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.

Appendix D presents the proof of Theorem 4.4 which concerns the PVOT structural break test. Finally, Appendix E contains omitted figures from the main paper.

Recall how the PVOT is constructed. We assume $T_n(\lambda) \geq 0$, and that large values are indicative of $H_1$. Let $p_n(\lambda)$ be a p-value or asymptotic p-value based on $T_n(\lambda)$: $p_n(\lambda)$ 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 $\Lambda$:

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

The p-value [PV] test with nominal level $\alpha$ for a chosen value of $\lambda$ is (1):

**PV Test:** reject $H_0$ if $p_n(\lambda) < \alpha$, otherwise fail to reject $H_0$. \hfill (2)

Now assume $\Lambda$ 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 $\alpha \in (0, 1)$:

**PVOT:** $P_n^*(\alpha) \equiv \int_{\Lambda} I (p_n(\lambda) < \alpha) d\lambda, \quad (3)$

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

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

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

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

b. Under $H_1^L$ weak convergence (11) is valid with $c(\lambda) = E[w_t^2(\lambda)]/(E[e_t^2 w_t^2(\lambda)])^{1/2} > 0$ 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. \( \lfloor z \rfloor \) 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. \( \epsilon > 0 \) 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 \( \lambda \). Thus, in this section \( \lambda \) is assumed to be part of the data generating process. Assume \( f(y, \theta, \lambda) > 0 \) almost everywhere on \( S \times \Theta \times \Lambda \), for some subset \( S \subseteq \mathbb{R}^{nk} \), \( \Theta \) is a compact subset of \( \mathbb{R}^s \) containing \( \theta_0 \), 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 \( \lambda \) is part of the data generating process only under \( H_1 \). Consider a sequence of local alternatives \( H_{0,\lambda}^l \) 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 \( \alpha \) test of \( H_0 \) for some imputed \( (b, \lambda) \), and as in Andrews and Ploberger (1994) let \( Q_\lambda \) be for each \( \lambda \) be an absolutely continuous probability measure and \( \mathbb{R}^s \), and let \( J \) be an integrable weight function on \( \Lambda \). For example, the LR statistic is \( \xi(\mathcal{Y}_n, b, \lambda) = I(f(\mathcal{Y}_n, \theta_0 + \mathcal{N}_n^{-1}b, \lambda) > c_{n,\alpha}(b, \lambda)) \) where \( c_{n,\alpha}(b, \lambda) \) is the asymptotic level \( \alpha \) critical value, hence \( E[\xi(\mathcal{Y}_n, b, \lambda)] \to \alpha \) under \( H_0 \). Andrews and Ploberger (1994) require \( Q_\lambda \) to be a Gaussian density that depends on \( \lambda \) in order to show that their exp-LM statistic is optimal. We allow \( Q_\lambda \) to depend on \( \lambda \) 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 \( \{f(y, \theta_0 + \mathcal{N}_n^{-1}b, \lambda) : (b, \lambda) \in \mathbb{R}^s \times \Lambda\}_{n \geq 1} \)

---

\(^1\) 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 \( \Sigma_\lambda \) that depends on \( \lambda \). They also use Lebesgue measure \( J \) for the weight on \( \Lambda \) in their simulations as a default tactic when information about the true \( \lambda \) under \( H_1 \) is not available.
has weighted average local power (cf. Andrews and Ploberger, 1994)

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

Now let \( g(y) \) be any joint probability measure that is positive on \( \mathbb{R}^{nk} \) \( a.e. \), define the expectations operator \( E_g[m(\mathcal{Z})] \equiv \int_{\mathbb{R}^{nk}} m(z) g(z) \, dz \) for an arbitrary scalar mapping \( m \), and define:

\[
d\omega(y, \theta_0 + N_n^{-1}b, \lambda) \equiv \frac{f(y, \theta_0 + N_n^{-1}b, \lambda)}{g(y)} dQ_\lambda(b) dJ(\lambda).
\]

We do not require \( d\omega(y, \theta_0 + N_n^{-1}b, \lambda) \) 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 \( E_g \):

\[
\int_{\Lambda} \int_{\mathbb{R}^s} \left[ \int_{\mathbb{R}^{nk}} \xi(y, b, \lambda) f(y, \theta_0 + N_n^{-1}b, \lambda) \, dy \right] dQ_\lambda(b) dJ(\lambda)
= \int_{\mathbb{R}^{nk}} \left[ \int_{\Lambda} \int_{\mathbb{R}^s} \xi(y, b, \lambda) f(y, \theta_0 + N_n^{-1}b, \lambda) \, dy \right] g(y) dJ(\lambda)
= E_g \left[ \int_{\mathbb{R}^{nk}} \xi(y, b, \lambda) d\omega(y, \theta_0 + N_n^{-1}b, \lambda) \right].
\]

We will call the above integral under expectations,

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

the \( \omega \)-PVOT since it gives the \( \omega \) measure of the subset of local drift \( b \) and nuisance parameter \( \lambda \) on which a test based on \( \xi(\mathcal{Y}_n, b, \lambda) \) rejects \( H_0 \) in favor of \( f(y, \theta_0 + N_n^{-1}b, \lambda) \). 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 \( \lambda \) and local drift \( b \), and \( d\omega \) need not be Lebesgue measure.\(^2\) Weighted average local power is therefore a mean \( \omega \)-PVOT:

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

The \( \omega \)-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 \((J, Q_\lambda)\), if and only if it has the highest mean \( \omega \)-PVOT. This seems natural since relative to all other tests an optimal test should

\(^2\)Recall the PVOT \( \mathcal{P}_\pi^{(\alpha)}(\alpha) \equiv \int_{\Lambda} I(p_n(\lambda) < \alpha) \, d\lambda \) uses Lebesgue measure.
spend more time in the rejection region, over the nuisance parameter $\lambda$ and local drift $b$.

As a special case, the probability measure

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

(6)

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

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

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

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

(8)

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

$$= \int_{\Lambda} \int_{\mathbb{R}^s} E_{f_n(b,\lambda)} \left[ \mathcal{P}^{(\omega)}_\xi (\mathcal{Y}_n) \right] dQ_\lambda(b) dJ(\lambda).$$

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

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

**Remark 1.** By the Neyman-Pearson Lemma and Proposition B.1, the LR test has the highest weighted average mean $\omega$-PVOT amongst asymptotic level $\alpha$ tests of $H_0$ against the sequence of simple alternatives $\{f(y, \theta_0 + N_n^{-1} b, \lambda) : (b, \lambda) \in \mathbb{R}^s \times \Lambda\}_{n \geq 1}$. 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 $I(f(y, \theta_0 + N_n^{-1} b, \lambda)/f(y, \theta_0) > c_{n,\alpha}(b, \lambda))$ in order to rewrite weighted average power in terms of the $\omega$-PVOT, hence we are restricted to testing $H_0$ against the sequence of alternatives $\{f(y, \theta_0 + N_n^{-1} b, \lambda) : (b, \lambda) \in \mathbb{R}^s \times \Lambda\}_{n \geq 1}$. Evidently there does not exist a comparable result showing PVOT optimality of Andrews and Ploberger’s (1994)
smoothed LR test $\xi(Y_n) \equiv I(\int_A \int_{R^b} f(y, \theta_0 + N_n^{-1}b, \lambda)dQ_\lambda(b)dJ(\lambda)/f(y, \theta_0) > c_{n, \alpha})$ of $H_0$ against the sequence of local alternatives $\{\int_A \int_{R^b} f(y, \theta_0 + N_n^{-1}b, \lambda)/f(y, \theta_0)dQ_\lambda(b)dJ(\lambda)\}_{n \geq 1}$. Logically, we cannot obtain a PVOT on $\Lambda$ for a smoothed test statistic like $\xi$, the sequence of local alternatives $\{\xi\}$ that are not smoothed on $\Lambda$, precisely by measuring how often the non-smoothed PV test rejects on $\Lambda$. By comparison, Andrews and Ploberger (1994) only treat test statistics like $\xi(Y_n) \in \{0, 1\}$ which involve presmoothing over the nuisance parameter $\lambda$ and drift $b$.

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

Define the operator $E_{f(\theta_1, \lambda)}[\omega] \equiv \int_{R^b} zf(z, \theta_1, \lambda)dz$, and define the corresponding $\omega$-PVOT $P_{\xi}^{(\omega)}(Y_n) \equiv \int_A \xi(Y_n, \lambda)d\omega(Y_n, \theta_1, \lambda)$. If the probability measure in (6) is now $g(y) = \int_A f(y, \theta_1, \lambda)dJ(\lambda)$, then $d\omega(y, \theta_1, \lambda) = f(y, \theta_1, \lambda)dJ(\lambda) / (\int_A f(y, \theta_1, \lambda)dJ(\lambda))$ 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 $\int_A E_{f(\theta_1, \lambda)}[P_{\xi}^{(\omega)}(Y_n)]dJ(\lambda)$, the weighted average mean $\omega$-PVOT, where the mean is evaluated under $H_1$.*

Now use Lebesgue measure $J(\lambda)$ on $\Lambda$, 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 $\omega$-PVOT now reduces to $P_{\xi}^{(\omega)}(Y_n) = \int_A \xi(Y_n, \lambda)d\lambda$, which is simply PVOT (3). Power under the null, of course, is trivial: by construction $\int_{R^b} \xi(y, \lambda)f(y, \theta_0)d\lambda = P(\xi(Y_n, \lambda) = 1) = P(p_n(\lambda) < \alpha) \to \alpha$ for each $\lambda$, hence by Fubini’s Theorem and bounded convergence $\int_A E_{f(\theta_1, \lambda)}[P_{\xi}^{(\omega)}(Y_n)]dJ(\lambda) \to \alpha$.

This reveals that the PVOT $\int_A \xi(Y_n, \lambda)d\lambda$ as in (3) is just the power relevant $\omega$-PVOT evaluated under the null with Lebesgue measure. Thus, PVOT $\int_A \xi(Y_n, \lambda)d\lambda$ is simply a point estimate of the $p$-value test weighted average probability of rejection, identically $\int_A E_{f(\theta_1, \lambda)}[P_{\xi}^{(\omega)}(Y_n)]dJ(\lambda)$, evaluated under $H_0$. This probability is no larger than $\alpha$ when $H_0$ is in fact true, hence if the PVOT $\int_A \xi(Y_n, \lambda)d\lambda \leq \alpha$ then we have evidence that either $H_0$ is correct, or global power is trivial. Conversely, $\int_A \xi(Y_n, \lambda)d\lambda > \alpha$ 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 $\Lambda$ with measure greater than $\alpha$, in which case, $\int_A \xi(Y_n, \lambda)d\lambda \leq \alpha$ 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

\[ y_t = f(x_t, \zeta_0) + \epsilon_t, \]

where \( \zeta_0 \) lies in the interior of \( \mathcal{Z} \), 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 \( f : \mathbb{R}^k \times \mathcal{Z} \rightarrow \mathbb{R} \) is a known response function. Assume \( \{e_t, x_t, y_t\} \) are stationary for simplicity. Let \( \Psi \) be a 1-1 bounded mapping from \( \mathbb{R}^k \) to \( \mathbb{R}^k \), let \( F : \mathbb{R} \rightarrow \mathbb{R} \) be analytic and non-polynomial (e.g. exponential or logistic), and assume \( \lambda \in \Lambda \), a compact subset of \( \mathbb{R}^k \). Misspecification sup \( \zeta \in \mathcal{Z} \) \( P(\mathbb{E}[y_t|x_t] = f(x_t, \zeta)) < 1 \) implies \( E[e_tF(\lambda'\Psi(x_t))] \neq 0 \forall \lambda \in \Lambda/S \), where \( S \) 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 \( H_0 : \mathbb{E}[y_t|x_t] = f(x_t, \zeta_0) \) a.s. is

\[ T_n(\lambda) = \left( \frac{1}{\hat{v}_n(\lambda)} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} e_t(\hat{\zeta}_n)F(\lambda'\Psi(x_t)) \right)^2 \]

where \( e_t(\zeta) \equiv y_t - f(x_t, \zeta) \).

The estimator \( \hat{\zeta}_n \) is \( \sqrt{n} \)-consistent of a strongly identified \( \zeta_0 \), and \( \hat{v}_n^2(\lambda) \) is a consistent estimator of \( E[1/\sqrt{n} \sum_{t=1}^{n} e_t(\hat{\zeta}_n)F(\lambda'\Psi(x_t))]^2 \). By application of Theorem 3.3, below, under regularity conditions detailed below the asymptotic p-value is \( p_n(\lambda) \equiv 1 - \tilde{F}_0(T_n(\lambda)) \) where \( \tilde{F}_0 \) 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

\[ y_t = f(x_t, \zeta_0) + \beta_0 F(\lambda'\Psi(x_t)) + \epsilon_t. \]  \( \tag{9} \)

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 \( \{c(\lambda)\} \):

\[ \{T_n(\lambda) : \lambda \in \Lambda\} \Rightarrow^* \{(Z(\lambda) + bc(\lambda))^2 : \lambda \in \Lambda\} \]  \( \tag{11} \)
where \( \{Z(\lambda) + c(\lambda)b\} \) 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. \( \{y_t, x_t, \varepsilon_t\} \) are stationary; \( E|y_t|^{4+\varepsilon} < \infty \) and \( E|\varepsilon_t|^{4+\varepsilon} \) for tiny \( \varepsilon > 0 \); \( E[\varepsilon_t|x_t] = 0 \) a.s. under \( H_1^t \); \( E[\inf_{\lambda \in \Lambda} w_t^2(\lambda)] > 0 \), \( E[x_t^2 \inf_{\lambda \in \Lambda} w_t^2(\lambda)] > 0 \), and \( \inf_{\lambda \in \Lambda} ||(\partial / \partial \lambda)E[\varepsilon_t^2 F(\lambda^2 \Psi(x_t))]|^2 || > 0 \); \( \{x_t, \varepsilon_t\} \) are \( \beta \)-mixing with mixing coefficients \( \beta_n = O(h^{-4-\delta}) \) for tiny \( \delta > 0 \).

b. Response Function: \( f: \mathbb{R}^k \times \mathcal{Z} \rightarrow \mathbb{R} \); \( f(\cdot, \zeta) \) is twice continuously differentiable; \( (\partial / \partial \zeta)^i f(x, \zeta) \) are Borel measurable for each \( \zeta \in \mathcal{Z} \) and \( i = 0, 1, 2 \); write \( h_t(\zeta) = (\partial / \partial \zeta)^i f(x_t, \cdot) \) for \( i = 0, 1, 2 \); \( E[\sup_{\zeta \in \mathcal{Z}} |h_t(\zeta)|^{4+\delta}] < \infty \) 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; \( \Psi \) is one-to-one and bounded.

d. Variance Estimator:

\[
\hat{\sigma}^2_n(\lambda) \equiv \frac{1}{n} \sum_{s,t=1}^n K((s-t)/\gamma_n) \varepsilon_s(\hat{\zeta}_n) \varepsilon_t(\hat{\zeta}_n) \hat{w}_{n,s}(\lambda, \hat{\zeta}_n) \hat{w}_{n,t}(\lambda, \hat{\zeta}_n)
\]

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

e. Plug-In: \( \zeta_0 \) is an interior point of \( \mathcal{Z} \), and \( \hat{\zeta}_n \equiv \arg\min_{\zeta \in \mathcal{Z}} \{1/n \sum_{t=1}^n (y_t - f(x, \zeta))^2\} \).

**Remark 3.** The kernel variance \( \hat{\sigma}^2_n(\lambda) \) form follows from a standard expansion of \( 1/\sqrt{n} \sum_{t=1}^n \varepsilon_t(\hat{\zeta}_n) F(\lambda^2 \Psi(x_t)) \) around \( \zeta_0 \) under \( H_0 \). We exploit a kernel estimator in order to prove uniform convergence of \( \hat{\sigma}^2_n(\lambda) \) 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 \( \mathcal{I}_n \equiv \int_{\delta_n}^{\infty} \{\|K(x)| + |K(-x)|\} dx = o(1/\sqrt{n}) \) 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 \( \sup_{\lambda \in \Lambda} |\hat{\sigma}^2_n(\lambda) - v^2(\lambda)| \overset{p}{\rightarrow} 0 \) 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 $\{T_n(\lambda) : \lambda \in \Lambda\} \Rightarrow^{*} \{Z(\lambda)^2 : \lambda \in \Lambda\}$ where $\{Z(\lambda) : \lambda \in \Lambda\}$ is a zero mean Gaussian process with a version that has almost surely uniformly continuous sample paths, and covariance kernel

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

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:

$$g_t(\zeta) \equiv \frac{\partial}{\partial \zeta} f(x_t, \zeta) \quad \text{and} \quad F_t(\lambda) \equiv F(\lambda\Psi(x_t)) \quad (13)$$

$$\varphi_t(\lambda) \equiv F_t(\lambda)^2 - E[F_t(\lambda)g_t'] \times (E[g_t g_t'])^{-1} \times E[F_t(\lambda)g_t]$$

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

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

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

$$M_t(h, \lambda) \equiv m_t(\lambda)m_{t-h}(\lambda) - E[m_t(\lambda)m_{t-h}(\lambda)]$$

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

In the first two lemmas we only exploit a subset of Assumption 2. In particular, we do not require $\epsilon_t$ in model (9) to satisfy $E[\epsilon_t|x_t] = 0$ a.s. We only require $\epsilon_t$ to be stationary, mixing, $E[\epsilon_t] = 0$ and $E[\epsilon_t g_t] = 0$. First, the kernel estimator $\hat{v}^2_n(\lambda)$ 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; $E[|\epsilon_t|^{4+t} < \infty$ for tiny $t > 0$; $E[\epsilon_t] = 0$ and $E[\epsilon_t g_t] = 0$; $\{x_t, \epsilon_t\}$ are $\beta$-mixing with mixing coefficients $\beta_h = O(1/(h^2\ln(h)))$. Let $H_1^L$ hold. Then $v^2(\lambda) < \infty$ and $\sup_{\lambda \in \Lambda} |\hat{v}^2_n(\lambda) - v^2(\lambda)| \rightarrow 0.$
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 $E[\epsilon_t(\partial/\partial \zeta)g_t] = 0$:

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

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

Lemma C.3 (Functional CLT). Under Assumption 2 $\{\hat{Z}_n(\lambda)\} \Rightarrow^* \{Z(\lambda) + bE[\varphi_t(\lambda)]/v(\lambda)\}$, where $\{Z(\lambda)\}$ 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 $H^*_1: \beta_0 = b/n^{1/2}$ 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, $\sup_{\lambda \in \Lambda} |\hat{v}_t^2(\lambda) - v^2(\lambda)| \xrightarrow{p} 0$. Hence, by Lemmas C.2 and C.3:

$$\left\{ \frac{1}{\hat{v}_t(\lambda) n^{1/2}} \sum_{t=1}^{n} \epsilon_t(\hat{\zeta}_n) F_t(\lambda(\Psi(x_t))) : \lambda \in \Lambda \right\} \Rightarrow^* \left\{ \bar{Z}(\lambda) + \frac{b}{v(\lambda)} E[\varphi_t(\lambda)] \right\},$$

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 $E[\varphi_t(\lambda)]/v(\lambda) = E[w_t^2(\lambda)]/(E[\epsilon_t^2 w_t^2(\lambda)])^{1/2}$. QED.

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

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

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

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 $L_{2,\beta}(P)$ of $L_2(P)$ 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 $\beta$-mixing property, measurability of $\mathcal{M}_t(h,\lambda)$, boundedness of $F$, $E[\sup_{\zeta \in \mathbb{Z}} |g_t(\zeta)|^{4+\varepsilon}] < \infty$ and $E|\varepsilon_t|^{4+\varepsilon} < \infty$, by exploiting a well known Cramér-Wold theorem argument and a $\beta$-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 $\int_0^1 |\ln(N_\delta(\varepsilon,\mathcal{H} \times \Lambda, || \cdot ||_2)|^{1/2}d\varepsilon < \infty$ with $L_2$-bracketing numbers $N_\delta(\varepsilon,\mathcal{H} \times \Lambda, || \cdot ||_{2,\beta})$. 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)

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

By Minkowski’s inequality:

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

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

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

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

11
Furthermore, since $H$ is compact, by boundedness of $F$ and $E|\epsilon_t|^{1+t} < \infty$:

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

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

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

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

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

Similarly, it follows

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

This, combined with (15), proves (14). QED.

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

$$\hat{M}_{n,t,h}(\lambda, \zeta) \equiv e_t(\zeta)\hat{w}_{n,t}(\lambda, \zeta)e_{t-h}(\zeta)\hat{w}_{n,t-h}(\lambda, \zeta)$$

and

$$\hat{M}_{n,h}(\lambda, \zeta) \equiv e_t(\zeta)\hat{w}_{n}(\lambda, \zeta)e_{t-h}(\zeta)\hat{w}_{n-h}(\lambda, \zeta).$$

We do not show that $\hat{M}_{n,h}(\lambda, \zeta)$ depends on $n$ through $e_t$ under $H_1^L$. By construction $\hat{\omega}_n^2(\lambda) = \ldots$
\[
\sum_{h=1-n}^{n-1} \mathcal{K}(h/\gamma_n) \hat{M}_{n,h}(\lambda, \hat{\zeta}_n) \text{ and } v^2(\lambda) = \sum_{h=-\infty}^{\infty} E[\mathcal{M}_{t,h}(\lambda)]. \text{ We have:}
\]

\[
v_n^2(\lambda) - v^2(\lambda) = \sum_{h=-\infty}^{\infty} E[\mathcal{M}_{t,h}(\lambda)] - v^2(\lambda)
\]

\[
+ \left\{ \sum_{h=1-n}^{n-1} \mathcal{K}(h/\gamma_n)E[\hat{\mathcal{M}}_h(\lambda)] - \sum_{h=-\infty}^{\infty} E[\mathcal{M}_{t,h}(\lambda)] \right\}
\]

\[
+ \sum_{h=1-n}^{n-1} \mathcal{K}(h/\gamma_n) \left\{ \hat{\mathcal{M}}_h(\lambda) - E[\hat{\mathcal{M}}_h(\lambda)] \right\}
\]

\[
+ \sum_{h=1-n}^{n-1} \mathcal{K}(h/\gamma_n) \left\{ \hat{\mathcal{M}}_h(\lambda, \hat{\zeta}_n) - \hat{\mathcal{M}}_h(\lambda) \right\}
\]

\[
+ \sum_{h=1-n}^{n-1} \mathcal{K}(h/\gamma_n) \left\{ \hat{\mathcal{M}}_{n,h}(\lambda, \hat{\zeta}_n) - \hat{\mathcal{M}}_h(\lambda, \hat{\zeta}_n) \right\}
\]

\[
= \mathcal{A}_n(\lambda) + \mathcal{B}_n(\lambda) + \mathcal{C}_n(\lambda) + \mathcal{D}_n(\lambda) + \mathcal{E}_n(\lambda) + \mathcal{F}_n(\lambda).
\]

By construction \( \mathcal{A}_n(\lambda) = 0 \). We show that each remaining term converges to zero in probability, uniformly on \( \Lambda \). In view of \( E|\epsilon_t|^{4+\epsilon} < \infty \) 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 \( \zeta_0 \). We have:

\[
\sup_{\lambda \in \Lambda} |\mathcal{B}_n(\lambda)| = \sup_{\lambda \in \Lambda} \left| \sum_{h=-\infty}^{\infty} E\left[ \mathcal{M}_{t,h}(\lambda) \right] - \sum_{h=1-n}^{n-1} \mathcal{K}(h/\gamma_n) \frac{n-h}{n} E\left[ \mathcal{M}_{t,h}(\lambda) \right] \right|
\]

\[
\leq \sup_{\lambda \in \Lambda} \left| \sum_{h=-\infty}^{\infty} E\left[ \mathcal{M}_{t,h}(\lambda) \right] - \sum_{h=1-n}^{n-1} \mathcal{K}(h/\gamma_n) E\left[ \mathcal{M}_{t,h}(\lambda) \right] \right|
\]

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

\[
\leq \sup_{\lambda \in \Lambda} \left| \sum_{h=1-n}^{n-1} \left\{ E\left[ \mathcal{M}_{t,h}(\lambda) \right] - \mathcal{K}(h/\gamma_n) E\left[ \mathcal{M}_{t,h}(\lambda) \right] \right\} \right| + \sup_{\lambda \in \Lambda} \left| \sum_{|h| \geq n} E\left[ \mathcal{M}_{t,h}(\lambda) \right] \right|
\]
Similarly, hence, \( \delta > 0 \) where the last inequality follows from Lemma 1.3 in Ibragimov (1962) for some \( C \). Therefore \( \sup \lambda \in \Lambda \{1 - K(h/\gamma_n)\} E[M_{t,h}(\lambda)] \) for tiny \( \gamma \) as in view of 2. Now use \( \alpha_h = O(1/(h^2 \ln(h))) \) to deduce

\[
\sum_{h=1}^{n-1} |1 - K(h/\gamma_n)| \alpha_h^{\delta/(2+\delta)} \leq K \sum_{h=1}^{n-1} |1 - K(h/\gamma_n)| \frac{1}{h^{2\delta/(2+\delta)} (\ln(h))^{\delta/(2+\delta)}},
\]

Hence, \( \sum_{h=1}^{n-1} |1 - K(h/\gamma_n)| \alpha_h^{2\delta/(2+\delta)} \to 0 \) 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 \( \sup_{\lambda \in \Lambda} |B_n(\lambda)| \to 0 \).

**Step 2 (C(\lambda))**: Drop \( \zeta_0 \). Let \( \{\delta_n\} \) be a sequence of numbers, \( \delta_n \in \{1, ..., n-1\}, \delta_n = o(n) \).
Then:

\[
\sup_{\lambda \in \Lambda} |C_n(\lambda)| = \sup_{\lambda \in \Lambda} \left| \sum_{h=1-n}^{n-1} \mathcal{K}(h/\gamma_n) \left\{ \hat{\mathcal{M}}_h(\lambda) - E \left[ \hat{\mathcal{M}}_h(\lambda) \right] \right\} \right|
\]

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

\[
\leq \sum_{h=1-n}^{n-1} |\mathcal{K}(h/\gamma_n)| \sup_{\lambda \in \Lambda} \left| \frac{1}{n} \sum_{t=|h|+1}^{n} \{ \mathcal{M}_{t,h}(\lambda) - E [\mathcal{M}_{t,h}(\lambda)] \} \right|
\]

\[
= \sum_{|h| \leq \delta_n} |\mathcal{K}(h/\gamma_n)| \sup_{\lambda \in \Lambda} \left| \frac{1}{n} \sum_{t=|h|+1}^{n} \{ \mathcal{M}_{t,h}(\lambda) - E [\mathcal{M}_{t,h}(\lambda)] \} \right|
\]

\[
+ \sum_{|h| = \delta_n+1}^{n-1} |\mathcal{K}(h/\gamma_n)| \sup_{\lambda \in \Lambda} \left| \frac{1}{n} \sum_{t=|h|+1}^{n} \{ \mathcal{M}_{t,h}(\lambda) - E [\mathcal{M}_{t,h}(\lambda)] \} \right|
\]

\[
= a_n + b_n
\]

say. Consider \( a_n \), and define

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

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 \( \{ W(x) : -\infty < x < \infty \} \) that satisfies

\[
\left\{ \sup_{h \in \{ 1, \ldots, \delta_n \}} W_n(h/\gamma_n) \right\} \Rightarrow^* \left\{ \sup_{x \in \mathbb{R}} W(x) \right\} = O_p(1)
\]

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

\[
a_n = \frac{1}{\gamma_n} \sum_{|h| \leq \delta_n} |\mathcal{K}(h/\gamma_n)| \sup_{\lambda \in \Lambda} \left| \frac{1}{\sqrt{n}} \sum_{t=|h|+1}^{n} \{ \mathcal{M}_{t,h}(\lambda) - E [\mathcal{M}_{t,h}(\lambda)] \} \right|
\]
\[
\leq \frac{1}{\gamma_n} \sum_{|h| \leq \delta_n} |\mathcal{K}(h/\gamma_n)| \sup_{h \in \{1, \ldots, \delta_n\}} \sup_{\lambda \in \Lambda} \left| \frac{1}{\sqrt{n}} \sum_{t=|h|+1}^{n} \{ \mathcal{M}_{t,h}(\lambda) - E[\mathcal{M}_{t,h}(\lambda)] \} \right|
\]

\[
= o_p \left( \int_{-\infty}^{\infty} |\mathcal{K}(x)| \, dx \right) = o_p(1).
\]

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 \( \beta \)-mixing property to deduce

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

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

**Step 3** (\( D_n(\lambda) \)): Use \( e_t = bn^{-1/2} F_t(\lambda) + \epsilon_t \) under \( H^L_1 \) to deduce

\[
\left| \frac{1}{\sqrt{n}} \sum_{t=|h|+1}^{n} \{ \epsilon_t \epsilon_{t-h} - \epsilon_t \epsilon_{t-h} \} w_{t-h}(\lambda) w_t(\lambda) \right| \leq |b| \frac{1}{n} \sum_{t=|h|+1}^{n} \{ |\epsilon_t F_{t-h}(\lambda) + \epsilon_{t-h} F_t(\lambda)| w_{t-h}(\lambda) w_t(\lambda) |
\]

\[
+ b^2 \frac{1}{n^{3/2}} \sum_{t=|h|+1}^{n} |F_t(\lambda) F_{t-h}(\lambda)| \times |w_{t-h}(\lambda) w_t(\lambda)|
\]

\[
\leq |b| \frac{1}{n} \sum_{t=1}^{n} \{ |\epsilon_t F_{t-h}(\lambda) + \epsilon_{t-h} F_t(\lambda)| w_{t-h}(\lambda) w_t(\lambda) |
\]

\[
+ b^2 \frac{1}{n^{3/2}} \sum_{t=1}^{n} |F_t(\lambda) F_{t-h}(\lambda)| \times |w_{t-h}(\lambda) w_t(\lambda)|
\]

\[
\equiv \mathcal{X}_n(\lambda, h),
\]

16
say. Since $|\epsilon_tF_{t-h}(\lambda) + \epsilon_{t-h}F_t(\lambda)| \times |w_{t-h}(\lambda)w_t(\lambda)|$ and $|F_t(\lambda)F_{t-h}(\lambda)| \times |w_{t-h}(\lambda)w_t(\lambda)|$ are stationary, ergodic and integrable, it follows that

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

Let

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

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)

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

Therefore

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

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

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

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

$$= a_n,$$

say. By Lemma C.2.b, for some $\mathcal{O}_p(1)$ that is not a function of $\lambda$:

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

The $\beta$-mixing decay rate $\beta_i = O(1/(i^2 \ln(i)))$ combined with $E[\epsilon_tg_t] = 0$ imply $E[1/\sqrt{n} \sum_{t=1}^{n} \epsilon_tg_t]^2 < \infty$ by application of Lemma 1.3 in Ibragimov (1962). This implies $1/\sqrt{n} \sum_{t=1}^{n} \epsilon_tg_t = \mathcal{O}_p(1)$. 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:

$$
\sup_{\lambda \in \Lambda} \left\| \frac{1}{n} \sum_{t=\lfloor h \rfloor + 1}^{n} \frac{\partial}{\partial \zeta} \mathcal{M}_{t,h}(\lambda, \zeta_n^*) \right\| \leq K \frac{1}{n} \sum_{t=\lfloor h \rfloor + 1}^{n+h} |e_{t-h}| \left( K + K \sup_{\zeta \in \mathbb{R}} \left\| \frac{\partial}{\partial \zeta} f(x_t, \zeta) \right\| \right)
$$

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

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

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

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

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

$$
= o_p \left( \frac{1}{\gamma_n} \sum_{|h| \leq \delta_n} |K(h/\gamma_n)| \right) = o_p(1).
$$

**Step 5 ($F_n(\lambda)$):** 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

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

(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:

\[
\sup_{\lambda \in \Lambda} |F_n(\lambda)| \leq \frac{1}{\gamma_n} \sum_{h=1-n}^{n-1} |\mathcal{K}(h/\gamma_n)| \gamma_n \left\{ \frac{1}{n} \sum_{t=1}^{n} |e_t(\hat{\zeta}_n)e_{t-h}(\hat{\zeta}_n)| \right\} \times \sup_{\lambda \in \Lambda} \left| \hat{w}_{n,t}(\lambda, \hat{\zeta}_n)\hat{w}_{n,t-h}(\lambda, \hat{\zeta}_n) - w_t(\lambda, \hat{\zeta}_n)w_{t-h}(\lambda, \hat{\zeta}_n) \right| .
\]

Define

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

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

Re-arrange terms in (18) to deduce:

\[
\sup_{\lambda \in \Lambda} \left| \hat{w}_{n,t}(\lambda, \hat{\zeta}_n)\hat{w}_{n,t-h}(\lambda, \hat{\zeta}_n) - w_t(\lambda, \hat{\zeta}_n)w_{t-h}(\lambda, \hat{\zeta}_n) \right|
\leq \sup_{\lambda \in \Lambda} \left| \hat{F}_n(\lambda, \zeta) - F(\lambda, \zeta) \right| \times \left( \left| g_t(\hat{\zeta}_n) \right| + \left| g_{t-h}(\hat{\zeta}_n) \right| \right)
\]

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

Now combine (17) and (18):

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

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 \( \{X_n(\lambda, h) : \lambda, h \in \Lambda \times H_N \} \), \( \sup_{\lambda, h \in \Lambda \times H_N} \|X_n(\lambda, h)\| = O_p(1) \):

\[
\sup_{\lambda, h \in \Lambda \times H_N} \left\| \frac{1}{n} \sum_{t=1}^{n} e_t(\hat{\zeta}_n)e_{t-h}(\hat{\zeta}_n) \right\| : \left\{ \left| g_t(\hat{\zeta}_n) \right| + \left| g_{t-h}(\hat{\zeta}_n) \right| + \left| g_t(\hat{\zeta}_n)g_{t-h}(\hat{\zeta}_n) \right| \right\} - X_n(\lambda, h) \xrightarrow{p} 0,
\]

hence

\[
\sup_{\lambda, h \in \Lambda \times H_N} \frac{1}{n} \sum_{t=1}^{n} e_t(\hat{\zeta}_n)e_{t-h}(\hat{\zeta}_n) \left\{ \left| g_t(\hat{\zeta}_n) \right| + \left| g_{t-h}(\hat{\zeta}_n) \right| + \left| g_t(\hat{\zeta}_n)g_{t-h}(\hat{\zeta}_n) \right| \right\} = O_p(1). \quad (20)
\]

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

\[
\frac{1}{\gamma_n} \sum_{h=1-n}^{n-1} |K(h/\gamma_n)| \frac{1}{n} \sum_{t=1}^{n} e_t(\hat{\zeta}_n)e_{t-h}(\hat{\zeta}_n) \left\{ \left| g_t(\hat{\zeta}_n) \right| + \left| g_{t-h}(\hat{\zeta}_n) \right| + \left| g_t(\hat{\zeta}_n)g_{t-h}(\hat{\zeta}_n) \right| \right\}
\]

\[
\leq K \int_{-\infty}^{\infty} |K(x)| \, dx + o_p(1).
\]

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

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 \( \beta \)-mixing with mixing coefficients \( \beta_i = O(1/(i^2 \ln(i))) \). Then:

a. \( \hat{\zeta}_n \xrightarrow{p} \zeta_0 \)

b. \( \sup_{\lambda \in \Lambda} \left\| \sqrt{n} \left( \hat{\zeta}_n - \zeta_0 \right) - (E[g_t g_t])^{-1} \left( \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \epsilon_t g_t + b \times E[F_t(\lambda)g_t] \right) \right\| \xrightarrow{p} 0. \)

**Proof.** We first derive some required asymptotic properties. Recall \( \beta \)-mixing implies ergodicity. We have:

\[
\sup_{\zeta \in \mathbb{Z}} \left\| \frac{1}{n} \sum_{t=1}^{n} \epsilon_t(\zeta)g_t(\zeta) - E[\epsilon_t(\zeta)g_t(\zeta)] \right\| \xrightarrow{p} 0 \quad (22)
\]
Consider (22), and write \( w_t(\zeta) = \epsilon_t(\zeta) g_t(\zeta) - E[\epsilon_t(\zeta) g_t(\zeta)] \). Pointwise convergence \( 1/n \sum_{t=1}^{n} w_t(\zeta) \overset{p}{\to} 0 \) 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 \( ||w_t(\zeta) - w_t(\hat{\zeta})||_2 \leq K||\zeta - \hat{\zeta}|| \) (cf., Newey, 1991, Corollary 3.1). By the mean value theorem, Jensen and Minkowski inequalities, and the assumed moment bounds:

\[
\|w_t(\zeta) - w_t(\hat{\zeta})\|_2 \leq 2 \|\epsilon_t(\zeta) g_t(\zeta) - \epsilon_t(\hat{\zeta}) g_t(\hat{\zeta})\|_2
\]

\[
= 2 \sup_{\zeta \in \mathbb{Z}} \left\| \frac{\partial}{\partial \zeta} \epsilon_t(\zeta) g_t(\zeta) + \epsilon_t(\zeta) \frac{\partial}{\partial \zeta} g_t(\zeta) \right\|_2 \times \|\zeta - \hat{\zeta}\| \leq K \|\zeta - \hat{\zeta}\|.
\]

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

\[ o_p(1) = \sup_{\zeta \in \mathbb{Z}} \left\| \frac{1}{n} \sum_{t=1}^{n} \epsilon_t(\zeta) g_t(\zeta) - E[\epsilon_t(\zeta) g_t(\zeta)] \right\| \]

\[ \geq \left\| \frac{1}{n} \sum_{t=1}^{n} \epsilon_t(\hat{\zeta}_n) g_t(\hat{\zeta}_n) - E[\epsilon_t(\hat{\zeta}_n) g_t(\hat{\zeta}_n)] \right\| = \left\| E[\epsilon_t(\hat{\zeta}_n) g_t(\hat{\zeta}_n)] \right\|.
\]

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

**Claim (b).** By the least squares first order condition, and the mean value theorem, for some
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

\[
\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \epsilon_t(\hat{\zeta}_n) F_t(\lambda) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left( g_t - f(x_t, \zeta_0) \right) F_t(\lambda) - \left\{ \frac{1}{n} \sum_{t=1}^{n} F_t(\lambda) g_t(\zeta_0^*) \right\}' \sqrt{n} (\hat{\zeta}_n - \zeta_0)
\]

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

Proof of Lemma C.3. Write

\[
\hat{Z}_n(\lambda) \equiv \frac{1}{v(\lambda)} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \epsilon_t w_t(\lambda) + b \frac{1}{v(\lambda)} \frac{1}{n} \sum_{t=1}^{n} \varphi_t(\lambda) = Z_n(\lambda) + A_n(\lambda)
\]

\[
A(\lambda) \equiv b \frac{1}{v(\lambda)} E[\varphi_t(\lambda)]
\]
Under Assumption 2 $v^2(\lambda) = E[\epsilon_t^2w_t^2(\lambda)] > 0$, and $E[Z_n(\lambda)Z_n(\tilde{\lambda})]$ is identically the covariance kernel (12). It suffices to prove $\{Z_n(\lambda)\} \Rightarrow^* \{Z(\lambda)\}$ and $\sup_\lambda ||A_n(\lambda) - E[A(\lambda)]|| \xrightarrow{p} 0$.

**Step 1 ($Z_n(\lambda)$).** The proof follows the lines of the proof of Lemma C.4. Convergence in the finite dimensional distributions of $\{Z_n(\lambda)\}$ 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):

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

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

$$E \left[ \left( Z_n(\lambda) - Z_n(\tilde{\lambda}) \right)^2 \right] = E \left[ \epsilon_t^2 \left\{ \frac{w_t(\lambda)}{v(\lambda)} - \frac{w_t(\tilde{\lambda})}{v(\lambda)} \right\}^2 \right]. \quad (28)$$

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

$$E \left[ \epsilon_t^2 \left\{ \frac{w_t(\lambda)}{v(\lambda)} - \frac{w_t(\tilde{\lambda})}{v(\lambda)} \right\}^2 \right] = E \left[ \epsilon_t^2 \left\{ \frac{\partial}{\partial \lambda^*} \frac{w_t(\lambda)}{v(\lambda)} \left( \lambda - \lambda^* \right) \right\}^2 \right] \quad (29)$$

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

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:

$$\sup_{\lambda \in \Lambda} \left( E \left[ \epsilon_t^2 \left( \sum_{i=1}^k \frac{\partial}{\partial \lambda_i} \frac{w_t(\lambda)}{v(\lambda)} \right)^2 \right] \right)^{1/2}$$

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

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

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

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

The required uniform law follows by adapting the proof of (22) in the proof of Lemma C.2. QED.

D Theorem 4.4: Structural Break Test

D.1 Model and Test Statistic

The model we consider for simplicity is

\[ y_t = \beta'_t x_t + \epsilon_t \text{ where } \beta_t \in \mathbb{R}^{k_\beta} \text{ may depend on } t \geq 1, \text{ and } k_\beta \geq 1. \quad (31) \]

Let \(\beta_t \in B\), a compact subset of \(\mathbb{R}^{k_\beta}\). We want to test for parameter constancy

\[ H_0 : \beta_t = \beta_0 \ \forall t \geq 1, \]

against a one-time change point

\[ H_1 : \beta_t = \beta_1 \text{ for } t = 1, \ldots, [\lambda n] \text{ and } \beta_t = \beta_2 \text{ for } t \geq [\lambda n] + 1. \]

It is a simple extension to allow for a subset of regressors to have coefficients that do not change (see Andrews, 1993).

The parameters \(\beta_t\) are constants, and \(\lambda\) is a nuisance parameter under \(H_1\). Let \(\theta = [\beta'_1, \beta'_2]'\) and define regressors for each sample partition:

\[ x_{n,t}(\lambda) \equiv [x'_t I (1 \leq t \leq [\lambda n]), x'_t I ([\lambda n] + 1 \leq t \leq n)]'. \]
Define the unrestricted least squares estimator:

\[ \hat{\theta}_n(\lambda) = \left[ \hat{\beta}_{1,n}(\lambda)', \hat{\beta}_{2,n}(\lambda) \right]' = \arg \min_{\theta \in \mathbb{R}^{2k_\beta}} \sum_{t=1}^{n} (y_t - \theta' x_{n,t}(\lambda))^2. \]

Andrews’ (1993) Wald statistic is:

\[ T_n(\lambda) = \left( \hat{\beta}_{n,1}(\lambda) - \hat{\beta}_{n,2}(\lambda) \right)' \left\{ \hat{V}_{n,i}(\lambda)/\lambda + \hat{V}_{n,i}(\lambda)/(1 - \lambda) \right\}^{-1} \left( \hat{\beta}_{n,1}(\lambda) - \hat{\beta}_{n,2}(\lambda) \right). \]

The variance estimators \( \hat{V}_{n,i}(\lambda) \) can be defined using a partitioned or full sample (Andrews, 1993, eq’s (3.9)-(3.14)). Define

\[ m_t(\beta) \equiv (y_t - \beta' x_t) x_t \]

\[ m_{n,t,1}(\beta, \lambda) \equiv m_t(\beta) - \frac{1}{\lambda n} \sum_{t=1}^{\lambda n} m_t(\beta) \text{ and } m_{n,t,2}(\beta, \lambda) \equiv m_t(\beta) - \frac{1}{(1 - \lambda) n} \sum_{t=\lambda n+1}^{n} m_t(\beta) \]

\[ \mathcal{M}_{n,t,i}(\beta, \lambda) \equiv m_{n,t,i}(\beta, \lambda)m_{n,t,i}(\beta, \lambda)'. \]

In the partial sample case:

\[ \hat{V}_{n,i}(\lambda) \equiv \left\{ \hat{J}_{n,i}(\lambda)' \hat{S}_{n,i}^{-1}(\lambda) \hat{J}_{n,i}(\lambda) \right\}^{-1} \]

where:

\[ \hat{J}_{n,1}(\lambda) = \frac{1}{\lambda n} \sum_{t=1}^{\lambda n} x_t x'_t \text{ and } \hat{J}_{n,2}(\lambda) = \frac{1}{(1 - \lambda) n} \sum_{t=\lambda n+1}^{n} x_t x'_t \]

\[ \hat{S}_{n,1}(\lambda) = \frac{1}{\lambda n} \sum_{t=1}^{\lambda n} \mathcal{M}_{n,t,1}(\hat{\beta}_{n,1}(\lambda), \lambda)' \text{ and } \hat{S}_{n,2}(\lambda) = \frac{1}{(1 - \lambda) n} \sum_{t=\lambda n+1}^{n} \mathcal{M}_{n,t,2}(\hat{\beta}_{n,2}(\lambda), \lambda)'. \]

In the full sample case each \( \hat{V}_{n,i}(\lambda) \equiv \left\{ \hat{J}_{n}(\lambda)' \hat{S}_{n}^{-1}(\lambda) \hat{J}_{n}(\lambda) \right\}^{-1} \) where \( \hat{J}_{n}(\lambda) = \hat{J}_{n,1}(1) \) and \( \hat{S}_{n}(\lambda) = \hat{S}_{n,1}(1) \). The PVOT test is based on the asymptotic p-value \( P(T_n(\lambda) > \chi^2_{k_\beta}) \) where \( \chi^2_{k_\beta} \) is distributed \( \chi^2_{k_\beta} \).

Define

\[ \mathcal{V} \equiv \sigma^2 \mathcal{J}^{-1} \text{ and } \mathcal{J} \equiv \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} E[x_t x'_t] \text{ where } \sigma^2 = E[\epsilon_t^2]. \]

We assume \( \mathcal{J} \) exists by assumption below.
D.2 Theorem 4.4

The main result follows.

**Theorem 4.4.** Assume the following: (i) \( \beta_t \in \mathcal{B} \subset \mathbb{R}^{k_\beta} \) \( \forall t \in \mathbb{Z} \), and under the null \( \beta_0 \) lies in the interior of \( \mathcal{B} \). (ii) \( \{\epsilon_t, x_t, y_t\}_{t \geq 1} \) lies on a common probability space \((\Omega, \mathcal{F}, \mathcal{P})\). (iii) \( \epsilon_t \) is \( L_r \)-bounded for some \( r > 2 \) uniformly over \( t \); \( x_t \) is \( L_q \)-bounded for some \( q > 4 \) uniformly over \( t \); \( (\epsilon_t, x_t) \) are \( \mathcal{F}_t \)-measurable, \( \sigma(\mathcal{U}_{t \in \mathbb{N}} \mathcal{F}_t) \subseteq \mathcal{F} \) for \( t \geq 1 \), and \( \alpha \)-mixing with coefficients \( \alpha_h = O(h^{-r/(r-2)}) \); and \( \{x_t, y_t\}_{t \geq 1} \) are stationary under \( H_0 \). (iv) \( \{\epsilon_t, \mathcal{F}_t\} \) and \( \{\epsilon_t^2 - \sigma^2, \mathcal{F}_t\} \) are stationary martingale difference sequences. (v) \( \lim_{n \to \infty} 1/n \sum_{t=1}^n E[x_t x'_t] \) exists and is positive definite. Finally, (vi) \( E[\sup_{k,j \in \mathbb{N}} \{1/l \sum_{t=k+1}^{k+l} |y_t x_t| + K x_t^2\}] < \infty \).

a. Under \( H_0 \), \( \{T_n(\lambda) : \lambda \in \Lambda\} \Rightarrow^* \{T(\lambda) : \lambda \in \Lambda\} \) where \( T(\lambda) \equiv (\mathcal{B}_{k_\beta}(\lambda) - \lambda \mathcal{B}_{k_\beta}(1))' (\mathcal{B}_{k_\beta}(\lambda) - \lambda \mathcal{B}_{k_\beta}(1)) / |\lambda(1 - \lambda)| \), and \( \{\mathcal{B}_{k_\beta}(\lambda) : \lambda \in \Lambda\} \) is a \( k_\beta \times 1 \) vector of independent Brownian motions on \( \Lambda \). Therefore Assumption 1 holds, where \( \mathcal{T}(\lambda) \) is pointwise \( \chi^2(k_\beta) \) distributed.

b. Consider a sequence of local alternatives \( H_{1_t}^L : \beta_t = \beta_0 + \eta (t - [n\lambda]) / \sqrt{n} \) for fixed \( \eta \in \mathbb{R}^{k_\beta} \). Under \( H_{1_t}^L \), \( \{T_n(\lambda) : \lambda \in \Lambda\} \Rightarrow^* \mathcal{B}(\lambda, \eta)' \mathcal{B}(\lambda, \eta) \) where

\[
\mathcal{B}(\lambda, \eta) \equiv \frac{1}{\sqrt{\lambda(1 - \lambda)}} \left( \mathcal{B}_{k_\beta}(\lambda) - \lambda \mathcal{B}_{k_\beta}(1) \right) + \sqrt{\lambda(1 - \lambda)} \mathcal{V}^{-1/2} \eta,
\]

a pointwise non-central \( \chi^2(k_\beta) \) random variable with non-centrality \( \sqrt{\lambda(1 - \lambda)} \mathcal{V}^{-1/2} \eta \). The PVOT test is therefore consistent.

We require several preliminary results. Without loss of generality, \( \lambda n \) is treated as an integer whenever convenient for notational simplicity. This allows us, for example, to write compactly

\[
\frac{1}{\lambda n} \sum_{t=1}^{[\lambda n]} m_t(\beta) - E[m_t(\beta)] = \frac{1}{\lambda n} \sum_{t=1}^{[\lambda n]} \left( m_t(\beta) - E[m_t(\beta)] \right),
\]

and therefore treat the resulting error as \( ([\lambda n] / \lambda n - 1) \mathcal{E} m_t([\beta]) = 0 \). We write processes \( \{\mathcal{A}(\lambda) : \lambda \in \Lambda\} \) as \( \{\mathcal{A}(\lambda)\} \) to reduce notation.

**Lemma D.1.** Assume the conditions of Theorem 4.4.a hold. Write \( x_t = [x_{i,t}]_{i=1}^{k_\beta} \).

a. For each \( \xi_t \in \{\epsilon_t, x_{i,t}, \epsilon_t x_{i,t} x_{j,t}, x_{i,t} x_{j,t}, \epsilon_t^2 x_{i,t} x_{j,t}, x_{i,t} x_{j,t} x_{k,t} x_{l,t}\} \), \( i, j, k, l \in \{1, \ldots, k_\beta\} \):

\[
\sup_{\lambda \in \Lambda} \left| \frac{1}{n} \sum_{t=1}^{[\lambda n]} (\xi_t - E[\xi_t]) \right| \overset{p}{\to} 0 \text{ and } \sup_{\lambda \in \Lambda} \left| \frac{1}{n} \sum_{t=[\lambda n] + 1}^{n} (\xi_t - E[\xi_t]) \right| \overset{p}{\to} 0.
\]

(32)
The martingale difference assumption implies:

Claim (b).

\[
\sum_{t=1}^{n} x_{n,t}(\lambda) \epsilon_t \] \Rightarrow \ast \left\{ \begin{array}{c}
B_{\lambda}(\lambda) \\
(1 - B_{\lambda}(\lambda)) \end{array} \right\},
\]

where \( B_{\lambda}(\lambda) \) is a \( k_\lambda \times 1 \) vector of independent Brownian motions on \( \Lambda \).

**Proof.**

Claim (a). Let \( \epsilon > 0 \) be a tiny number. Recall that \( \{\xi_t, \mathcal{F}_t\} \) is an \( L_p \)-mixingale, \( p \geq 1 \), if \( ||E[\xi_t|\mathcal{F}_{t-h}]||_p \leq c_t \psi_h \) and \( ||\xi_t - E[\xi_t|\mathcal{F}_{t+h}]||_p \leq c_t \psi_{h+1} \), for some \( c_t \geq 0 \) that may depend on \( t \), and \( \psi_h > 0 \) that depends on displacement \( h \). See Andrews (1988), cf. McLeish (1975).

Under the stated mixing and moment conditions, and by measurability, each \( \xi_t \) forms an \( L_{1+\epsilon} \)-bounded \( L_1 \)-mixingale \( \{\xi_t - E[\xi_t|\mathcal{F}_t]\} \), with constants \( c = K \sup_{t \in \mathbb{Z}} ||\xi_t||_r < \infty \) for each \( t \) and some \( r > 4 \), and coefficients \( \psi_h = K \alpha_h^{-1/r} = O(h^{-r/(r-2)}) = o(h^{-1/2}) \). See McLeish (1975, Lemma 2.1) and Andrews (1988, Example 4). Now use \( \sum_{t=[\lambda n]+1}^{n} = \sum_{t=1}^{n} - \sum_{t=1}^{[\lambda n]} \) and Lemma A2 in Andrews (1993) to yield (32).

Claim (b). The martingale difference assumption implies:

\[
E \left[ \left( \frac{1}{\sqrt{n}} \sum_{t=1}^{n} x_{n,t}(\lambda) \epsilon_t \right) \left( \frac{1}{\sqrt{n}} \sum_{t=1}^{n} x_{n,t}(\lambda) \epsilon_t \right)^\prime \right] = \sigma^2 \left[ \begin{array}{cc}
\frac{1}{n} \sum_{t=1}^{[\lambda n]} E[x_t x_t'] & 0 \\
0 & \frac{1}{n} \sum_{t=[\lambda n]+1}^{n} E[x_t x_t']
\end{array} \right] \rightarrow \sigma^2 \left[ \begin{array}{cc}
\lambda \mathcal{J} & 0 \\
0 & (1 - \lambda) \mathcal{J}
\end{array} \right].
\]

The claim therefore follows from arguments in Andrews (1993, p. 849). \( \square \)

**Lemma D.2.** Under the conditions of Theorem 4.4.b:

\[
\left\{ \sqrt{n} \left( \hat{\theta}_n(\lambda) - \theta_0 \right) \right\} \Rightarrow \ast \left\{ \begin{array}{c}
\frac{1}{\sqrt{n}} \mathcal{V}^{1/2} B_p(\lambda) + \eta \\
\frac{1}{1-\lambda} \mathcal{V}^{1/2} (B_p(1) - B_p(\lambda))
\end{array} \right\}
\]

**Proof.** The least squares estimator satisfies \( 0 = 1/n \sum_{t=1}^{n} (y_t - \hat{\theta}_n(\lambda)' x_{n,t}(\lambda)) x_{n,t}(\lambda). \) Under \( H_1^L \), \( \beta_t = \beta_0 + \eta \mathbb{1}(t \leq [n\lambda]) / \sqrt{n} \) for fixed \( \eta \in \mathbb{R}^{k_\lambda} \), hence:

\[
\sqrt{n} \left( \hat{\theta}_n(\lambda) - \theta_0 \right) = \left( \frac{1}{n} \sum_{t=1}^{n} x_{n,t}(\lambda)' x_{n,t}(\lambda) \right)^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} x_{n,t}(\lambda) \epsilon_t
\] (33)

\(^3\)Under \( H_1^L \), the observed \( y_t = y_{n,t} \) forms a triangular array. We do not show this in order to reduce notation.
\[ + \left( \frac{1}{n} \sum_{t=1}^{n} x_{n,t}(\lambda)x_{n,t}(\lambda)' \right)^{-1} \frac{1}{n} \sum_{t=1}^{n} x_{n,t}(\lambda)x_{n,t}(\lambda)' I (t \leq [n\lambda]) \eta \]

\[ = C_n(\lambda) + D_n(\lambda). \]

By construction

\[ \frac{1}{n} \sum_{t=1}^{n} x_{n,t}(\lambda)x_{n,t}(\lambda)' = \begin{bmatrix} \frac{1}{\lambda n} \sum_{t=1}^{[\lambda n]} x_t x'_t & 0 \\ 0 & (1-\lambda) \frac{1}{(1-\lambda) n} \sum_{t=[\lambda n]+1}^{n} x_t x'_t \end{bmatrix} \]

\[ \frac{1}{n} \sum_{t=1}^{n} x_{n,t}(\lambda)x_{n,t}(\lambda)' I (t \leq [n\lambda]) = \begin{bmatrix} \frac{1}{\lambda n} \sum_{t=1}^{[\lambda n]} x_t x'_t & 0 \\ 0 & 0 \end{bmatrix}. \]

Hence, by Lemma D.1.a:

\[ \sup_{\lambda \in \Lambda} \left\| \frac{1}{n} \sum_{t=1}^{n} \begin{bmatrix} x_{n,t}(\lambda)x_{n,t}(\lambda)' \end{bmatrix} - \begin{bmatrix} E[x_t x'_t] & 0 \\ 0 & E[x_t x'_t] \end{bmatrix} \right\|_p \to 0 \quad (34) \]

\[ \sup_{\lambda \in \Lambda} \left\| \frac{1}{n} \sum_{t=1}^{n} \begin{bmatrix} x_{n,t}(\lambda)x_{n,t}(\lambda)' I (t \leq [n\lambda]) \end{bmatrix} - \begin{bmatrix} E[x_t x'_t] & 0 \\ 0 & 0 \end{bmatrix} \right\|_p \to 0. \]

Furthermore, by definition and the stated assumptions \(1/n \sum_{t=1}^{n} E[x_t x'_t] \to J\).

The first uniform probability limit in (34) and weak convergence Lemma D.1.b ensure

\[ \left\| C_n(\lambda) - \begin{bmatrix} \frac{1}{\lambda} \sqrt{V/2} & 0 \\ 0 & \frac{1}{1-\lambda} \sqrt{V/2} \end{bmatrix} \left[ \frac{1}{\sigma} J^{-1/2} \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{\sigma} J^{-1/2} \end{bmatrix} \right] \frac{1}{\sqrt{n}} \sum_{t=1}^{n} x_{n,t}(\lambda)\epsilon_t \right\|_p \to 0, \]

hence Lemma D.1.b yields:

\[ \{C_n(\lambda)\} \Rightarrow \begin{bmatrix} \frac{1}{\lambda} \sqrt{V/2} & 0 \\ 0 & \frac{1}{1-\lambda} \sqrt{V/2} \end{bmatrix} \left[ \frac{1}{\sigma} J^{-1/2} \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{\sigma} J^{-1/2} \end{bmatrix} \right] \{B_p(\lambda) - B_p(1) \} \quad (35) \]

Further, uniform laws of large numbers (34), and

\[ \begin{bmatrix} \frac{1}{n} \sum_{t=1}^{[\lambda n]} E[x_t x'_t] & 0 \\ 0 & \frac{1}{n} \sum_{t=[\lambda n]+1}^{n} E[x_t x'_t] \end{bmatrix} \begin{bmatrix} \frac{1}{n} \sum_{t=1}^{[\lambda n]} E[x_t x'_t] & 0 \\ 0 & 0 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}. \]
imply
\[
\sup_{\lambda \in \Lambda} \left\| D_n(\lambda) - \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \right\| \overset{p}{\to} 0. \tag{36}
\]

Combine (33), (35) and (36) to deduce:
\[
\sqrt{n} \left( \hat{\theta}_n(\lambda) - \theta_0 \right) \Rightarrow \left\{ \begin{bmatrix} \frac{1}{\lambda} \gamma^{1/2} & 0 \\ 0 & \frac{1}{1-\lambda} \gamma^{1/2} \end{bmatrix} \times \begin{bmatrix} B_p(\lambda) \\ B_p(1) - B_p(\lambda) \end{bmatrix} \right\} + \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \eta
\]
\[
= \left\{ \begin{bmatrix} \frac{1}{\lambda} \gamma^{1/2} B_p(\lambda) + \eta \\ \frac{1}{1-\lambda} \gamma^{1/2} (B_p(1) - B_p(\lambda)) \end{bmatrix} \right\}.
\]

This completes the proof. \textit{QED}.

Define \( m_t(\beta) \equiv (y_t - \beta'x_t)x_t \) and:

\[
S_{n,1}(\beta, \lambda) \equiv \frac{1}{\lambda n} \sum_{t=1}^{[\lambda n]} m_{n,t,1}(\beta, \lambda)m_{n,t,1}(\beta, \lambda)'
\]
\[
S_{n,2}(\beta, \lambda) \equiv \frac{1}{(1-\lambda) n} \sum_{t=[\lambda n]+1}^{n} m_{n,t,2}(\beta, \lambda)m_{n,t,2}(\beta, \lambda)'
\]
\[
\hat{S}_{n,1}(\beta, \lambda) \equiv \frac{1}{\lambda n} \sum_{t=1}^{[\lambda n]} E [m_t(\beta)m_t(\beta)'] - \left( \frac{1}{\lambda n} \sum_{t=1}^{[\lambda n]} E [m_t(\beta)] \right) \left( \frac{1}{\lambda n} \sum_{t=1}^{[\lambda n]} E [m_t(\beta)] \right)'
\]
\[
\hat{S}_{n,2}(\beta, \lambda) \equiv \frac{1}{(1-\lambda) n} \sum_{t=[\lambda n]+1}^{n} E [m_t(\beta)m_t(\beta)']
\]
\[
- \left( \frac{1}{(1-\lambda) n} \sum_{t=[\lambda n]+1}^{n} E [m_t(\beta)] \right) \left( \frac{1}{(1-\lambda) n} \sum_{t=[\lambda n]+1}^{n} E [m_t(\beta)] \right)'.
\]

\textbf{Lemma D.3.} Under the conditions of Theorem 4.4, \( \sup_{\lambda \in \Lambda, \beta \in \mathcal{B}} \| S_{n,i}(\beta, \lambda) - \hat{S}_{n,i}(\beta, \lambda) \| \overset{p}{\to} 0. \)

\textbf{Proof.} Assume \( x_t \) and \( \beta \) are scalars to reduce notation. We prove the claim for \( S_{n,1}(\beta, \lambda) \), the proof for \( S_{n,2}(\beta, \lambda) \) being similar. Expand:

\[
S_{n,1}(\beta, \lambda) - \hat{S}_{n,1}(\beta, \lambda)
\]
\[
\frac{1}{\lambda n} \sum_{t=1}^{[\lambda n]} \left( m_t(\beta) - \frac{1}{\lambda n} \sum_{t=1}^{[\lambda n]} m_t(\beta) \right)^2 - \frac{1}{\lambda n} \sum_{t=1}^{[\lambda n]} E \left[ m_t^2(\beta) \right] + \left( \frac{1}{\lambda n} \sum_{t=1}^{[\lambda n]} E \left[ m_t(\beta) \right] \right)^2 \\
= \frac{1}{\lambda n} \sum_{t=1}^{[\lambda n]} m_t^2(\beta) - \left( \frac{1}{\lambda n} \sum_{t=1}^{[\lambda n]} m_t(\beta) \right)^2 - \frac{1}{\lambda n} \sum_{t=1}^{[\lambda n]} E \left[ m_t^2(\beta) \right] + \left( \frac{1}{\lambda n} \sum_{t=1}^{[\lambda n]} E \left[ m_t(\beta) \right] \right)^2 \\
= \frac{1}{\lambda n} \sum_{t=1}^{[\lambda n]} \left\{ m_t^2(\beta) - E \left[ m_t^2(\beta) \right] \right\} - \left\{ \left( \frac{1}{\lambda n} \sum_{t=1}^{[\lambda n]} m_t(\beta) \right)^2 - \left( \frac{1}{\lambda n} \sum_{t=1}^{[\lambda n]} E \left[ m_t(\beta) \right] \right)^2 \right\} \\
= \frac{1}{\lambda n} \sum_{t=1}^{[\lambda n]} \left\{ m_t^2(\beta) - E \left[ m_t^2(\beta) \right] \right\} \\
- \left\{ \frac{1}{\lambda n} \sum_{t=1}^{[\lambda n]} m_t(\beta) - \frac{1}{\lambda n} \sum_{t=1}^{[\lambda n]} E \left[ m_t(\beta) \right] \right\} \left\{ \frac{1}{\lambda n} \sum_{t=1}^{[\lambda n]} m_t(\beta) + \frac{1}{\lambda n} \sum_{t=1}^{[\lambda n]} E \left[ m_t(\beta) \right] \right\}.
\]

By assumption \( J \equiv \lim_{n \to \infty} 1/n \sum_{t=1}^{n} E[x_t^2] < \infty, \ E[\epsilon_t x_t] = 0, \ \lambda \in \Lambda \) a compact subset of \((0,1),\) and \( \beta \in B \) a compact set. Therefore

\[
\lim_{n \to \infty} \sup_{\lambda \in \Lambda, \beta \in B} \left| \frac{1}{\lambda n} \sum_{t=1}^{[\lambda n]} E \left[ m_t(\beta) \right] \right| \leq K \lim_{n \to \infty} \sup_{\lambda \in \Lambda} \left\{ \frac{1}{\lambda n} \sum_{t=1}^{[\lambda n]} E \left[ x_t^2 \right] \right\} \leq K \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} E \left[ x_t^2 \right] = KJ < \infty.
\]

It therefore suffices to prove

\[
\sup_{\lambda \in \Lambda, \beta \in B} \left| \frac{1}{\lambda n} \sum_{t=1}^{[\lambda n]} \left\{ m_t(\beta) - E \left[ m_t(\beta) \right] \right\} \right| \xrightarrow{p} 0 \quad \text{and} \quad \sup_{\lambda \in \Lambda, \beta \in B} \left| \frac{1}{\lambda n} \sum_{t=1}^{[\lambda n]} \left\{ m_t^2(\beta) - E \left[ m_t^2(\beta) \right] \right\} \right| \xrightarrow{p} 0. \quad (38)
\]

Note that

\[
\frac{1}{\lambda n} \sum_{t=1}^{[\lambda n]} \left\{ m_t(\beta) - E \left[ m_t(\beta) \right] \right\} = \frac{1}{\lambda n} \sum_{t=1}^{[\lambda n]} \epsilon_t x_t + \beta_1 - \beta \frac{1}{\lambda n} \sum_{t=1}^{[\lambda n]} \left\{ x_t^2 - E \left[ x_t^2 \right] \right\}
\]

and

\[
\frac{1}{\lambda n} \sum_{t=1}^{[\lambda n]} \left\{ m_t^2(\beta) - E \left[ m_t^2(\beta) \right] \right\}
\]

30
\[
\frac{1}{\lambda n} \sum_{t=1}^{[\lambda n]} \left\{ \epsilon_t^2 x_t^2 - E [\epsilon_t^2 x_t^2] \right\} + 2 (\beta_1 - \beta) \frac{1}{\lambda n} \sum_{t=1}^{[\lambda n]} \epsilon_t x_t^3 + (\beta_1 - \beta)^2 \frac{1}{\lambda n} \sum_{t=1}^{[\lambda n]} \{ x_t^4 - E [x_t^4] \}.
\]

The uniform laws (38) therefore follow from the assumption that \( \Lambda \subset (0, 1) \) is bounded away from 0, \((\beta_1, \beta) \in \mathcal{B} \) a compact set, \( E|\epsilon_t|x_t| = 0 \) a.s., and Lemma D.1.a. QED.

**Lemma D.4.** Under the conditions of Theorem 4.4 \( \sup_{\lambda \in \Lambda} ||\hat{V}_{n,1}(\lambda) - \mathcal{V}|| \overset{p}{\to} 0 \), where \( \mathcal{V} \equiv \sigma^2 \mathcal{J}^{-1} \).

**Proof.** We prove the claim for \( \hat{V}_{n,1}(\lambda) \), the proof for \( \hat{V}_{n,2}(\lambda) \) being similar. It suffices to prove \( \sup_{\lambda \in \Lambda} ||\hat{J}_{n,1}(\lambda) - \mathcal{J}|| \overset{p}{\to} 0 \) where \( \mathcal{J} \equiv \lim_{n \to \infty} 1/n \sum_{t=1}^{n} E[x_t x_t'] \) is positive definite by assumption, and \( \sup_{\lambda \in \Lambda} ||\hat{S}_{n,1}(\lambda) - \sigma^2 \mathcal{J}|| \overset{p}{\to} 0 \).

By Lemma D.1.a, and the assumption that \( \Lambda \subset (0, 1) \) is bounded away from 0:

\[
\sup_{\lambda \in \Lambda} \left\| \frac{1}{\lambda n} \sum_{t=1}^{[\lambda n]} \{ x_t x_t' - E [x_t x_t'] \} \right\| \leq K \sup_{\lambda \in \Lambda} \left\| \frac{1}{n} \sum_{t=1}^{\lambda n} \{ x_t x_t' - E [x_t x_t'] \} \right\| \overset{p}{\to} 0.
\]

The required limit \( \sup_{\lambda \in \Lambda} ||\hat{J}_{n,1}(\lambda) - \mathcal{J}|| \overset{p}{\to} 0 \) therefore follows from \( \mathcal{J} \equiv 1/n \sum_{t=1}^{n} E[x_t x_t'] \).

Next, for \( \hat{S}_{n,1}(\lambda) \) note that \( E[m_t] = 0 \) and \( E[m_t m_t'] = \sigma^2 E[x_t x_t'] \) by the martingale difference assumptions. Now use \( \sup_{\lambda \in \Lambda} ||\hat{\theta}(\lambda) - \theta_0|| \overset{p}{\to} 0 \) by Lemma D.2 and the mapping theorem, and continuity of \( E[m_t(\cdot)] \) and \( E[m_t(\cdot)m_t(\cdot)] \), to yield

\[
\sup_{\lambda \in \Lambda} \left\| E \left[ m_t(\hat{\theta}_{n,1}(\lambda)) \right] \right\| \to 0
\]

and

\[
\sup_{\lambda \in \Lambda} \left\| E \left[ m_t(\hat{\theta}_{n,1}(\lambda))m_t(\hat{\theta}_{n,1}(\lambda)) - \sigma^2 E [x_t x_t'] \right] \right\| \to 0.
\]

By convergence of Cesàro means, and \( \lambda \in \Lambda \) a compact subset of \( (0, 1) \), it therefore follows that:

\[
\sup_{\lambda \in \Lambda} \left\| \frac{1}{\lambda n} \sum_{t=1}^{[\lambda n]} E \left[ m_t(\hat{\theta}_{n,1}(\lambda)) \right] \right\| \leq K \frac{1}{n} \sum_{t=1}^{n} \sup_{\lambda \in \Lambda} \left\| E \left[ m_t(\hat{\theta}_{n,1}(\lambda)) \right] \right\| \to 0
\]

and

\[
\sup_{\lambda \in \Lambda} \left\| \frac{1}{\lambda n} \sum_{t=1}^{[\lambda n]} E \left[ m_t(\hat{\theta}_{n,1}(\lambda))m_t(\hat{\theta}_{n,1}(\lambda)) - \sigma^2 \hat{J}_{n,1}(\lambda) \right] \right\| \to 0.
\]

31
\[
\begin{align*}
&= \sup_{\lambda \in \Lambda} \left\| \frac{1}{\lambda n} \sum_{t=1}^{\lceil \lambda n \rceil} \left\{ E \left[ m_t(\hat{\beta}_{n,1}(\lambda))m_t(\hat{\beta}_{n,1}(\lambda)) - \sigma^2 E [x_t x_t'] \right] \right\} \right. \\
\leq& \ K \frac{1}{n} \sum_{t=1}^{n} \sup_{\lambda \in \Lambda} \left\| \left\{ E \left[ m_t(\hat{\beta}_{n,1}(\lambda))m_t(\hat{\beta}_{n,1}(\lambda)) - \sigma^2 E [x_t x_t'] \right] \right\} \right. \to 0.
\end{align*}
\]

This implies that for \( \hat{S}_{n,1}(\beta, \lambda) \) defined in (37):
\[
\sup_{\lambda \in \Lambda} \left\| \hat{S}_{n,1}(\hat{\beta}_{n,1}, \lambda) - \sigma^2 \hat{J}_{n,1}(\lambda) \right\| \to 0.
\]

Now combine \( \sup_{\lambda \in \Lambda} \left\| \hat{S}_{n,1}(\lambda) - \hat{S}_{n,1}(\hat{\beta}_{n,1}, \lambda) \right\| \overset{p}{\to} 0 \) by application of Lemma D.3, \( \sup_{\lambda \in \Lambda} \left\| \hat{J}_{n,1}(\lambda) \to 0 \right\| \), and (39) to yield \( \sup_{\lambda \in \Lambda} \left\| \hat{S}_{n,1}(\lambda) - \sigma^2 J \right\| \to 0 \) as required. \( \textbf{QED} \).

**Proof of Theorem 4.4.** Claim \( (a) \) follows from \( (b) \) under \( \eta = 0 \). Consider \( (b) \). By Lemma D.2 and the mapping theorem
\[
\{ \sqrt{n} \left( \hat{\beta}_{n,1}(\lambda) - \hat{\beta}_{n,2}(\lambda) \right) \} \Rightarrow^* \left\{ \frac{1}{\lambda} \sqrt{n} \beta_1 - \frac{1}{1 - \lambda} \sqrt{n} (\beta_1(1) - \beta_1(\lambda)) + \eta \right\},
\]

where
\[
\frac{1}{\lambda} \beta_1 - \frac{1}{1 - \lambda} (\beta_1(1) - \beta_1(\lambda)) = \frac{1}{\lambda (1 - \lambda)} \left\{ (1 - \lambda) \beta_1(\lambda) - \lambda \beta_1(1) \right\}
\]

Therefore
\[
\{ \sqrt{n} \left( \hat{\beta}_{n,1}(\lambda) - \hat{\beta}_{n,2}(\lambda) \right) \} \Rightarrow^* \left\{ \frac{1}{\lambda (1 - \lambda)} \beta_1 + \sqrt{n} \eta \right\}.
\]

Be construction, \( \beta_1(\lambda) \) is independent of \( \beta_1(1) - \beta_1(\lambda) \), hence
\[
E \left[ \left\{ (1 - \lambda) \beta_1(\lambda) - \lambda \beta_1(1) \right\} \{ (1 - \lambda) \beta_1(\lambda) - \lambda \beta_1(1) \} \right]
\]

\[
= (1 - \lambda)^2 \lambda I_{k^2} + \lambda^2 (1 - \lambda) I_{k^2} = \lambda (1 - \lambda) I_{k^2}.
\]

32
Therefore

\[
\left\{ \left( \frac{1}{\lambda (1 - \lambda)} \right)^{1/2} \right\} \sqrt{n} \left( \hat{\beta}_{n,1}(\lambda) - \hat{\beta}_{n,2}(\lambda) \right)
\]

\[
\Rightarrow^{*} \left\{ \frac{1}{\sqrt{\lambda (1 - \lambda)}} \left\{ B_p(\lambda) - \lambda B_p(1) \right\} + \sqrt{\lambda (1 - \lambda)} V^{-1/2} \eta \right\}
\]

where \((B_p(\lambda) - \lambda B_p(1))/\sqrt{\lambda (1 - \lambda)}\) is for each \(\lambda\) an \(N(0, I_{k_p})\) random vector. The claim now follows from Lemma D.4 and the mapping theorem. Q.E.D.
E Figures

Figure E.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 $b \in [0, 7]$
Figure E.2: Test of Omitted Nonlinearity Example p-Value Sample Paths (Occupation Time = ∫Λ I (p_n(λ) < α) dλ for α ∈ {.01, .05, .10})

(a) Null is true: \( y_t \) is iid linear

(b) Null is false: \( y_t \) is iid quadratic

(c) Null is true: \( y_t \) is linear AR

(d) Null is false: \( y_t \) is Self-Exciting AR

The sample size is \( n = 250 \). The models are linear \( y_t = 2x_t + \epsilon_t \) or quadratic \( y_t = 2x_t + .1x_t^2 + \epsilon_t \), where \( \{x_t, \epsilon_t\} \) are iid variables; and AR(1) \( y_t = .9x_t + \epsilon_t \) or Self-Exciting Threshold AR(1) \( y_t = .9x_t - .4x_tI(x_t > 0) + \epsilon_t \), where \( x_t = y_{t-1} \) and \( \epsilon_t \) is iid.
Figure E.3: GARCH Test Example p-Value Sample Paths (Occupation Time $= \int_{\Lambda} I(p_n(\lambda) < \alpha) \, d\lambda$ for $\alpha \in \{.01, .05, .10\}$)

(a) Null is true: $\delta_0 = 0$

(b) Null is true: $\delta_0 = 0$

(c) Null is false: $\delta_0 > 0$

(d) Null is false: $\delta_0 > 0$

The sample size is $n = 250$. The model is $y_t = \sigma_t \epsilon_t$ and $\sigma_t^2 = \omega_0 + \delta_0 y_{t-1}^2 + \lambda_0 \sigma_{t-1}^2$ with parameter values $\omega_0 = 1$, $\lambda_0 = .6$, and $\delta_0 = 0$ or .3, where $\epsilon_t$ is iid. The discretized PVOT with smoothed p-value is $P_n^{(s)}(\alpha, N_n) = 1/(N_n - R_n + 1) \sum_{j=1}^{N_n - R_n + 1} I(p_n^{(s)}(R_n) < \alpha)$ with moving average $p_n^{(s)}(R_n) = 1/R_n \sum_{i=j}^{R_n+j-1} p_n(\lambda_i)$. We use $R_{250} = 35$. p-value (sm) is the smoothed p-value.
Figure E.4: Structural Break Example p-Value Sample Paths (Occupation Time = \( \int_{\lambda} I(p_n(\lambda) < \alpha) d\lambda \) for \( \alpha \in \{.01, .05, .10\} \))

(a) Null is true: \( \beta_t = .4 \) \( \forall t \)

(b) Null is true: \( \beta_t = .4 \) \( \forall t \)

(c) Null is false: \( \beta_t = .4 \) for \( 1 \leq t \leq \lceil \lambda n \rceil \), and \( \beta_t = .6 \) for \( t > \lceil \lambda n \rceil \)

(d) Null is false: \( \beta_t = .4 \) for \( 1 \leq t \leq \lceil \lambda n \rceil \), and \( \beta_t = .6 \) for \( t > \lceil \lambda n \rceil \)

The sample size is \( n = 250 \). The model is \( y_t = \sigma_t \epsilon_t \) and \( \sigma_t^2 = \omega_0 + \delta_0 y_{t-1}^2 + \lambda_0 \sigma_{t-1}^2 \) with parameter values \( \omega_0 = 1, \lambda_0 = .6, \) and \( \delta_0 = 0 \) or .3, where \( \epsilon_t \) is iid. The discretized PVOT with smoothed p-value is \( P_n^{(s)}(\alpha, N_n) = 1/(N_n - R_n + 1) \sum_{j=1}^{N_n - R_n + 1} I(p_{n,j}^{(s)}(R_n) < \alpha) \) with moving average \( p_{n,j}^{(s)}(R_n) = 1/R_n \sum_{i=j}^{R_n+j-1} p_n(\lambda_i) \). We use \( R_{250} = 35 \). p-value (sm) is the smoothed p-value.
F Bibliography


