominated convergence that

$$\frac{1}{\gamma_n}\sum_{h=1-n}^{n-1}|\mathcal{K}(h/\gamma_n)|\frac{1}{n}\sum_{t=1}^{n}\left|e_t(\hat{\zeta}_n)e_{t-h}(\hat{\zeta}_n)\right|\left\{\left|g_t(\hat{\zeta}_n)\right|+\left|g_{t-h}(\hat{\zeta}_n)\right|+\left|g_t(\hat{\zeta}_n)g_{t-h}(\hat{\zeta}_n)\right|\right\} \tag{21}$$
$$\leq K\int_{-\infty}^{\infty}|\mathcal{K}(x)|\,dx+o_p(1).$$

Finally, use $\gamma_n=o(\sqrt{n})$ with (19)-(21) to conclude $\sup_{\lambda\in\Lambda}|\mathcal{F}_n(\lambda)|=o_p(1)$. The completes the proof. $\mathcal{QED}$.

The proof of Lemma C.2 follows from consistency and asymptotic linearity of the NLLS estimator.

Lemma C.5 (NLLS consistency and linearity)**.** *Let Assumption 2.b-e hold. Let $\{y_t,x_t,\epsilon_t\}$ be stationary; $E|\epsilon_t|^{4+\iota}<\infty$ for tiny $\iota>0$; $E[\epsilon_t]=0$, $E[\epsilon_t g_t]=0$ and $E[\epsilon_t(\partial/\partial\zeta)g_t]=0$; $\{x_t,\epsilon_t\}$ are β-mixing with mixing coefficients $\beta_i=O(1/(i^2\ln(i))$. Then:*

a. $\hat{\zeta}_n\xrightarrow{p}\zeta_0$

$$b. \quad \sup_{\lambda \in \Lambda} \left\| \sqrt{n} \left(\hat{\zeta}_n - \zeta_0 \right) - \left(E \left[g_t g_t' \right] \right)^{-1} \left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \epsilon_t g_t + b \times E \left[F_t(\lambda) g_t \right] \right) \right\| \xrightarrow{p} 0.$$

Proof. We first derive some required asymptotic properties. Recall β-mixing implies ergodicity. We have:

$$\sup_{\zeta \in Z} \left\| \frac{1}{n} \sum_{t=1}^{n} \epsilon_t(\zeta) g_t(\zeta) - E \left[\epsilon_t(\zeta) g_t(\zeta) \right] \right\| \xrightarrow{p} 0 \tag{22}$$

$$\sup_{\zeta \in Z} \left\| \frac{1}{n} \sum_{t=1}^{n} \epsilon_t(\zeta) \frac{\partial}{\partial \zeta} g_t(\zeta) - E \left[\epsilon_t(\zeta) \frac{\partial}{\partial \zeta} g_t(\zeta) \right] \right\| \xrightarrow{p} 0 \tag{23}$$

$$\sup_{\zeta \in \mathcal{Z}} \left\| \frac{1}{n} \sum_{t=1}^{n} g_t(\zeta) g_t(\zeta)' - E \left[g_t(\zeta) g_t(\zeta)' \right] \right\| \xrightarrow{p} 0 \tag{24}$$

$$\sup_{\lambda \in \Lambda, \zeta \in \mathcal{Z}} \left\| \frac{1}{n} \sum_{t=1}^{n} F_t(\lambda) g_t(\zeta) - E \left[F_t(\lambda) g_t(\zeta) \right] \right\| \xrightarrow{p} 0 \tag{25}$$

$$\sup_{\lambda \in \Lambda} \left| \frac{1}{n} \sum_{t=1}^{n} F_t(\lambda)^2 - E \left[F_t(\lambda)^2 \right] \right| \xrightarrow{p} 0. \tag{26}$$

Consider (22), and write $w_t(\zeta) \equiv \epsilon_t(\zeta) g_t(\zeta) - E[\epsilon_t(\zeta) g_t(\zeta)]$. Pointwise convergence $1/n \sum_{t=1}^{n} w_t(\zeta) \xrightarrow{p} 0$ holds in view of stationarity, ergodicity and integrability $E|w_t(\zeta)| < \infty$. It remains to verify stochastic equicontinuity (see, e.g., Newey, 1991, Theorem 2.1), which holds if $||w_t(\zeta) - w_t(\tilde{\zeta})||_2 \leq K||\zeta - \tilde{\zeta}||$ (cf., Newey, 1991, Corollary 3.1). By the mean value theorem, Jensen and Minkowski inequalities, and the assumed moment bounds:

$$\begin{aligned} \left\| w_t(\zeta) - w_t(\tilde{\zeta}) \right\|_2 &\leq 2 \left\| \epsilon_t(\zeta) g_t(\zeta) - \epsilon_t(\tilde{\zeta}) g_t(\tilde{\zeta}) \right\|_2 \\ &= 2 \sup_{\zeta \in \mathcal{Z}} \left\| \frac{\partial}{\partial \zeta} \epsilon_t(\zeta) g_t(\zeta) + \epsilon_t(\zeta) \frac{\partial}{\partial \zeta} g_t(\zeta) \right\|_2 \times \left\| \zeta - \tilde{\zeta} \right\| \leq K \left\| \zeta - \tilde{\zeta} \right\|. \end{aligned}$$

A similar argument applies to each remaining claim.

Claim (a). Use (22), and $1/n \sum_{t=1}^{n} \epsilon_t(\hat{\zeta}_n) g_t(\hat{\zeta}_n) = 0$ by construction of $\hat{\zeta}_n$, to deduce

$$\begin{aligned} o_p(1) &= \sup_{\zeta \in Z} \left\| \frac{1}{n} \sum_{t=1}^{n} \epsilon_t(\zeta) g_t(\zeta) - E \left[\epsilon_t(\zeta) g_t(\zeta) \right] \right\| \\ &\geq \left\| \frac{1}{n} \sum_{t=1}^{n} \epsilon_t(\hat{\zeta}_n) g_t(\hat{\zeta}_n) - E \left[\epsilon_t(\hat{\zeta}_n) g_t(\hat{\zeta}_n) \right] \right\| = \left\| E \left[\epsilon_t(\hat{\zeta}_n) g_t(\hat{\zeta}_n) \right] \right\|. \end{aligned}$$

Therefore $||E[\epsilon_t(\hat{\zeta}_n)g_t(\hat{\zeta}_n)]|| \rightarrow 0$. Now invoke continuity of $E[\epsilon_t(\zeta)g_t(\zeta)]$, and $E[\epsilon_t g_t] = 0$ by supposition, to conclude $||\hat{\zeta}_n - \zeta_0|| \xrightarrow{p} 0$.

Claim (b). By the least squares first order condition, and the mean value theorem, for some ζ_n^*, $||\zeta_n^* - \zeta_0|| \leq ||\hat{\zeta}_n - \zeta_0||$:

$$\begin{aligned} 0 &= \frac{1}{n}\sum_{t=1}^{n} \epsilon_t(\hat{\zeta}_n)g_t(\hat{\zeta}_n) \\ &= \frac{1}{n}\sum_{t=1}^{n}\left(\epsilon_t + b\frac{1}{\sqrt{n}}F_t(\lambda)\right)g_t + \left\{-\frac{1}{n}\sum_{t=1}^{n} g_t(\zeta_n^*)g_t(\zeta_n^*)' + \frac{1}{n}\sum_{t=1}^{n}\epsilon_t(\zeta_n^*)\frac{\partial}{\partial\zeta}g_t(\zeta_n^*)\right\}\left(\hat{\zeta}_n - \zeta_0\right). \end{aligned}$$

Uniform laws (22) and (24), consistency $\hat{\zeta}_n \xrightarrow{p} \zeta_0$, and continuity imply

$$\frac{1}{n}\sum_{t=1}^{n}\epsilon_t(\zeta_n^*)(\partial/\partial\zeta)g_t(\zeta_n^*) \xrightarrow{p} E\left[\epsilon_t\frac{\partial}{\partial\zeta}g_t\right] = 0 \text{ and } \frac{1}{n}\sum_{t=1}^{n}g_t(\zeta_n^*)g_t(\zeta_n^*)' \xrightarrow{p} E[g_t g_t'].$$

Therefore:

$$\sqrt{n}\left(\hat{\zeta}_n - \zeta_0\right) = \left(E\left[g_t g_t'\right]\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\epsilon_t g_t + b\frac{1}{n}\sum_{t=1}^{n}F_t(\lambda)g_t\right)(1 + o_p(1))$$

where $o_p(1)$ is not a function of λ. Now use limit (25) to complete the proof. $\mathcal{QED}$.

Proof of Lemma C.2. Expand $1/\sqrt{n}\sum_{t=1}^{n}\epsilon_t(\hat{\zeta}_n)F_t(\lambda)$ based on the mean-value-theorem to deduce

$$\begin{aligned} \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\epsilon_t(\hat{\zeta}_n)F_t(\lambda) &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(y_t - f(x_t,\zeta_0)\right)F_t(\lambda) - \left\{\frac{1}{n}\sum_{t=1}^{n}F_t(\lambda)g_t(\zeta_n^*)\right\}'\sqrt{n}\left(\hat{\zeta}_n - \zeta_0\right) \\ &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\epsilon_t - b\frac{1}{\sqrt{n}}F_t(\lambda)\right)F_t(\lambda) - \left\{\frac{1}{n}\sum_{t=1}^{n}F_t(\lambda)g_t(\zeta_n^*)\right\}'\sqrt{n}\left(\hat{\zeta}_n - \zeta_0\right), \end{aligned}$$

where $||\zeta_n^* - \zeta_0|| \leq ||\hat{\zeta}_n - \zeta_0||$. ULLN (25), consistency Lemma C.5.a and continuity imply $\sup_{\lambda\in\Lambda}||1/n\sum_{t=1}^{n}F_t(\lambda)g_t(\zeta_n^*) - E[F_t(\lambda)g_t]|| \xrightarrow{p} 0$. The claim now follows by invoking asymptotic linearity Lemma C.5.b, ULLN (26), and re-arranging terms. $\mathcal{QED}$.

Proof of Lemma C.3. Write

$$\tilde{\mathcal{Z}}_n(\lambda) \equiv \frac{1}{v(\lambda)}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\epsilon_t w_t(\lambda) + b\frac{1}{v(\lambda)}\frac{1}{n}\sum_{t=1}^{n}\varphi_t(\lambda) = \mathcal{Z}_n(\lambda) + \mathcal{A}_n(\lambda)$$

$$\mathcal{A}(\lambda) \equiv b\frac{1}{v(\lambda)}E\left[\varphi_t(\lambda)\right]$$

Under Assumption 2 $v^2(\lambda) = E[\epsilon_t^2 w_t^2(\lambda)] > 0$, and $E[\mathcal{Z}_n(\lambda)\mathcal{Z}_n(\tilde{\lambda})]$ is identically the covariance kernel (12). It suffices to prove $\{\mathcal{Z}_n(\lambda)\} \Rightarrow^* \{\mathcal{Z}(\lambda)\}$ and $\sup_\lambda ||\mathcal{A}_n(\lambda) - E[\mathcal{A}(\lambda)]|| \xrightarrow{p} 0$.

Step 1 ($\mathcal{Z}_n(\lambda)$). The proof follows the lines of the proof of Lemma C.4. Convergence in the finite dimensional distributions of $\{\mathcal{Z}_n(\lambda)\}$ follows from Assumption 2.a,b, and, for example, Theorem 2.1 in Ibragimov (1975) combined with Theorem 1.a in Bradley (1993). The claim then follows from Doukhan, Massart, and Rio (1995, Theorem 1, eq. (2.17), Application 4) if we demonstrate (see, e.g. Pollard, 1984):

$$\left(E\left[\left(\mathcal{Z}_n(\lambda) - \mathcal{Z}_n(\tilde{\lambda})\right)^2\right]\right)^{1/2} \leq K\left\|\lambda - \tilde{\lambda}\right\|_2 \ \forall \lambda, \tilde{\lambda} \in \Lambda. \tag{27}$$

In view of $E[\epsilon_t|x_t] = 0$ *a.s.* and stationarity:

$$E\left[\left(\mathcal{Z}_n(\lambda) - \mathcal{Z}_n(\tilde{\lambda})\right)^2\right] = E\left[\epsilon_t^2\left\{\frac{w_t(\lambda)}{v(\lambda)} - \frac{w_t(\tilde{\lambda})}{v(\tilde{\lambda})}\right\}^2\right]. \tag{28}$$

By the mean-value-theorem, for some λ^*, $||\lambda^* - \lambda|| \leq ||\lambda - \tilde{\lambda}||$:

$$\begin{aligned} E\left[\epsilon_t^2\left\{\frac{w_t(\lambda)}{v(\lambda)} - \frac{w_t(\tilde{\lambda})}{v(\tilde{\lambda})}\right\}^2\right] &= E\left[\epsilon_t^2\left\{\frac{\partial}{\partial \lambda'}\frac{w_t(\lambda^*)}{v(\lambda^*)}\left(\lambda - \tilde{\lambda}\right)\right\}^2\right] \\ &= E\left[\epsilon_t^2\left\{\sum_{i=1}^k \frac{\partial}{\partial \lambda_i}\frac{w_t(\lambda^*)}{v(\lambda^*)}\left(\lambda_i - \tilde{\lambda}_i\right)\right\}^2\right] \\ &\leq \sup_{\lambda \in \Lambda} E\left[\epsilon_t^2 \sum_{i=1}^k \left(\frac{\partial}{\partial \lambda_i}\frac{w_t(\lambda)}{v(\lambda)}\right)^2\right] \sum_{i=1}^k \left(\lambda_i - \tilde{\lambda}_i\right)^2. \end{aligned} \tag{29}$$

The last line exploits the the Cauchy-Schwartz inequality. Since $\inf_{\lambda \in \Lambda} v(\lambda)^2 > 0$, and $F(u)$ and $(\partial/\partial u)F(u)$ are uniformly bounded under Assumption 2, it follows by the moment bounds of Assumption 2 that:

$$\begin{aligned} &\sup_{\lambda \in \Lambda}\left(E\left[\epsilon_t^2 \sum_{i=1}^k \left(\frac{\partial}{\partial \lambda_i}\frac{w_t(\lambda)}{v(\lambda)}\right)^2\right]\right)^{1/2} \\ &= \sup_{\lambda \in \Lambda}\frac{1}{v(\lambda)}\left(E\left[\epsilon_t^2 \sum_{i=1}^k \left(\frac{\partial}{\partial \lambda_i}w_t(\lambda) - w_t(\lambda)\frac{1}{v(\lambda)}\frac{\partial}{\partial \lambda_i}v(\lambda)\right)^2\right]\right)^{1/2} \\ &\leq K \sup_{\lambda \in \Lambda}\left(E\left[\epsilon_t^2 \sum_{i=1}^k \left(\left|\frac{\partial}{\partial \lambda_i}w_t(\lambda)\right| + K\left|w_t(\lambda)\right|\left|\frac{\partial}{\partial \lambda_i}v(\lambda)\right|\right)^2\right]\right)^{1/2} \leq K. \end{aligned} \tag{30}$$

Combine (28)-(30) to deduce (27).

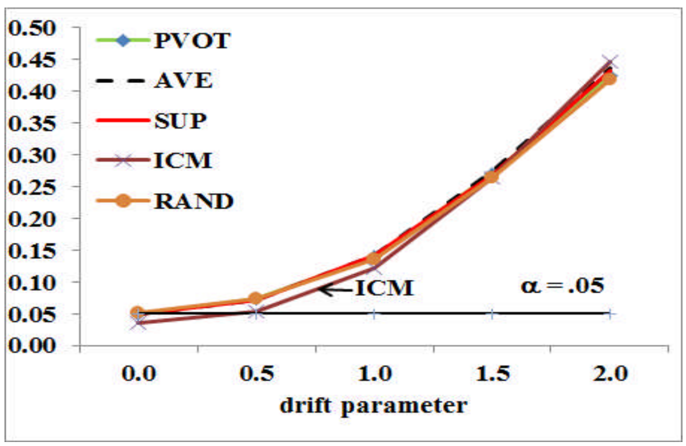
Step 2 ($\mathcal{A}_n(\lambda)$). Since $\inf_{\lambda \in \Lambda} v(\lambda)^2 > 0$, and $b \in \mathbb{R}$ is a constant, it follows:

$$\sup_{\lambda \in \Lambda} \left| b \frac{1}{v(\lambda)} \frac{1}{n} \sum_{t=1}^{n} (\varphi_t(\lambda) - E[\varphi_t(\lambda)]) \right| \leq K \sup_{\lambda \in \Lambda} \left| \frac{1}{n} \sum_{t=1}^{n} (\varphi_t(\lambda) - E[\varphi_t(\lambda)]) \right|.$$

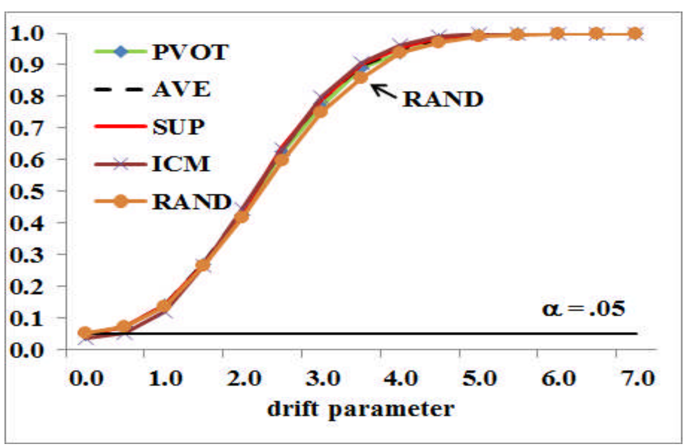
The required uniform law follows by adapting the proof of (22) in the proof of Lemma C.2. $\mathcal{QED}$.

D Figures

Figure D.1: Local Power for PVOT, Randomized, Average, Supremum and ICM Tests of Omitted Nonlinearity : null model is $y_t = \beta_0 x_t + \epsilon_t$



(a) Local power over drift $b \in [0, 2]$



(b) Local power over drift $b \in [0, 7]$

E Bibliography

Andrews, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858.

Andrews, D. W. K., and W. Ploberger (1994): "Optimal Tests when a Nuisance Parameter is Present Only under the Alternative," *Econometrica*, 62, 1383–1414.

Bierens, H. J. (1990): "A Consistent Conditional Moment Test of Functional Form," *Econometrica*, pp. 1443–1458.

Bierens, H. J., and W. Ploberger (1997): "Asymptotic Theory for Integrated Conditional Moment Tests," *Econometrica*, 65, 1129–1151.

Bradley, R. (1993): "Equivalent Mixing Conditions for Random Fields," *Annals of Probability*, 21, 1921–1926.

Doukhan, P., P. Massart, and E. Rio (1995): "Invariance Principles for Absolutely Regular Empirical Processes," *Annales de l'Institut Henri Poincare B*, 31, 393–427.

Dudley, R. M. (1967): "The Sizes of Compact Subsets of Hilbert Space and Continuity of Gaussian Processes," *Journal of Functional Analysis*, 1, 290–330.

——— (1978): "Central Limit Theorems for Empirical Measures," *Annals of Probability*, 6, 899–929.

Hill, J. B., and M. Aguilar (2013): "Moment Condition Tests for Heavy Tailed Time Series," *Journal of Econometrics*, 172, 255–274.

Hoffman-Jϕrgensen, J. (1991): "Convergence of Stochastic Processes on Polish Spaces," in *Various Publications Series 39*. Aarhus Universitet, Aarhus, Denmark, mimeo.

Ibragimov, I. (1962): "Some Limit Theorems for Stationary Processes," *Theory of Probability and its Applications*, 7, 349–382.

——— (1975): "A Note on the Central Limit Theorem for Dependent Random Variables," *Theory of Probability and its Applications*, 20, 135–141.

Newey, W. K. (1991): "Uniform Convergence in Probability and Stochastic Equicontinuity," *Econometrica*, 59, 1161–1167.

Petersen, K. (1983): *Ergodic Theory.* Cambridge Univ. Press.

Pollard, D. (1984): *Convergence of Stochastic Processes.* Springer, New York.

——— (1990): "Empirical Processes: Theory and Applications," *Regional Conference Series in Probability and Statistics*, 2.

Stinchcombe, M., and H. White (1998): "Consistent Specification Testing with Nuisance Parameters Present Only Under the Alternative," *Econometric Theory*, 14, 295–325.

van der Vaart, A., and J. Wellner (1996): *Weak Convergence and Empirical Processes.* Springer, New York.