Supplemental Material for “Weak-Identification Robust Wild Bootstrap applied to a Consistent Model Specification Test”

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

September 18, 2019

Contents

A Outline and Assumptions .......................................................... 3
B Supporting Lemmata ................................................................. 7
C Proof of Theorem 4.1 ................................................................. 20
D Identification Category Selection Type 2 P-Value ......................... 23
E Robust Critical Values ............................................................. 29
   E.1 Least Favorable and ICS Critical Values .................................. 30
      E.1.1 Least Favorable Critical Value .................................... 30
      E.1.2 Identification Category Selection Type 1 ...................... 31
      E.1.3 Identification Category Selection Type 2 ...................... 31
   E.2 Asymptotics for Robust Critical Values ...................................... 32
   E.3 Computation of $c_{1-a,n}(\lambda)$ ........................................ 32
   E.4 Proofs ........................................................................ 35
F Bibliography ......................................................................... 37

List of Tables

1 Identification Cases .................................................................. 30
2 STAR Test Rejection Frequencies: Sample Size $n = 100$, $\sigma = 1$ .. 39
3 STAR Test Rejection Frequencies: Sample Size $n = 250$, $\sigma = 1$ .... 40

∗Dept. of Economics, University of North Carolina, Chapel Hill; www.unc.edu/~jbhill; jbhill@email.unc.edu
STAR Test Rejection Frequencies: Sample Size $n = 500$, $\sigma = 1$
A Outline and Assumptions

Appendix B contains proofs of the supporting lemmata from the main paper. In Appendix C we prove Theorem 4.1. Appendix D details the Identification Category Selection Type 2 [ICS-2] p-value. Appendix E presents bootstrapped identification category robust critical values, with asymptotic theory.

Recall the model

\[ y_t = \zeta_0 x_t + \beta_0 g(x_t, \pi_0) + \epsilon_t = f(\theta_0, x_t) + \epsilon_t \] where \( x_t \in \mathbb{R}^{k_x} \) and \( \theta \equiv [\zeta', \beta', \pi]' \).  \hspace{1cm} (A.1)

The variable \( y_t \) is a scalar, \( x_t \in \mathbb{R}^{k_x} \) are covariates with finite \( k_x \geq 2 \), \( g : \mathbb{R}^{k_x} \times \Pi \to \mathbb{R}^{k_\beta} \) is a known function, and \( \zeta_0 \in \mathcal{Z}, \beta_0 \in \mathcal{B} \) and \( \pi_0 \in \Pi \), where \( \mathcal{B}, \mathcal{Z} \) and \( \Pi \) are compact subsets of \( \mathbb{R}^{k_\beta}, \mathbb{R}^{k_x} \) and \( \mathbb{R}^{k_\pi} \) respectively for finite \( k_\pi \geq 1 \). The covariates \( x_t \) include a constant term and at least one stochastic regressor. Assume \( E[\epsilon_t] = 0 \) and \( E[\epsilon_t^2] \in (0, \infty) \) for some unique \( \theta_0 \in \Theta \equiv \mathcal{Z} \times \mathcal{B} \times \Pi \).

Let \( y_t \) exist on \( (\Omega, \mathcal{P}, \mathcal{F}) \), where \( \mathcal{F} \equiv \sigma(\cup_{t \in \mathbb{Z}} \mathcal{F}_t) \) and \( \mathcal{F}_t \equiv \sigma(y_r : r \leq t) \). Assume \( \Theta \) has the form \( \{ \theta \equiv [\beta', \zeta', \pi]' : \beta \in \mathcal{B}, \zeta \in \mathcal{Z}(\beta), \pi \in \Pi \} \), where \( \mathcal{B}, \mathcal{Z}(\beta) \) for each \( \beta \), and \( \Pi \) are compact subsets. Recall:

\[ \psi \equiv [\beta', \zeta', \pi]' \in \Psi \equiv \{ (\beta, \zeta) : \beta \in \mathcal{B}, \zeta \in \mathcal{Z}(\beta) \} \].

The true parameter space \( \Theta^* = \psi^* \times \Pi^* = \{ \theta \equiv [\beta', \zeta', \pi]' : \beta \in \mathcal{B}^*, \zeta \in \mathcal{Z}^*(\beta), \pi \in \Pi^* \} \) lies in the interior of \( \Theta \), it contains \( \theta_0 \equiv [\beta_0', \zeta_0', \pi_0'] \), and \( 0 \in \mathcal{B}^* \).

Recall definitions and constructions:

\[ \mathcal{B}(\beta) = \begin{bmatrix} I_{k_\psi} & 0_{k_\psi \times 2} \\ 0_{2 \times k_\psi} & \| \beta \| \times I_2 \end{bmatrix} \hspace{1cm} \text{(A.2)} \]

and

\[ d_{\psi, t}(\pi) \equiv [g(x_t, \pi)', x_t']' \] and \( d_{\theta, t}(\omega, \pi) \equiv \left[ g(x_t, \pi)', x_t', \omega' \frac{\partial}{\partial \pi} g(x_t, \pi) \right]' \) and \( d_{\theta, t} \equiv d_{\theta, t}(\omega_0, \pi_0) \)

\[ b_{\psi}(\pi, \lambda) = E \left[ F(\lambda' \mathcal{W}(x_t)) \, d_{\psi, t}(\pi) \right] \]

\[ b_{\theta}(\omega, \pi, \lambda) = E \left[ F(\lambda' \mathcal{W}(x_t)) \, d_{\theta, t}(\omega, \pi) \right] \] and \( b_{\theta}(\lambda) \equiv E \left[ F(\lambda' \mathcal{W}(x_t)) \, d_{\theta, t} \right] \)

\[ \mathcal{H}_{\psi}(\pi) \equiv E \left[ d_{\psi, t}(\pi) \, d_{\psi, t}(\pi)' \right] \] and \( \mathcal{H}_{\theta}(\omega, \pi) \equiv E \left[ d_{\theta, t}(\omega, \pi) \, d_{\theta, t}(\omega, \pi)' \right] \)

\[ \mathcal{H}_{\theta} \equiv \mathcal{H}_{\theta}(\omega_0, \pi_0) = E \left[ d_{\theta, t} \, d_{\theta, t}' \right] \]
\[ \mathcal{K}_{\psi,t}(\pi, \lambda) \equiv F(\lambda'W(x_t)) - \mathbf{b}_{\psi}(\pi, \lambda) \mathcal{H}_{\psi}^{-1}(\pi)d_{\psi,t}(\pi) \]
\[ \mathcal{K}_{\theta,t}(\lambda) \equiv F(\lambda'W(x_t)) - \mathbf{b}_{\theta}(\lambda) \mathcal{H}_{\theta}^{-1}(\beta_n/\|\beta_n\|, \pi_0) \quad \text{and} \quad \mathcal{K}_{\theta,t}(\lambda; a, m) \equiv \sum_{i=1}^{m} \alpha_i \mathcal{K}_{\theta,t}(\lambda_i), \]
and
\[ G_{\psi,n}(\theta) = \sqrt{n} \left\{ \frac{\partial}{\partial \psi} Q_n(\theta) - E \left[ \frac{\partial}{\partial \psi} Q_n(\theta) \right] \right\} = -\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \{ \epsilon_t(\theta)d_{\psi,t}(\pi) - E[\epsilon_t(\theta)d_{\psi,t}(\pi)] \} \]
\[ G_{\theta,n}(\theta) = \mathcal{B}(\beta_n)^{-1} \sqrt{n} \left\{ \frac{\partial}{\partial \theta} Q_n(\theta) - E \left[ \frac{\partial}{\partial \theta} Q_n(\theta) \right] \right\} = -\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \{ \epsilon_t(\theta)d_{\theta,t}(\omega(\beta), \pi) - E[\epsilon_t(\theta)d_{\theta,t}(\omega(\beta), \pi)] \}, \]
and
\[ D_{\psi}(\pi) \equiv -\frac{\partial}{\partial \beta_0} E[\epsilon_t(\theta)d_{\psi,t}(\pi)] = -E[d_{\psi,t}(\pi)g(x_t, \pi_0)'] \quad \text{and} \quad H_{\psi}(\pi) \equiv E[d_{\psi,t}(\pi)d_{\psi,t}(\pi)'], \]
and
\[ \hat{\mathcal{H}}_n = \frac{1}{n} \sum_{t=1}^{n} d_{\theta,t}(\omega(\hat{\beta}_n), \hat{\pi}_n)d_{\theta,t}(\omega(\hat{\beta}_n), \hat{\pi}_n)' \quad \text{where} \quad \omega(\beta) \equiv \left\{ \frac{\beta/\|\beta\|}{1_{k_\beta}/\|1_{k_\beta}\|} \quad \text{if} \quad \beta \neq 0 \right. \]
\[ \begin{align*}
\hat{b}_{\theta,n}(\omega, \pi, \lambda) & = \frac{1}{n} \sum_{t=1}^{n} F(\lambda'W(x_t)) d_{\theta,t}(\omega, \pi) \\
\hat{\epsilon}_t^2(\hat{\theta}_n, \lambda) & = \frac{1}{n} \sum_{t=1}^{n} \{ F(\lambda'W(x_t)) - \hat{b}_{\theta,n}(\omega(\hat{\beta}_n), \hat{\pi}_n, \lambda) \hat{\mathcal{H}}_n^{-1} d_{\theta,t}(\omega(\hat{\beta}_n), \hat{\pi}_n) \}^2 \\
\hat{\mathcal{H}}_n & = \frac{1}{n} \sum_{t=1}^{n} d_{\theta,t}(\omega(\hat{\beta}_n), \hat{\pi}_n)d_{\theta,t}(\omega(\hat{\beta}_n), \hat{\pi}_n) \\
\hat{\mathcal{S}}_n & \equiv \hat{\mathcal{H}}_n^{-1} \hat{\mathcal{S}}_n \hat{\mathcal{H}}_n^{-1}, \nonumber
\end{align*} \]
and
\[ E_{\psi,n}(\pi; a, r) \equiv \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \epsilon_t \sum_{i=1}^{m} \alpha_i r'd_{\psi,t}(\pi_i) \quad \text{and} \quad E_{\theta,n}(\omega; \pi, a, r) \equiv \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \epsilon_t \sum_{i=1}^{m} \alpha_i r'd_{\theta,t}(\omega_i, \pi_i) \]
\[ C_{\mathcal{G}}_{\psi,n}(\lambda; a, r) \equiv r_1 \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \sum_{i=1}^{m} \alpha_i \{ \epsilon_t(\psi_n, \pi_i)\mathcal{K}_{\psi,t}(\pi_i, \lambda_i) - E[\epsilon_t(\psi_n, \pi_i)\mathcal{K}_{\psi,t}(\pi_i, \lambda_i)] \} + r_2 \sum_{i=1}^{m} \alpha_i G_{\psi,n}(\psi_n, \pi_i). \]
Recall the statistic used to determine whether $b$ is finite:

$$A_n \equiv \left( \frac{1}{k_\beta} n^{\beta_\beta} \frac{\hat{\Sigma}_{\beta,\beta,n}^{-1}}{\beta_{\beta,\beta,n}} \right)^{1/2}$$  \hfill (A.6)

where $\hat{\Sigma}_{\beta,\beta,n}$ is the upper $(p+1) \times (p+1)$ block of $\hat{\Sigma}_n$.

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 Euclidean norm; $| \cdot |_p$ is the $L_p$-norm. $K > 0$ is a finite constant whose value may change from place to place. 0 denotes almost everywhere. $\Rightarrow^*$ denotes weak convergence on $l_\infty$, the space of bounded functions with sup-norm topology, in the sense of Hoffman-Jørgensen (1984, 1991), cf. Dudley (1978) and Pollard (1984, 1990).

Recall that by probability subadditivity, for stochastic measurable $(A,B) \geq 0$ and any $a \in (0, \infty)$:

$$P(A + B > a) \leq P(A > a2) + P(B > a/2).$$  \hfill (A.7)

**Assumption 1** (data generating process, test weight).

a. Identification:

(i) Under $H_0$, $E[\varepsilon_t|x_t] = 0$ a.s. and $E[\varepsilon_t^2|x_t] = \sigma_0^2$ a.s., a finite positive constant.

(ii) Under $(i, b)$: $E[ (y_t - \zeta'_0 x_t) d_{\psi,t}(\pi) ] = 0$ for unique $\psi_0 = [0_{k_\beta}, \zeta'_0]'$ in the interior of $\Psi^*$. Under $(ii, \omega_0)$: $E[\varepsilon_t(\theta_0) \times d_{\theta,t}(\omega_0, \pi_0)] = 0$ for unique $\theta_0 = [\beta_0', \zeta_0', \pi_0']'$ in the interior of $\Theta^* = \Psi^* \times \Pi^*$.

b. Memory and Moments: $\{\varepsilon_t, x_t\}$ are $L_p$-bounded for some $p > 6$, strictly stationary, and $\beta$-mixing with mixing coefficients $\beta_i = O(l^{-qp(q-p)-\iota})$ for some $q > p$ and tiny $\iota > 0$.

c. Response $g(x, \pi)$ and Test Weight $F(\lambda W(x))$:

(i) $g(\cdot, \pi)$ is Borel measurable for each $\pi$; $g(\cdot, \pi)$ is twice continuously differentiable in $\pi \in \mathbb{R}^{k_*}$; $g(x_t, \pi)$ is a non-degenerate random variable for each $\pi \in \Pi$.

(ii) $F : \mathbb{R} \to \mathbb{R}$ is analytic, non-polynomial, and $W$ is one-to-one and bounded.

(iii) $E[\sup_{\pi \in \Pi} |(\partial/\partial \pi)^j g(x_t, \pi)|^6] < \infty$ and $E[\sup_{\lambda \in \Lambda} |(\partial/\partial \lambda)^j F(\lambda W(x_t))|^6] < \infty$ for $i = 0, 1, 2$ and $j = 0, 1$.

d. Long-Run Variances:

(i) Under $(i, b)$ with $|b| < \infty$ let $\liminf_{n \to \infty} E[\inf_{a, r, \theta} (r^* \sum_{i=1}^{m} \alpha_i G_{\psi, n}(\theta_i))^2] > 0$ and $\limsup_{n \to \infty} E[\sup_{a, r, \theta} (r^* \sum_{i=1}^{m} \alpha_i G_{\psi, n}(\theta_i))^2] < \infty$.

(ii) Under $(ii, \omega_0)$ let $\liminf_{n \to \infty} E[\inf_{a, r, \theta} (r^* \sum_{i=1}^{m} \alpha_i G_{\theta, n}(\theta_i))^2] > 0$ and $\limsup_{n \to \infty} E[\sup_{a, r, \theta} (r^* \sum_{i=1}^{m} \alpha_i G_{\theta, n}(\theta_i))^2] < \infty$. 

5
(iii) $E[\inf_{r,\omega,\pi}(r'd_{\theta,t}(\omega, \pi))^2] > 0$ and $E[\sup_{r,\omega,\pi}(r'd_{\theta,t}(\omega, \pi))^2] = \infty$; $E[\inf_{r,\pi}(r'd_{\psi,t}(\pi))^2] > 0$ and $E[\sup_{r,\pi}(r'd_{\psi,t}(\pi))^2] < \infty$.

(iv) $\lim \inf_{n \to \infty} \inf_{a,r,\pi} E[\mathcal{E}_{\psi,n}(a, r)] > 0$ and $\lim \sup_{n \to \infty} \sup_{a,r,\pi} E[\mathcal{E}_{\psi,n}(a, r)] < \infty$; and $\lim \inf_{n \to \infty} \inf_{a,r,\omega,\pi} E[\mathcal{E}_{\psi,n}(\omega, \pi; a, r)] > 0$ and $\lim \sup_{n \to \infty} \sup_{a,r,\omega,\pi} E[\mathcal{E}_{\psi,n}(\omega, \pi; a, r)] < \infty$.

(v) Under $\mathcal{C}(i, b)$ with $||b|| < \infty$, $\lim \inf_{n \to \infty} E[\sup_{a,r,\lambda} \mathcal{E}_\psi,\mathcal{E}_{\psi,n}(\lambda; a, r)] < \infty$.

(vi) Under $\mathcal{C}(ii, \omega_0)$, $E[\sup_{a,r,\lambda}(1/\sqrt{n} \sum_{t=1}^{\infty} \epsilon_t K_{\psi,t}(\lambda; a, m))] < \infty$ for each $m$.

e. True Parameter Space:

(i) $\Theta^* \equiv \{ (\beta, \zeta, \pi) : \beta \in \mathcal{B}, \zeta \in \mathcal{Z}(\beta), \pi \in \Pi^* \}$ is compact.

(ii) $0_{\beta, \lambda} \in \text{int}(\mathcal{B}^*)$.

(iii) For some set $\mathcal{Z}_0^*$ and some $\delta > 0$, $\mathcal{Z}^*(\beta) = \mathcal{Z}_0^* \forall ||\beta|| < \delta$.

f. Optimization Parameter Space:

(i) $\Theta \equiv \{ (\beta, \zeta, \pi) : \beta \in \mathcal{B}, \zeta \in \mathcal{Z}(\beta), \pi \in \Pi \}$ and $\Theta^* \subset \text{int}(\Theta)$.

(ii) $(\Theta, \mathcal{B}, \Pi)$ are compact, and $\mathcal{Z}(\beta)$ is compact for each $\beta$. (iii) For some set $\mathcal{Z}_0$ and some $\delta > 0$, $\mathcal{Z}(\beta) = \mathcal{Z}_0 \forall ||\beta|| < \delta$ and $\mathcal{Z}_0^* \subset \text{int}(\mathcal{Z}_0)$.

Assumption 2 (identification of $\pi$). Let drift case $\mathcal{C}(i, b)$ hold with $||b|| < \infty$. (a) Each sample path of the process $\{ \xi_{\psi}(\pi, b) : \pi \in \Pi \}$ in some set $\mathcal{A}(b)$ with $P(\mathcal{A}(b)) = 1$ is minimized over $\Pi$ at a unique point $\pi^*(b)$ that may depend on the sample path. (b) $P(\tau_{\beta}(\pi^*(b), b) = 0) = 0$.

Assumption 3 (non-degenerate scale on $\Lambda$-a.e.).

a. Let $\mathcal{C}(i, b)$ with $||b|| < \infty$ hold. Then $P(E[\inf_{\pi \in \Pi} \{ \epsilon_t^2(\psi_0, \pi) \}|x_i] > 0) = 1$. There exists a Borel measurable function $\mu : \mathbb{R}^{k_x} \to \mathbb{R}$ such that $\kappa_t(\omega, \pi) \equiv [\mu(x_t), d_{\theta,t}^*(\omega, \pi)]'$ has nonsingular $E[\kappa_t(\omega, \pi)\kappa_t(\omega, \pi)']$ uniformly on $\{ \omega \in \mathbb{R}^{k_x} : \omega' \omega = 1 \} \times \Pi$.

b. Let $\mathcal{C}(ii, \omega_0)$ hold. Then $P(E[\epsilon_t^2|x_i] > 0) = 1$. There exists a Borel measurable function $\mu : \mathbb{R}^{k_x} \to \mathbb{R}$ such that $\kappa_t \equiv [\mu(x_t), d_{\theta,t}^*]'$ has a nonsingular $E[\kappa_t\kappa_t^*]$.

Assumption 4 (non-degenerate scale). Let $\inf_{\omega \in \mathbb{R}^{k_x}, \omega' \omega = 1, \pi \in \Pi} v^2(||\beta_0||, \omega, \zeta_0, \pi, \lambda) > 0 \forall \lambda \in \Lambda$ under identification case $\mathcal{C}(i, b)$ with $||b|| < \infty$, and under $\mathcal{C}(ii, \omega_0)$ let $v^2(\theta_0^+, \lambda) > 0 \forall \lambda \in \Lambda$.

Assumption 5 (p-value). a. $\mathcal{F}_{\lambda,h}(c)$ is continuous a.e. on $[0, \infty)$, $\forall h \in \mathcal{H}$. b. The ICS-1 threshold sequence $\{ \kappa_n \}$ satisfies $\kappa_n \to \infty$ and $\kappa_n = o(\sqrt{n})$.

Assumption 6. The test weight $\{ F(w) : w \in \mathbb{R} \}$ and distribution functions $\{ F_{n,\lambda}(c) : \lambda \in \Lambda, c \in [0, \infty) \}$ and $\{ F_{n,\lambda,\psi}(c) : \lambda \in \Lambda, c \in [0, \infty) \}$ belong to the $\mathcal{V}(\mathcal{C})$ class.
B Supporting Lemmata

All subsequent Gaussian processes have almost surely uniformly continuous and bounded sample paths, hence in many case we just say Gaussian process. Let \( l(A) \) and \( l(A) \) denote the minimum and maximum eigenvalue of matrix \( A \).

**Lemma B.1.** Under \( \mathcal{C}(i,b) \) and Assumption 1, \( \{ G_{\psi,n}(\theta) : \theta \in \Theta \} \Rightarrow^{*} \{ G_{\psi}(\theta) : \theta \in \Theta \} \), a zero mean Gaussian process with almost surely uniformly continuous and bounded sample paths and covariance \( E[G_{\psi}(\theta)G_{\psi}(\theta')] \), \( ||E[G_{\psi}(\theta)G_{\psi}(\theta')]|| < \infty \).

**Proof.** Recall \( \Theta \) is compact and therefore bounded. Weak convergences to a Gaussian process with almost surely uniformly continuous and bounded sample paths therefore requires convergence in finite dimensional distributions, and stochastic equicontinuity (see, e.g., Dudley, 1978; Pollard, 1990).

Let \( m \in \mathbb{N} \), \( \alpha \in \mathbb{R}^m \) and \( r \in \mathbb{R}^{k_x+k_y} \) be arbitrary, with \( \alpha'\alpha = 1 \) and \( r'r = 1 \). Under Assumption 1.b,c \( \sum_{i=1}^{m} \alpha_i \epsilon_i(\theta_i) r'd_{\psi,i}(\theta_i) \) is, for any \( m \)-tuple \( \{ \theta_1, \ldots, \theta_m \} \) of points \( \theta_i \) in \( \Theta \), strictly stationary, \( L_\psi \)-bounded, \( p > 4 \), and \( \beta \)-mixing with coefficients \( \beta_l = O(l^{-\frac{p}{(\alpha-q)p})^\iota} \) for some \( \iota > 0 \) and \( q > p \). Hence \( E[(\sum_{i=1}^{m} \alpha_i r'G_{\psi,n}(\theta_i))^2] = O(1) \) (McLeish, 1975, Theorem 1.6, Lemma 2.1). Long run variance Assumption 1.d(i) and Theorem 1.4 in Ibragimov (1962) therefore yield: \( \sum_{i=1}^{m} \alpha_i r'G_{\psi,n}(\theta_i) \rightarrow_{d} N(0, \lim_{n \rightarrow \infty} E[(\sum_{i=1}^{m} \alpha_i r'G_{\psi,n}(\theta_i))^2]) \). Convergence in finite dimensional distributions now follows from the Cramér-Wold theorem.

Stochastic equicontinuity for \( r'G_{\psi,n}(\theta) \) holds if \( \forall (\epsilon, \eta) > 0 \) there exists \( \delta > 0 \) such that:

\[
\lim_{n \rightarrow \infty} P_n(r, \delta, \eta) = \lim_{n \rightarrow \infty} P \left( \sup_{\theta, \tilde{\theta} \in \Theta, ||\theta - \tilde{\theta}|| \leq \delta} \left| r'G_{\psi,n}(\theta) - r'G_{\psi,n}(\tilde{\theta}) \right| > \eta \right) < \epsilon.
\]

(B.8)

We adapt arguments developed in Arcones and Yu (1994, proof of Theorem 2.1 and Lemma 2.1) to prove (B.8). This requires the notion of the \( V-C \) subgraph class of functions, denoted \( \mathcal{V}(\mathcal{C}) \). See Vapnik and Červonenkis (1971) and Dudley (1978, Section 7), and see Pollard (1984, Chapter II.4) for the closely related polynomial discrimination class. By the implication of probability subadditivity (A.7) and \( r'r = 1 \), it suffices to prove the claim for each element of \( G_{\psi,n}(\theta) = [G_{\psi,n,i}(\theta)]_{i=1}^{k_x+k_y} \).

\( G_{\psi,n,i}(\theta) \) lies in \( \mathcal{V}(\mathcal{C}) \) because it is continuous, hence the covering numbers satisfy \( \mathcal{N}(\epsilon, \mathcal{K}, ||\cdot||_2) < \alpha \epsilon^{-b} \) for all \( \epsilon \in (0,1) \) and some \( a, b > 0 \) (e.g. Lemma 7.13 in Dudley, 1978, and Lemma II.25 in Pollard, 1984). Furthermore, under Assumption 1.b,c each \( G_{\psi,n,i}(\theta) \) is \( L_\psi \)-bounded, \( r \equiv p/2 > 2 \), and \( \beta \)-mixing with coefficients \( \beta_l = O(l^{-\frac{q}{(\alpha-q)p})^\iota} \), \( q > p > 6 \) and tiny \( \iota > 0 \). By simple algebra it follows \( \beta_l = O(l^{-\frac{p}{(\alpha-q)p}}) = O(l^{-\frac{p}{(\alpha-q)p}}) \) because \( p/(p-4) < qp/(q - p) \). Therefore \( \{ G_{\psi,n,i}(\theta) : \theta \in \Theta \} \) is stochastically equicontinuous by Lemma 2.1 in Arcones and Yu (1994, see
Lemma B.2. Under $C(i,b)$ and Assumption 1, $\sup_{\pi \in \Pi} ||\hat{H}_{\psi,n}(\pi) - H_{\psi}(\pi)|| \xrightarrow{P} 0$, where $\ell(H_{\psi}(\pi)) > 0$ and $i(H_{\psi}(\pi)) < \infty$ for each $\pi \in \Pi$.

Proof. Pointwise $\hat{H}_{\psi,n}(\pi) \xrightarrow{P} H_{\psi}(\pi)$ under Assumption 1.b,c since $d_{\psi,t}(\kappa)$ is stationary, $L_2$-bounded, and ergodic by the $\beta$-mixing property. Further, $\ell(H_{\psi}(\pi)) > 0$ and $i(H_{\psi}(\pi)) < \infty$ for each $\pi \in \Pi$ respectively follow from $\inf_{r,t=1} E [(r'd_{\psi,t}(\pi))^2] > 0$ under Assumption 1.d(iii), and $||H_{\psi}(\pi)|| < \infty$ under envelope bounds Assumption 1.c and compactness of $\Theta$.

It remains to show $\hat{H}_{\psi,n}(\pi) - H_{\psi}(\pi)$ is stochastically equicontinuous. By the mean-value-theorem and Cauchy-Schwartz inequality:

$$E \left[ \sup_{\pi,\tilde{\pi} \in \Pi: ||\pi - \tilde{\pi}|| \leq \delta} \left| \hat{H}_{\psi,n}(\pi) - \hat{H}_{\psi,n}(\tilde{\pi}) \right| \right]$$

$$\leq 2E \left[ \sup_{\pi \in \Pi} \frac{\partial}{\partial \pi} d_{\psi,t}(\pi) \sup_{\pi \in \Pi} |d_{\psi,t}(\pi)'| \right] \times \delta$$

$$\leq 2 \left( E \left[ \sup_{\pi \in \Pi} \left| \frac{\partial}{\partial \pi} g(x_t, \pi) \right|^2 \right] \right)^{1/2} \left( E \left[ \left( \sup_{\pi \in \Pi} |g(x_t, \pi)| + |x_t| \right)^2 \right] \right)^{1/2} \times \delta \equiv \mathcal{K}\delta,$$

where $\mathcal{K} \geq 0$ is implicitly defined and $\delta > 0$. The right hand side is bounded by $L_2$-boundedness of $x_t$, $\sup_{\pi \in \Pi} |g(x_t, \pi)|$ and $\sup_{\pi \in \Pi} (\partial/\partial \pi) g(x_t, \pi)$ under Assumption 1.b,c. Hence $\mathcal{K} \in [0, \infty)$. Therefore, assuming $\mathcal{K} > 0$, $\forall(\epsilon, \eta) > 0$ there exists $\delta$, $0 < \delta < \epsilon/\mathcal{K}$, such that by Markov’s inequality

$$\lim_{n \to \infty} P \left( \sup_{\pi,\tilde{\pi} \in \Pi: ||\pi - \tilde{\pi}|| \leq \delta} \left\{ \hat{H}_{\psi,n}(\pi) - H_{\psi}(\pi) \right\} - \left\{ \hat{H}_{\psi,n}(\tilde{\pi}) - H_{\psi}(\tilde{\pi}) \right\} > \eta \right) < \epsilon. \quad (B.9)$$

If $\mathcal{K} = 0$ then $\forall(\epsilon, \eta) > 0$ and any $\delta \in (0, \infty)$ (B.9) holds. This yields stochastic equicontinuity, completing the proof. $\mathcal{QED}$

Lemma B.3. Under $C(ii,\omega_0)$ and Assumption 1, $\{G_{\theta,n}(\theta) : \theta \in \Theta\} \Rightarrow^{*} \{G_{\theta}(\theta) : \theta \in \Theta\}$, a zero mean Gaussian process with almost surely uniformly continuous and bounded sample paths.

Proof. The arguments used to prove Lemma B.1 carry over verbatim, except long run variance Assumption 1.d(ii) is used in place of Assumption 1.d(i). $\mathcal{QED}$

Corollary B.4. Let $\theta_n \equiv [\beta_n', \zeta_0', \pi_0']'$ be the sequence of true values under local drift $\{\beta_n\}$. Under $C(ii,\omega_0)$ and Assumption 1, $\sqrt{n}\mathfrak{B}(\beta_n)^{-1}(\partial/\partial \theta)Q_n(\theta_n) \xrightarrow{d} G_0$, a zero mean Gaussian law with a
finite, positive definite covariance $E[G_\theta G_\theta']$, and has a version that has almost surely uniformly continuous and bounded sample paths. Moreover, $E[G_\theta G_\theta'] = \sigma^2 E[d_{\theta,t}d'_{\theta,t}]$ under $H_0$.

**Proof.** By the definition of $G_{\theta,n}(\theta_n)$:

$$\sqrt{n} \mathbb{B}((\beta_n)^{-1} \frac{\partial}{\partial \theta} Q_n(\theta_n)) = G_{\theta,n}(\theta_n) + \sqrt{n} E[\epsilon_t(\theta_n)d_{\theta,t}(\beta_n/\|\beta_n\|,\pi_0)].$$

Combine Lemma B.3, $\theta_n \to \theta_0$, the fact that $\theta_n$ is non-random, and continuity to yield $G_{\theta,n}(\theta_n) \overset{d}{\to} G_\theta \equiv G_\theta(\theta_0)$. By identification Assumption 1.a(ii) and the fact that $\theta_n \equiv [\beta_n',\zeta_0',\pi_0]'$ is the sequence of true values under local drift $\{\beta_n\}$, it follows that $E[\epsilon_t(\theta_n)d_{\theta,t}(\beta_n/\|\beta_n\|,\pi_0)] = 0$. This proves $\sqrt{n} \mathbb{B}((\beta_n)^{-1}(\partial/\partial \theta) Q_n(\theta_n)) \overset{d}{\to} G_\theta$.

Finally, since $\theta_n \equiv [\beta_n',\zeta_0',\pi_0]'$ is the sequence of true values, under $H_0$, note that

$$G_{\theta,n}(\theta_n) = -\frac{1}{\sqrt{n}} \sum_{t=1}^n \left\{ \epsilon_t d_{\theta,t}(\omega(\beta_n),\pi) - E[\epsilon_t(\theta_n)d_{\theta,t}(\omega(\beta_n),\pi)] \right\}$$

$$= -\frac{1}{\sqrt{n}} \sum_{t=1}^n \epsilon_t d_{\theta,t}(\omega(\beta_n),\pi).$$

Hence in view of stationarity $E[G_{\theta,n}(\theta_n)G_{\theta,n}(\theta_n)'] = \sigma^2 E[d_{\theta,t}(\beta_n/\|\beta_n\|,\pi_0)d_{\theta,t}(\beta_n/\|\beta_n\|,\pi_0)'].$

Since $\beta_n/\|\beta_n\| \to \omega_0$ and $\|\omega_0\| = 1$, under Assumption 1.b,c:

$$\mathfrak{d}_t \equiv \limsup_{n \to \infty} \sup_{r' \in \mathbb{R}} \left( r' \left[ g(x_t,\pi)', x_t, \frac{\beta_n'}{\|\beta_n\|} \frac{\partial}{\partial \pi} g(x_t,\pi) \right] \right)^2$$

exists and $E[\mathfrak{d}_t] < \infty$. Dominated convergence now yields $E[d_{\theta,t}(\beta_n/\|\beta_n\|,\pi_0)d_{\theta,t}(\beta_n/\|\beta_n\|,\pi_0)'] \to E[d_{\theta,t}d'_{\theta,t}]$, hence $E[G_{\theta,n}(\theta_n)G_{\theta,n}(\theta_n)'] \to \sigma^2 E[d_{\theta,t}d'_{\theta,t}]$. This implies $\sqrt{n} \mathbb{B}((\beta_n)^{-1}(\partial/\partial \theta) Q_n(\theta_n)) \overset{d}{\to} G_\theta$, with asymptotic variance $\sigma^2 E[d_{\theta,t}d'_{\theta,t}]. \ \Box$

**Lemma B.5.** Under $C(ii,\omega_0)$ and Assumption 1, $\hat{\mathcal{H}}_n \overset{p}{\to} \mathcal{H}_\theta$, and $i(\mathcal{H}_\theta) > 0$ and $i'(\mathcal{H}_\theta) < \infty$.

**Proof.** By the construction of $\hat{\mathcal{H}}_n \equiv 1/n \sum_{t=1}^n d_{\theta,t}(\omega(\beta_n),\pi_0)d_{\theta,t}(\omega(\beta_n),\pi_0)'$ and $\mathcal{H}_\theta \equiv E[d_{\theta,t}d'_{\theta,t}]$, and $d_{\theta,t}(\omega,\pi) \equiv [g(x_t,\pi)', x_t, \omega'(\partial/\partial \pi)g(x_t,\pi)'].$ after adding and subtracting like terms, we have for any $r = [r'_\beta, r'_\pi]'$, $r_\beta \in \mathbb{R}^{k\beta}, r_\pi \in \mathbb{R}^{k\pi}$:

$$r' (\hat{\mathcal{H}}_n - \mathcal{H}_\theta) r = \frac{1}{n} \sum_{t=1}^n \left( r'_\beta g(x_t,\pi_0) + r'_\pi x_t + r'_\pi \frac{\partial}{\partial \pi} g(x_t,\pi_0) \right)^2$$

$$- E \left( \left( r'_\beta g(x_t,\pi_0) + r'_\pi x_t + r'_\pi \frac{\partial}{\partial \pi} g(x_t,\pi_0) \right)^2 \right).$$

9
\begin{align*}
+ \frac{1}{n} \sum_{t=1}^{n} \left( r'_\pi \frac{\partial}{\partial \pi} g(x_t, \pi_0) \left( \frac{\beta_n}{\|\beta_n\|} - \omega_0 \right) \right)^2 \\
+ 2 \frac{1}{n} \sum_{t=1}^{n} \left( r'_\pi g(x_t, \pi_0) + r'_x x_t + r'_\pi \frac{\partial}{\partial \pi} g(x_t, \pi_0) \omega_0 \right) \times r'_\pi \frac{\partial}{\partial \pi} g(x_t, \pi_0) \left( \frac{\beta_n}{\|\beta_n\|} - \omega_0 \right).
\end{align*}

The Assumption 1.b,c envelop moment and mixing properties imply each summand is a summation of stationary, ergodic and integrable random variables. Further \( \beta_n/\|\beta_n\| - \omega_0 \to 0 \) by assumption. The ergodic theorem now yields \( r'(\hat{\mathcal{H}}_n - \mathcal{H}_\theta) r \overset{p}{\to} 0 \).

Finally, \( \xi(\mathcal{H}_\theta) > 0 \) and \( \bar{\iota}(\mathcal{H}_\theta) < \infty \) follow from Assumption 1.c,d(iii). \( \mathcal{QED} \)

Define the augmented parameter, and its space:

\[ \theta^+ \equiv \left[ \|\beta\|, \omega', \zeta', \pi' \right]' \]

\[ \Theta^+ \equiv \left\{ \theta^+ \in \mathbb{R}^{k_x+k_\zeta+k_\pi+1} : \theta^+ = (\|\beta\|, \omega(\beta), \zeta, \pi) : \beta \in \mathcal{B}, \zeta \in \mathcal{Z}(\beta), \pi \in \Pi \right\}. \]

Define

\[ \epsilon_\iota(\theta^+) \equiv y_t - \zeta' x_t - \|\beta\| \omega' g(x_t, \pi), \]

and:

\[ \hat{\mathcal{H}}_n(\theta^+) \equiv \frac{1}{n} \sum_{t=1}^{n} d_{\theta,t}(\omega(\beta), \pi) d_{\theta,t}(\omega(\beta), \pi)', \hat{\mathcal{V}}_n(\theta^+) \equiv \frac{1}{n} \sum_{t=1}^{n} \epsilon_\iota^2(\theta^+) d_{\theta,t}(\omega(\beta), \pi) d_{\theta,t}(\omega(\beta), \pi)'. \]

Hence \( \hat{\mathcal{H}}_n(\theta^+) = \hat{\mathcal{H}}_n \) and \( \hat{\mathcal{V}}_n(\theta^+) = \hat{\mathcal{V}}_n \). Define

\[ \mathcal{H}_\theta(\theta^+) \equiv E [d_{\theta,t}(\omega, \pi) d_{\theta,t}(\omega, \pi)'] \quad \text{and} \quad \mathcal{V}(\theta^+) \equiv E \left[ \epsilon_\iota^2(\theta^+) d_{\theta,t}(\omega, \pi) d_{\theta,t}(\omega, \pi)' \right]. \]

In the interest of decreasing (some) notation we use the same argument \( \theta^+ \) for both \( \hat{\mathcal{H}}_n(\theta^+) \) and \( \hat{\mathcal{V}}_n(\theta^+) \), although \( \hat{\mathcal{H}}_n(\theta^+) \) only depends on \( (\omega(\beta), \pi) \).

**Lemma B.6.** Under Assumption 1 \( \sup_{\theta^+ \in \Theta^+} \|\hat{\mathcal{H}}_n(\theta^+)-\mathcal{H}_\theta(\theta^+)\| \overset{p}{\to} 0 \), \( \sup_{\pi \in \Pi} \|\hat{\mathcal{D}}_{\psi,n}(\pi, \pi_0) - \mathcal{D}_{\psi}(\pi)\| \overset{p}{\to} 0 \), and \( \sup_{\theta^+ \in \Theta^+, \pi} \|\hat{\mathcal{V}}_n(\theta^+) - \mathcal{V}(\theta^+)\| \overset{p}{\to} 0 \), where \( \inf_{\theta^+ \in \Theta^+} \mathcal{H}_\theta(\theta^+) > 0 \), \( \bar{\iota}(\mathcal{H}_\theta) < \infty \), \( \inf_{\theta^+ \in \Theta^+} \bar{\iota}(\mathcal{V}(\theta^+)) > 0 \), and \( \bar{\iota}(\mathcal{V}_\theta) < \infty \).

**Proof.** We prove the claim for \( \hat{\mathcal{V}}_n(\theta^+) \), the proofs for \( \hat{\mathcal{H}}_n(\theta^+) \) and \( \hat{\mathcal{D}}_{\psi,n}(\pi, \pi_0) \) being similar. Pointwise convergence follows from mixing (hence ergodicity) and moment properties in Assumption 1.b,c.

Uniform convergence is proven if we show stochastic equicontinuity: \( \forall (\epsilon, \eta) > 0 \) there exists
\( \delta > 0 \) such that:

\[
\lim_{n \to \infty} \mathcal{P}_n(r, \delta, \eta) = \lim_{n \to \infty} P \left( \sup_{\theta^+, \hat{\theta}^+ \in \Theta^+ : |\theta^+ - \hat{\theta}^+| \leq \delta} \left| \frac{\hat{V}_n(\theta^+) - V(\theta^+)}{\hat{V}_n(\hat{\theta}) - V(\hat{\theta}^+)} \right| > \eta \right) < \epsilon.
\]

First note that:

\[
E \left[ \sup_{\theta^+, \hat{\theta}^+ \in \Theta^+ : |\theta^+ - \hat{\theta}^+| \leq \delta} \left| \frac{\hat{V}_n(\theta^+) - \hat{V}_n(\hat{\theta})}{\hat{V}_n(\theta^+) - \hat{V}_n(\hat{\theta}^+)} \right| \right]
\]

\[
= \sup_{\theta^+, \hat{\theta}^+ \in \Theta^+ : |\theta^+ - \hat{\theta}^+| \leq \delta} \left| \frac{1}{n} \sum_{t=1}^{n} \left\{ \epsilon_t^2(\theta^+) d_{\theta,t}(\omega, \pi) d_{\theta,t}(\omega, \pi)' - \epsilon_t^2(\hat{\theta}^+) d_{\theta,t}(\hat{\omega}, \hat{\pi}) d_{\theta,t}(\hat{\omega}, \hat{\pi})' \right\} \right|
\]

\[
\leq \sup_{\theta^+, \hat{\theta}^+ \in \Theta^+ : |\theta^+ - \hat{\theta}^+| \leq \delta} \left| \frac{1}{n} \sum_{t=1}^{n} \left\{ \epsilon_t^2(\theta^+) - \epsilon_t^2(\hat{\theta}^+) \right\} d_{\theta,t}(\omega, \pi) d_{\theta,t}(\omega, \pi)' \right|
\]

\[
+ \sup_{\theta^+, \hat{\theta}^+ \in \Theta^+ : |\theta^+ - \hat{\theta}^+| \leq \delta} \left| \frac{1}{n} \sum_{t=1}^{n} \epsilon_t^2(\hat{\theta}^+) \left\{ d_{\theta,t}(\omega, \pi) d_{\theta,t}(\omega, \pi)' - d_{\theta,t}(\hat{\omega}, \hat{\pi}) d_{\theta,t}(\hat{\omega}, \hat{\pi})' \right\} \right|.
\]

By the mean value theorem, and the moment properties of Assumption 1.b,c:

\[
E \left[ \sup_{\theta^+, \hat{\theta}^+ \in \Theta^+ : |\theta^+ - \hat{\theta}^+| \leq \delta} \left| \frac{1}{n} \sum_{t=1}^{n} \left\{ \epsilon_t^2(\theta^+) - \epsilon_t^2(\hat{\theta}^+) \right\} d_{\theta,t}(\omega, \pi) d_{\theta,t}(\omega, \pi)' \right| \right]
\]

\[
\leq 2E \left[ \sup_{\theta^+ \in \Theta^+} |\epsilon_t(\theta^+)| \sup_{\theta^+ \in \Theta^+} |d_{\theta,t}(\omega, \pi)|^3 \right] \times \delta \leq K \delta,
\]

and

\[
E \left[ \sup_{\theta^+, \hat{\theta}^+ \in \Theta^+ : |\theta^+ - \hat{\theta}^+| \leq \delta} \left| \frac{1}{n} \sum_{t=1}^{n} \epsilon_t^2(\hat{\theta}^+) \left\{ d_{\theta,t}(\omega, \pi) d_{\theta,t}(\omega, \pi)' - d_{\theta,t}(\hat{\omega}, \hat{\pi}) d_{\theta,t}(\hat{\omega}, \hat{\pi})' \right\} \right| \right]
\]

\[
\leq 2E \left[ \sup_{\theta^+ \in \Theta^+} |\epsilon_t^2(\theta^+) \sup_{\theta^+ \in \Theta^+} |d_{\theta,t}(\omega, \pi)| \sup_{\theta^+ \in \Theta^+} \left| \frac{\partial}{\partial \theta^+} d_{\theta,t}(\omega, \pi) \right| \right] \times \delta \leq K \delta,
\]

where

\[
\left| \frac{\partial}{\partial \theta^+} d_{\theta,t}(\omega, \pi) \right| \leq 2 \times \left| \frac{\partial}{\partial \pi} g(x_t, \pi) \right| + |\omega| \times \left| \frac{\partial^2}{\partial \pi \partial \pi} g(x_t, \pi) \right|.
\]
A similar set of steps shows
\[
\sup_{\theta^+, \tilde{\theta}^+ \in \Theta^+: \|\theta^+ - \tilde{\theta}^+\| \leq \delta} \left| V_n(\theta^+) - V_n(\tilde{\theta}) \right|
\]
\[
= \sup_{\theta^+, \tilde{\theta}^+ \in \Theta^+: \|\theta^+ - \tilde{\theta}^+\| \leq \delta} \left| E \left[ \epsilon_t^2(\theta^+) d_{\theta,t}(\omega, \pi_0) d_{\theta,t}(\omega, \pi_0)' \right] - E \left[ \epsilon_t^2(\tilde{\theta}^+) d_{\theta,t}(\tilde{\omega}, \tilde{\pi}) d_{\theta,t}(\tilde{\omega}, \tilde{\pi})' \right] \right|
\]
\[
\leq K \delta.
\]
Now invoke Markov’s and Minkowski inequalities to yield:
\[
\lim_{n \to \infty} P \left( \sup_{\theta^+, \tilde{\theta}^+ \in \Theta^+: \|\theta^+ - \tilde{\theta}^+\| \leq \delta} \left| \left\{ \hat{V}_n(\theta^+) - V(\theta^+) \right\} - \left\{ \hat{V}_n(\tilde{\theta}) - V(\tilde{\theta}) \right\} \right| > \eta \right)
\]
\[
\leq \lim_{n \to \infty} \frac{1}{\eta} E \left[ \sup_{\theta^+, \tilde{\theta}^+ \in \Theta^+: \|\theta^+ - \tilde{\theta}^+\| \leq \delta} \left| \hat{V}_n(\theta^+) - \hat{V}_n(\tilde{\theta}) \right| \right]
\]
\[
+ \lim_{n \to \infty} \frac{1}{\eta} \sup_{\theta^+, \tilde{\theta}^+ \in \Theta^+: \|\theta^+ - \tilde{\theta}^+\| \leq \delta} \left| \left\{ V(\theta^+) \right\} - V(\tilde{\theta}) \right| \n\]
\[
\leq K \delta.
\]
This proves stochastic equicontinuity (B.10) for any \( \delta \) such that \( 0 < \delta < \epsilon/K \). \( \text{QED} \)

Define
\[
a_n \equiv \begin{cases} \sqrt{n} & \text{if } C(i, b) \text{ and } \|b\| < \infty \\ \|\beta_n\|^{-1} & \text{if } C(i, b) \text{ and } \|b\| = \infty \end{cases}
\]
Recall
\[
\psi_{0,n} \equiv \left[ o_{k,n}' \right]_0,'
\]
hence \( Q_{0,n} \equiv Q_n(\psi_{0,n}, \pi) \) does not depend on \( \pi \). Define:
\[
Z_n(\pi) = -a_n \hat{H}_{\psi,n}^{-1}(\pi) \frac{\partial}{\partial \psi} Q_n(\psi_{0,n}, \pi).
\]
Under \( C(i, b) \), Lemma B.2 yields that \( \hat{H}_{\psi,n}(\pi) \) is positive definite uniformly on \( \Pi \), asymptotically with probability approaching one. Write \( Q_n^c(\pi) \equiv Q_n(\hat{\psi}(\pi), \pi) \).

**Lemma B.7.** Let drift case \( C(i, b) \) and Assumption 1 hold.

a. In general \( a_n(\hat{\psi}(\pi) - \psi_{0,n}) = Z_n(\pi). \)

b. \( a_n^2 \left\{ Q_n^c(\pi) - Q_{0,n} \right\} = -2^{-1} Z_n(\pi)' \hat{H}_{\psi,n}(\pi) Z_n(\pi) \) where \( Q_{0,n} \equiv Q_n(\psi_{0,n}, \pi) \).

**Proof.**
Claim a. By the definition of $\hat{\psi}_n(\pi), 0 = 1/n \sum_{t=1}^n \epsilon_t(\hat{\psi}_n(\pi), \pi) d_{\psi,t}(\pi)$. Now use $(\partial/\partial \psi)Q_n(\psi_{0,n}, \pi) = -1/n \sum_{t=1}^n \epsilon_t(\psi_{0,n}, \pi)d_{\psi,t}(\pi), \hat{\psi}_{\psi,n}(\pi) \equiv 1/n \sum_{t=1}^n d_{\psi,t}(\pi)d_{\psi,t}(\pi)'$, and linearity of the first order equation in $\hat{\psi}_n(\pi)$, to yield the desired result.

Claim b. The equality $x^2 - y^2 = (x - y)(x + y)$ and rudimentary algebra yield:

$$Q_n^c(\pi) - Q_{0,n} = -\frac{1}{n} \sum_{t=1}^n \epsilon_t(\psi_{0,n}, \pi) d_{\psi,t}(\pi)' \times (\hat{\psi}_n(\pi) - \psi_{0,n})$$

$$+ \frac{1}{2} (\hat{\psi}_n(\pi) - \psi_{0,n}) \times \hat{H}_{\psi,n}(\pi) \times (\hat{\psi}_n(\pi) - \psi_{0,n})$$

$$= -\frac{\partial}{\partial \psi} Q_n(\psi_{0,n}, \pi)' \times (\hat{\psi}_n(\pi) - \psi_{0,n}) + \frac{1}{2} (\hat{\psi}_n(\pi) - \psi_{0,n}) \times \hat{H}_{\psi,n}(\pi) \times (\hat{\psi}_n(\pi) - \psi_{0,n}).$$

Use (a) and the form of $Z_n(\pi)$ to deduce $a_n(\partial/\partial \psi)Q_n(\psi_{0,n}, \pi)'Z_n(\pi) = Z_n(\pi)\hat{H}^{-1}_{\psi,n}(\pi)Z_n(\pi)$ hence:

$$a_n^2 \{ Q_n^c(\pi) - Q_{0,n} \} = -a_n \frac{\partial}{\partial \psi} Q_n(\psi_{0,n}, \pi)'Z_n(\pi) + \frac{1}{2} Z_n(\pi)'\hat{H}_{\psi,n}(\pi)Z_n(\pi)$$

$$= -\frac{1}{2} Z_n(\pi) \times \hat{H}^{-1}_{\psi,n}(\pi) \times Z_n(\pi).$$

This proves the claim and completes the proof. QED

Define

$$\vartheta_\psi(\pi, \omega_0) \equiv -2^{-2} \omega_0' D_\psi(\pi)' H^{-1}_\psi(\pi) D_\psi(\pi) \omega_0$$

where

$$D_\psi(\pi) = -E [d_{\psi,t}(\pi)g(x_t, \pi_0)'].$$

Recall from the main the paper:

$$\xi_\psi(\pi, b) \equiv -\frac{1}{2} \{ G_\psi(\psi_{0,n}, \pi) + D_\psi(\pi)b \}' H^{-1}_\psi(\pi) \{ G_\psi(\psi_{0,n}, \pi) + D_\psi(\pi)b \}$$

The following is a key result for characterizing the asymptotic properties of $\hat{\pi}_n$ under weak identification.

Lemma B.8. Let drift case $C(i, b)$ and Assumption 1 hold.

a. If $||b|| < \infty$ then \(n(Q_n^c(\pi) - Q_{0,n}) : \pi \in \Pi\) \(\Rightarrow^* \{ \xi_\psi(\pi, b) : \pi \in \Pi \}.

b. If $||b|| = \infty$ and $\beta_n/||\beta_n|| \to \omega_0$ for some $\omega_0 \in \mathbb{R}^{k_3}$, $||\omega_0|| = 1$, then

$$\sup_{\pi \in \Pi} \left| \frac{1}{||\beta_n||^2} (Q_n^c(\pi) - Q_{0,n}) - \vartheta_\psi(\pi, \omega_0) \right| \overset{p}{\to} 0.$$
Proof.

Claim a. Recall $\mathcal{G}_{\psi,n}(\theta) = \sqrt{n}\{\partial/(\partial \psi)Q_n(\theta) + E[\epsilon_t(\psi_{0,n}, \pi)d_{\psi,t}(\pi)]\}$. By Lemma B.7.b and $||b|| < \infty$:

\[
n(\frac{C_n}{\sqrt{n}}(\pi), \pi) - Q_{0,n} = -\frac{1}{2}nZ_n(\pi)'\mathcal{H}_{\psi,n}(\pi)Z_n(\pi) = -\frac{1}{2}\sqrt{n}\frac{\partial}{\partial \psi}Q_n(\psi_{0,n}, \pi)'\mathcal{H}_{\psi,n}^{-1}(\pi)\sqrt{n}\frac{\partial}{\partial \psi}Q_n(\psi_{0,n}, \pi)
\]

\[
= -\frac{1}{2}\left\{\mathcal{G}_{\psi,n}(\psi_{0,n}, \pi) - \sqrt{n}E[\epsilon_t(\psi_{0,n}, \pi)d_{\psi,t}(\pi)]\right\}' \times \mathcal{H}_{\psi,n}^{-1}(\pi)
\]

\[
\times \left\{\mathcal{G}_{\psi,n}(\psi_{0,n}, \pi) - \sqrt{n}E[\epsilon_t(\psi_{0,n}, \pi)d_{\psi,t}(\pi)]\right\}.
\]

Further, by (C.18) in the proof of Theorem 4.1 in Appendix C:

\[
\sup_{\pi \in \Pi} \left|\sqrt{n}E[\epsilon_t(\psi_{0,n}, \pi)d_{\psi,t}(\pi)] + D_\psi(\pi)b\right| \to 0.
\]

Now use Lemma B.1 for $\mathcal{G}_{\psi,n}(\psi_{0,n}, \pi)$, and Lemma B.2 for $\mathcal{H}_{\psi,n}(\pi)$, to prove the claim. 

Claim b. Lemma B.7.b and the definition of $Z_n(\pi)$ lead to:

\[
a_n^2(\frac{C_n}{\sqrt{n}}(\pi) - Q_{0,n}) = -\frac{1}{2\sqrt{n}}\left\{\mathcal{G}_{\psi,n}(\psi_{0,n}, \pi) - \sqrt{n}E[\epsilon_t(\psi_{0,n}, \pi)d_{\psi,t}(\pi)]\right\}' \mathcal{H}_{\psi,n}^{-1}(\pi)
\]

\[
\times \left\{\mathcal{G}_{\psi,n}(\psi_{0,n}, \pi) - \sqrt{n}E[\epsilon_t(\psi_{0,n}, \pi)d_{\psi,t}(\pi)]\right\}.
\]

By (C.17) in the proof of Theorem 4.1:

\[
\sqrt{n}E[\epsilon_t(\psi_{0,n}, \pi)d_{\psi,t}(\pi)] = \sqrt{n}E[\{\epsilon_t(\psi_{0,n}, \pi) - \epsilon_t(\theta_n)\} d_{\psi,t}(\pi)] = E[\sqrt{n}\beta_n g(x_t, \pi_0)d_{\psi,t}(\pi)],
\]

hence $||\beta_n||^{-1}E[\epsilon_t(\psi_{0,n}, \pi)d_{\psi,t}(\pi)] = E[||\beta_n||^{-1}\beta_n g(x_t, \pi_0)d_{\psi,t}(\pi)]$, and therefore

\[
\sup_{\pi \in \Pi} \left|\frac{1}{||\beta_n||}E[\epsilon_t(\psi_{0,n}, \pi)d_{\psi,t}(\pi)] + D_\psi(\pi)\omega_0\right| \to 0.
\]

By supposition $\sqrt{n}||\beta_n|| \to \infty$, hence Lemma B.1 with the continuous mapping theorem, and Cramér’s Theorem, yield:

\[
\sup_{\pi \in \Pi} \left\|\frac{1}{\sqrt{n}}\mathcal{G}_{\psi,n}(\psi_{0,n}, \pi)\right\| \leq \frac{1}{\inf_{\pi \in \Pi} \sqrt{n}||\beta_n||} \sup_{\pi \in \Pi} ||\mathcal{G}_{\psi,n}(\psi_{0,n}, \pi)|| \to 0.
\]

Lemma B.2 applied to $\mathcal{H}_{\psi,n}(\pi)$, and the Slutsky theorem complete the proof. QED
Write $\epsilon_t(\psi, \pi) = y_t - \zeta' x_t - \beta' g(x_t, \pi)$ and define
\[
\begin{align*}
K_{\psi,t}(\pi, \lambda) &\equiv F (\lambda' \mathbf{W}(x_t)) - b_\psi(\pi, \lambda)' \mathbf{H}_\psi^{-1}(\pi) d_{\psi,t}(\pi) \\
K_{\theta,t}(\lambda) &\equiv F (\lambda' \mathbf{W}(x_t)) - b_\theta(\lambda)' \mathbf{H}_\theta^{-1} d_{\theta,t}(\beta_n/ \| \beta_n \|, \pi_0).
\end{align*}
\]
Recall $\psi_n$ is the (possibly drifting) true value of $\psi = [\beta', \zeta']'$ under $H_0$.

**Lemma B.9.** Let Assumption 1 hold.

a. Under $\mathcal{C}(i, b)$ with $\| b \| < \infty$:
\[
\left\{ \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \{ \epsilon_t(\psi_n, \pi)K_{\psi,t}(\pi, \lambda) - E[\epsilon_t(\psi_n, \pi)K_{\psi,t}(\pi, \lambda)] \} : \Pi, \Lambda \right\} \Rightarrow^* \{ Z_{\psi}(\pi, \lambda) : \Pi, \Lambda \},
\]
a zero mean Gaussian process with covariance kernel $E[Z_{\psi}(\pi, \lambda)Z_{\psi}(\tilde{\pi}, \tilde{\lambda})]$. Under $H_0$,
\[
\sup_{\pi \in \Pi, \lambda \in \Lambda} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \{ \epsilon_t(\psi_n, \pi)K_{\psi,t}(\pi, \lambda) - E[\epsilon_t(\psi_n, \pi)K_{\psi,t}(\pi, \lambda)] \} - \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \epsilon_t K_{\psi,t}(\pi, \lambda) \right| \overset{p}{\to} 0, \quad (B.11)
\]
and $E[Z_{\psi}(\pi, \lambda)Z_{\psi}(\tilde{\pi}, \tilde{\lambda})] = \sigma^2 E[K_{\psi,t}(\pi, \lambda)K_{\psi,t}(\tilde{\pi}, \tilde{\lambda})]$. 

b. Under $\mathcal{C}(i, \omega_0)$, $\{1/\sqrt{n} \sum_{t=1}^{n} \epsilon_t K_{\theta,t}(\lambda) : \lambda \in \Lambda \} \Rightarrow^* \{ Z_\theta : \lambda \in \Lambda \}$, a zero mean Gaussian process with covariance $E[Z_\theta(\lambda)Z_\theta(\tilde{\lambda})] = E[\epsilon_t^2 K_{\theta,t}(\lambda)K_{\theta,t}(\tilde{\lambda})]$ where $K_{\theta,t}(\lambda) \equiv F (\lambda' \mathbf{W}(x_t)) - b_\theta(\lambda)' \mathbf{H}_\theta^{-1} d_{\theta,t}$.

**Proof.** We only prove Claim (a). The proof for Claim (b) is nearly identical.

$\Pi, \Lambda$ are compact and therefore bounded. Weak convergences to a Gaussian process with *almost surely* uniformly continuous and bounded sample paths requires convergence in finite dimensional distributions, and stochastic equicontinuity (see, e.g., Dudley, 1978; Pollard, 1990).

Write compactly $\chi \equiv [\pi', \lambda']' \in \mathcal{X} \equiv \Pi \times \Lambda$, and define:
\[
\begin{align*}
\mathcal{E}_{\psi,t}(\psi_n, \chi) &\equiv \epsilon_t(\psi_n, \pi)K_{\psi,t}(\pi, \lambda) - E[\epsilon_t(\psi_n, \pi)K_{\psi,t}(\pi, \lambda)] \\
\mathcal{E}_{\psi,t}(\psi_n, \chi; a, m) &\equiv \sum_{i=1}^{m} \alpha_i \mathcal{E}_{\psi,t}(\psi_n, \chi_i)
\end{align*}
\]
where $m \in \mathbb{N}$, $a \in \mathbb{R}^m$ satisfies $a'a = 1$, and $\{\chi_1, ..., \chi_m\}$ is an $m$-tuple of points $\chi_i = [\pi_i', \lambda_i]' \in \mathcal{X}$.

Under Assumption 1.b,c $\mathcal{E}_{\psi,t}(\psi_n, \chi; a, m)$ has a zero mean, and is strictly stationary, $L_p$-bounded, $p > 4$, and $\beta$-mixing with coefficients $\beta_t = O(l^{-1(q/(q-p)-t)})$ for some $t > 0$ and $q > p$. Hence $E[(1/\sqrt{n} \sum_{t=1}^{n} \mathcal{E}_{\psi,t}(\psi_n, \chi; a, m))^2] = O(1)$ (McLeish, 1975, Theorem 1.6, Lemma 2.1). Long run variance Assumption 1.d(v) coupled with Assumption 4 imply $E[(\sum_{t=1}^{n} \mathcal{E}_{\psi,t}(\psi_n, \chi; a, m))^2] \to \infty$. 

15
Now invoke Theorem 1.4 in Ibragimov (1962) to yield:

\[ \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \mathcal{E}_{\psi,t}(\psi_{n}, \chi; a, m) \xrightarrow{d} N \left( 0, \lim_{n \to \infty} E \left( \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \mathcal{E}_{\psi,t}(\psi_{n}, \chi; a, m) \right)^{2} \right), \]

where \( \lim_{n \to \infty} E[\{1/\sqrt{n} \sum_{t=1}^{n} \mathcal{E}_{\psi,t}(\psi_{n}, \chi; a, m)\}^{2}] < \infty \). Convergence in finite dimensional distributions now follows by the Cramér-Wold theorem.

Next, after adding and subtracting \( \beta_{n} g(x_{t}, \pi_{0}) \):

\[ \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \mathcal{E}_{\psi,t}(\psi_{n}, \chi) \]

\[ = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \{ \epsilon_{t} K_{\psi,t}(\pi, \lambda) - E[\epsilon_{t} K_{\psi,t}(\pi, \lambda)] \} \]

\[ - \sqrt{n} \beta_{n}^{1/2} \sum_{t=1}^{n} \{ x_{t} \{ g(x_{t}, \pi) - g(x_{t}, \pi_{0}) \} K_{\psi,t}(\pi, \lambda) - E[ x_{t} \{ g(x_{t}, \pi) - g(x_{t}, \pi_{0}) \} K_{\psi,t}(\pi, \lambda) ] \} \]

\[ = 3_{n}(\pi, \lambda) + \mathfrak{X}_{n}(\pi, \lambda). \]

Under \( H_{0} \) and Assumption 1.a, \( E[\epsilon_{t} K_{\psi,t}(\pi, \lambda)] = 0 \) and

\[ E \left[ 3_{n}(\pi, \lambda)3_{n}(\bar{\pi}, \bar{\lambda}) \right] = E \left[ \epsilon_{t}^{2} K_{\psi,t}(\pi, \lambda) K_{\psi,t}(\bar{\pi}, \bar{\lambda}) \right] = \sigma^{2} E \left[ K_{\psi,t}(\pi, \lambda) K_{\psi,t}(\bar{\pi}, \bar{\lambda}) \right]. \]

Further, \( \sup_{\pi \in \Pi, \lambda \in \Lambda} |\mathfrak{X}_{n}(\pi, \lambda)| \xrightarrow{p} 0 \) by Lemma B.13. This proves (B.11).

Stochastic equicontinuity for \( \mathcal{E}_{\psi,t}(\psi_{n}, \chi) \) holds if \( \forall (\epsilon, \eta) > 0 \) there exists \( \delta > 0 \) such that:

\[ \lim_{n \to \infty} \mathcal{P}_{n}(r, \delta, \eta) = \lim_{n \to \infty} P \left( \sup_{\chi, \bar{\chi} \in \mathcal{X}, ||\chi - \bar{\chi}|| \leq \delta} |\mathcal{E}_{\psi,t}(\psi_{n}, \chi) - \mathcal{E}_{\psi,t}(\psi_{n}, \bar{\chi})| > \eta \right) < \epsilon. \]  

(B.12)

We again adapt arguments in Arcones and Yu (1994, proof of Theorem 2.1 and Lemma 2.1) in order to verify (B.12). \( \mathcal{E}_{\psi,t}(\psi_{n}, \chi) \) lies in the \( V-C \) subgraph class of functions \( \mathcal{V}(\mathcal{C}) \) because it is continuous, hence the covering numbers satisfy \( \mathcal{N}(\epsilon, \mathcal{K}, ||\cdot||_{2}) < a \epsilon^{-b} \) for all \( \epsilon \in (0, 1) \) and some \( a, b > 0 \) (e.g. Lemma 7.13 in Dudley, 1978, and Lemma II.25 in Pollard, 1984). Furthermore, under Assumption 1.b,c and by multiple uses of Minkowski and Hölder’s inequalities, it is easily verified that \( \mathcal{E}_{\psi,t}(\psi_{n}, \chi) \) is \( L_{r} \)-bounded, \( r \equiv p/2 > 2 \), and \( \beta \)-mixing with coefficients \( \beta_{l} = O(l^{-q(p/(q-p)-l)}), \)

\( q > p > 6 \) and tiny \( l > 0 \). By simple algebra it follows \( \beta_{l} = O(l^{-r/(r-2)}) = O(l^{-p/(p-4)}) \) because \( p/(p-4) < qp/(q-p) \). Therefore \( \{ \mathcal{E}_{\psi,t}(\psi_{n}, \chi) : \chi \in \mathcal{X} \} \) is stochastically equicontinuous Arcones and Yu (1994, Lemma 2.1, see especially eq. (2.13)). \( \square \)
Recall:

\[ d_{\psi,t}(\pi) \equiv [g(x_t, \pi)', x_t']' \quad \text{and} \quad d_{\theta,t}(\omega, \pi) \equiv \left[ g(x_t, \pi)', x_t', \omega' \frac{\partial}{\partial \pi} g(x_t, \pi) \right]' \]

\[ \hat{b}_{\psi,n}(\pi, \lambda) \equiv \frac{1}{n} \sum_{t=1}^{n} F(\lambda'W(x_t)) d_{\psi,t}(\pi) \quad \text{and} \quad b_{\psi}(\pi, \lambda) \equiv E \left[ F(\lambda'W(x_t)) \right] d_{\psi,t}(\pi) \]

\[ \hat{b}_{\theta,n}(\omega, \pi, \lambda) \equiv \frac{1}{n} \sum_{t=1}^{n} F(\lambda'W(x_t)) d_{\theta,t}(\omega, \pi) \quad \text{and} \quad b_{\theta}(\omega, \pi, \lambda) \equiv E \left[ F(\lambda'W(x_t)) \right] d_{\theta,t}(\omega, \pi) \]

**Lemma B.10.** Under Assumption 1, \( \sup_{\omega \in \mathbb{R}^k : ||\omega|| = 1, \pi \in \Pi, \lambda \in \Lambda} \left| \hat{b}_{\theta,n}(\omega, \pi, \lambda) - b_{\theta}(\omega, \pi, \lambda) \right| \xrightarrow{p} 0 \) and \( \sup_{\pi \in \Pi, \lambda \in \Lambda} \left| \hat{b}_{\psi,n}(\pi, \lambda) - b_{\psi}(\pi, \lambda) \right| \xrightarrow{p} 0 \).

**Proof.** Pointwise \( \hat{b}_{\psi,n}(\pi, \lambda) \xrightarrow{p} b_{\psi}(\pi, \lambda) \) follows from stationarity, ergodicity, and the Assumption 1 moment bounds. Next, we show stochastic equicontinuity: \( \forall (\epsilon, \eta) > 0 \) there exists \( \delta > 0 \) such that:

\[
\lim_{n \to \infty} \mathcal{P}_n(r, \delta, \eta) = \lim_{n \to \infty} \mathbb{P} \left( \sup_{\chi, \tilde{\chi} \in \mathcal{X} : ||\chi - \tilde{\chi}|| \leq \delta} \left| \left\{ \hat{b}_{\psi,n}(\chi) - b_{\psi}(\chi) \right\} - \left\{ \hat{b}_{\psi,n}(\tilde{\chi}) - b_{\psi}(\tilde{\chi}) \right\} \right| > \eta \right) < \epsilon.
\]

where \( \chi = [\lambda', \pi']' \in \mathcal{X} = \Lambda \times \Pi \). There exists \( \chi_* \in \mathcal{X} \), \( ||\chi - \chi_*|| \leq ||\chi - \tilde{\chi}|| \), such that:

\[
\left\{ \hat{b}_{\psi,n}(\chi) - b_{\psi}(\chi) \right\} - \left\{ \hat{b}_{\psi,n}(\tilde{\chi}) - b_{\psi}(\tilde{\chi}) \right\} = \frac{1}{n} \sum_{t=1}^{n} \frac{\partial}{\partial \chi} ((F(\lambda_*'W(x_t)) d_{\psi,t}(\pi_*)) - E[F(\lambda_*'W(x_t)) d_{\psi,t}(\pi_*)])' (\chi - \tilde{\chi}).
\]

The envelop moment bounds in Assumption 1 imply:

\[
E \left[ \sup_{\chi \in \mathcal{X}} \left| \frac{\partial}{\partial \chi} \left\{ (F(\lambda_*'W(x_t)) d_{\psi,t}(\pi_*)) - E[F(\lambda_*'W(x_t)) d_{\psi,t}(\pi_*))] \right\} \right| \right] \leq K < \infty.
\]

Now invoke Markov’s inequality to deduce \( \mathcal{P}_n(r, \delta, \eta) \leq \eta^{-1} K \delta < \epsilon \) for any \( 0 < \delta < \epsilon \eta/K \). QED

Define \( \Theta^+ \equiv \{ \theta^+ \in \mathbb{R}^k + k_{n+1} : \theta^+ = (||\beta||, \omega(\beta), \zeta, \pi) : \beta \in \mathcal{B}, \zeta \in \mathcal{Z}(\beta), \pi \in \Pi \} \) and

\[
\epsilon_t(\theta^+) \equiv y_t - \zeta' x_t - ||\beta|| \omega' g(x_t, \pi) \quad \text{and} \quad \hat{\mathcal{H}}_n(\omega, \pi) = \frac{1}{n} \sum_{t=1}^{n} d_{\theta,t}(\omega, \pi) d_{\theta,t}(\omega, \pi)'
\]

\[
\hat{\epsilon}^2_n(\theta^+, \lambda) = \frac{1}{n} \sum_{t=1}^{n} \epsilon^2_{t}(\theta^+) \left[ F(\lambda'W(x_t)) - \hat{b}_{\theta,n}(\theta, \omega, \lambda)' \hat{\mathcal{H}}_{n-1}^{-1}(\omega, \pi) d_{\theta,t}(\omega, \pi) \right]^2
\]

\[
\theta^2(\theta^+, \lambda) = E \left[ \epsilon^2_{t}(\theta) \left[ F(\lambda'W(x_t)) - b_{\theta}(\theta, \omega, \lambda)' \mathcal{H}_{n-1}^{-1}(\omega, \pi) d_{\theta,t}(\omega, \pi) \right]^2 \right].
\]
Lemma B.11. Under Assumption 1, \( \sup_{\theta^+ \in \Theta^+, \lambda \in \Lambda} \| \hat{v}_n^2(\theta^+, \lambda) - v^2(\theta^+, \lambda) \| \overset{P}{\to} 0. \)

Proof. Define

\[
v_n^2(\theta^+, \lambda) = \frac{1}{n} \sum_{t=1}^{n} \epsilon_t^2(\theta) \left\{ F(\lambda' W(x_t)) - b_\theta(\theta, \omega, \lambda)' H_{\theta}^{-1}(\omega, \pi) d_{\theta, t}(\omega, \pi) \right\}^2
\]

\[
\mathcal{C}_n(\theta^+) = \frac{1}{n} \sum_{t=1}^{n} \epsilon_t^2(\theta^+) d_{\theta, t}(\omega, \pi)
\]

\[
\mathcal{E}_n(\theta^+, \lambda) = \frac{1}{n} \sum_{t=1}^{n} \epsilon_t^2(\theta^+) d_{\theta, t}(\omega, \pi) F(\lambda' W(x_t))
\]

Then:

\[
\hat{v}_n^2(\theta^+, \lambda) - v_n^2(\theta^+, \lambda) = - \left\{ b_{\theta, n}(\theta, \omega, \lambda)' \hat{H}_n^{-1}(\omega, \pi) - b_\theta(\theta, \omega, \lambda)' H_{\theta}^{-1}(\omega, \pi) \right\}
\]

\[
\times \left\{ 2\mathcal{E}_n(\theta^+, \lambda) - \left( b_{\theta, n}(\theta, \omega, \lambda)' \hat{H}_n^{-1}(\omega, \pi) + b_\theta(\theta, \omega, \lambda)' H_{\theta}^{-1}(\omega, \pi) \right) \mathcal{C}_n(\theta^+) \right\}.
\]

By the same arguments used to prove Lemma B.6, \( \mathcal{C}_n(\theta^+) \overset{P}{\to} E[\epsilon_t^2(\theta^+) d_{\theta, t}(\omega, \pi)] \) uniformly on \( \Theta^+ \). Further, \( \mathcal{E}_n(\theta^+, \lambda) \overset{P}{\to} E[\epsilon_t^2(\theta^+) d_{\theta, t}(\omega, \pi) F(\lambda' W(x_t))] \) uniformly on \( \Theta^+ \times \Lambda \) because (i) point-wise converges follows from the assumed moment and mixing properties, and (ii) \( \mathcal{E}_n(\theta^+, \lambda) \) is stochastically equicontinuous by arguments in the proof of Lemma B.10 after simple alterations. Now apply Lemmas B.6 and B.10 to yield \( |\hat{v}_n^2(\theta^+, \lambda) - v_n^2(\theta^+, \lambda)| \overset{P}{\to} 0 \) uniformly on \( \Theta^+ \). Finally, \( v_n^2(\theta^+, \lambda) \overset{P}{\to} v^2(\theta^+, \lambda) \) uniformly on \( \Theta^+ \) by the same arguments in the proof of Lemma B.10. \( QED \)

Recall \( b_\theta(\omega, \pi, \lambda) \equiv E[F(\lambda' W(x_t)) d_{\theta, t}(\omega, \pi)] \), and define

\[
v^2(\lambda) \equiv v^2(\omega_0, \pi_0, \lambda)
\]

where:

\[
v^2(\omega, \pi, \lambda) \equiv E \left[ \epsilon_t^2(\psi_0, \pi) \left\{ F(\lambda' W(x_t)) - b_\theta(\omega, \pi, \lambda)' H_{\theta}^{-1}(\omega, \pi) d_{\theta, t}(\omega, \pi) \right\}^2 \right].
\]

Lemma B.12. Let Assumption 3 hold. Under \( C(i, b) \) with \( \|b\| < \infty \), the following set has Lebesgue measure zero:

\[
\left\{ \lambda \in \Lambda : \inf_{\omega, \pi \in \Pi} v^2(\omega, \pi, \lambda) = 0 \right\}.
\]

Under \( C(ii, \omega_0) \), the set \( \{ \lambda \in \Lambda : v^2(\lambda) = 0 \} \) has Lebesgue measure zero.

Proof. The proof under \( C(ii, \omega_0) \) is identical to Bieren’s (1990, Lemma 2). Consider weak
identification cases \( C(i, b) \) with \( ||b|| < \infty \). Assume

\[
S^* \equiv \left\{ \lambda \in \Lambda : \inf_{\omega' = 1, \pi \in \Pi} \nu^2(\omega, \pi, \lambda) = 0 \right\}
\]

has positive Lebesgue measure, and take any \( \lambda \in S^* \). Use \( P(E[\inf_{\pi \in \Pi} \{ \epsilon_t^2(\psi_0, \pi) \} | x_t] > 0) = 1 \) under Assumption 3 to deduce

\[
F(\lambda'W(x_t)) = b(\omega, \pi, \lambda)'H^{-1}_\theta(\omega, \pi)d_{\theta,t}(\omega, \pi) \text{ a.s.}
\]

Now use the Assumption 3.b Borel function \( \mu \) to yield that

\[
E[\mu(x_t)F(\lambda'W(x_t))] = E[\mu(x_t)d_{\theta,t}(\omega, \pi)'H^{-1}_\theta(\omega, \pi)b(\omega, \pi, \lambda)].
\]

Note \( b(\omega, \pi, \lambda) \equiv E[d_{\theta,t}(\omega, \pi)F(\lambda'W(x_t))] \) hence

\[
E[\mu(x_t)F(\lambda'W(x_t))] = E[\xi(\omega, \pi)'d_{\theta,t}(\omega, \pi) \times F(\lambda'W(x_t))],
\]

where \( \xi(\omega, \pi) \equiv H^{-1}_\theta(\omega, \pi)E[\mu(x_t)d_{\theta,t}(\omega, \pi)] \). This implies

\[
E[\{\mu(x_t) - \xi(\omega, \pi)'d_{\theta,t}(\omega, \pi)\} F(\lambda'W(x_t))] = 0. \tag{B.13}
\]

Since \( S^* \) has positive Lebesgue measure, equality in (B.13) applies for all \( \lambda \) in a subset with positive Lebesgue measure. Thus \( \mu(x_t) = \xi(\omega, \pi)'d_{\theta,t}(\omega, \pi) \) a.s. by Theorem 2.3 in Stinchcombe and White (1998). Hence \( E[\kappa_t(\omega, \pi)\kappa_t(\omega, \pi)'] \) is singular, where \( \kappa_t(\omega, \pi) \equiv [\mu(x_t), d_{\theta,t}(\omega, \pi)]' \), which contradicts Assumption 3.b(ii). \( \Box \)

Define

\[
\mathcal{M}_t(\pi, \lambda) \equiv \{g(x_t, \pi_0) - g(x_t, \pi)\} F(\lambda'W(x_t)) \text{ and } \tilde{\mathcal{M}}_t(\pi) \equiv \{g(x_t, \pi) - g(x_t, \pi_0)\} d_{\psi,t}(\pi)'.
\]

**Lemma B.13.** Under Assumption 1:

\[
\sup_{\pi \in \Pi, \lambda \in \Lambda} \left| \frac{1}{n} \sum_{t=1}^{n} \epsilon_t F(\lambda'W(x_t)) - E[\epsilon_t F(\lambda'W(x_t))] \right| \overset{P}{\to} 0,
\]

\[
\sup_{\pi \in \Pi, \lambda \in \Lambda} \left| \frac{1}{n} \sum_{t=1}^{n} \mathcal{M}_t(\pi, \lambda) - E[\mathcal{M}_t(\pi, \lambda)] \right| \overset{P}{\to} 0 \text{ where } \sup_{\pi \in \Pi, \lambda \in \Lambda} |E[\mathcal{M}_t(\pi, \lambda)]| < \infty
\]

\[
\sup_{\pi \in \Pi} \left| \frac{1}{n} \sum_{t=1}^{n} \tilde{\mathcal{M}}_t(\pi) - E[\tilde{\mathcal{M}}_t(\pi)] \right| \overset{P}{\to} 0 \text{ where } \sup_{\pi \in \Pi} |E[\tilde{\mathcal{M}}_t(\pi)]| < \infty
\]
Proof. In view of envelope moment bounds in 1.c, the argument is essentially identical to
the proof of Lemma B.10. QED.

C Proof of Theorem 4.1

Theorem 4.1. Let Assumptions 1 and 2 hold.

a. Under drift case $C(i, b)$ with $||b|| < \infty$, $(\sqrt{n}(\hat{\psi}_n(\pi) - \psi_n), \hat{\pi}_n) \overset{d}{\to} (\tau(\pi^*(b), b), \pi^*(b))$.

b. Under drift case $C(\omega_0)$, $\sqrt{n} \mathcal{B}(\hat{\beta}_n)(\hat{\theta}_n - \theta_n) \overset{d}{\to} -\mathcal{H}_\theta^{-1} \mathcal{G}_\theta$.

Proof.

Claim a.

Step 1: We first prove

$$\left\{ \sqrt{n}(\hat{\psi}_n(\pi) - \psi_n) : \Pi \right\} \Rightarrow \left\{ \tau(\pi, b) : \Pi \right\}.$$  (C.14)

Recall $\psi_{0,n} = [0'_{k_\beta}, \zeta_0']$. By Lemma B.7.a:

$$\sqrt{n}(\hat{\psi}_n(\pi) - \psi_n) = \sqrt{n}(\hat{\psi}_n(\pi) - \psi_{0,n}) + \sqrt{n}(\psi_{0,n} - \psi_n)$$

$$= -\hat{\mathcal{H}}^{-1}_{\psi,n}(\pi)\sqrt{n}\frac{\partial}{\partial \psi}Q_n(\psi_{0,n}, \pi) - \left[ \sqrt{n}\beta_n', 0'_{k_\beta} \right]'.$$  (C.15)

By the construction of $\mathcal{G}_{\psi,n}(\theta)$ in (A.3), we can write:

$$\sqrt{n}\frac{\partial}{\partial \psi}Q_n(\psi_{0,n}, \pi) = \mathcal{G}_{\psi,n}(\psi_{0,n}, \pi) - \sqrt{n}E[\epsilon_t(\psi_{0,n}, \pi)d_{\psi,t}(\pi)].$$  (C.16)

Assumption 1.a implies $E[\epsilon_t(\theta_n)d_{\psi,t}(\pi)] = 0$, hence:

$$\sqrt{n}E[\epsilon_t(\psi_{0,n}, \pi)d_{\psi,t}(\pi)] = \sqrt{n}E[\{\epsilon_t(\psi_{0,n}, \pi) - \epsilon_t(\theta_n)\} d_{\psi,t}(\pi)] = E[\sqrt{n}\beta_n'g(x_t, \pi_0)d_{\psi,t}(\pi)].$$  (C.17)

Therefore, by the definition of $D_{\psi}(\pi)$ in (A.4), and $\sqrt{n}\beta_n \to b$ with $||b|| < \infty$:

$$\sup_{\pi \in \Pi} \left| \sqrt{n}E[\epsilon_t(\psi_{0,n}, \pi)d_{\psi,t}(\pi)] + D_{\psi}(\pi)b \right| \to 0.$$  (C.18)

By Lemma B.2 $\sup_{\pi \in \Pi} ||\hat{\mathcal{H}}_{\psi,n}(\pi) - \mathcal{H}_{\psi}(\pi)|| \overset{p}{\to} 0$, where $\mathcal{H}_{\psi}(\pi)$ is bounded and positive definite uniformly on $\Pi$. Now combine (C.15)-(C.18) to yield:

$$\sup_{\pi \in \Pi} \left\| \sqrt{n}(\hat{\psi}_n(\pi) - \psi_n) - \left(-\mathcal{H}_{\psi}^{-1}(\pi) \{\mathcal{G}_{\psi,n}(\psi_{0,n}, \pi) + D_{\psi}(\pi)b\} - \left[ b, 0'_{k_\beta} \right]' \right) \right\| \overset{p}{\to} 0.$$  (C.19)
Therefore (C.14) follows by application of Lemma B.1.

**Step 2:** Now turn to \( \hat{\pi}_n \). Write \( Q_n^c(\pi) \equiv Q_n(\hat{\psi}_n(\pi), \pi) \). Let drift case \( C(i,b) \) hold with \( \|b\| < \infty \). By Lemma B.8.a \( \{n(Q_n^c(\pi), \pi) - Q_{0,n} : \Pi\} \Rightarrow^* \{\xi_\psi(\pi, b) : \Pi\} \), hence by the mapping theorem \( \{\arg\min_{\pi \in \Pi} \{n(Q_n^c(\pi) - Q_{0,n})\} - \arg\min_{\pi \in \Pi} \{\xi_\psi(\pi, b)\}\} \xrightarrow{p} 0 \). Therefore \( \hat{\pi}_n \xrightarrow{d} \pi^*(b) = \arg\min_{\pi \in \Pi} \{\xi_\psi(\pi, b)\} \) by the mapping theorem and Assumption 2.

**Step 3:** The proof is complete by showing joint weak convergence for \( \sqrt{n}(\hat{\psi}_n(\pi) - \psi_n) \) and \( \hat{\pi}_n \).

First, \( \sqrt{n}(\hat{\psi}_n(\pi) - \psi_n) \) and \( \hat{\pi}_n \) are continuous functions of \( G_{\psi,n}(\psi_{0,n}, \pi) \) and \( \hat{H}_{\psi,n}(\pi) \). The former follows from (C.15) and (C.16). In order to understand \( \hat{\pi}_n \), define

\[
\xi_{\psi,n}(\pi, b) \equiv -\frac{1}{2} \{G_{\psi,n}(\psi_{0,n}, \pi) + D_\psi(\pi)b\}' \hat{H}_{\psi,n}^{-1}(\pi) \{G_{\psi,n}(\psi_{0,n}, \pi) + D_\psi(\pi)b\}.
\]

By Lemmas B.1 and B.2 \( \{\xi_{\psi,n}(\pi, b) : \Pi\} \Rightarrow^* \{\xi_\psi(\pi, b) : \Pi\} \). Hence, by Lemma B.8.a and the mapping theorem

\[
\left| \arg\min_{\pi \in \Pi} \{n(Q_n^c(\pi) - Q_{0,n})\} - \arg\min_{\pi \in \Pi} \{\xi_{\psi,n}(\pi, b)\}\right| \xrightarrow{p} 0.
\]

In view of the argument above, this implies

\[
\left| \hat{\pi}_n - \arg\min_{\pi \in \Pi} \{\xi_{\psi,n}(\pi, b)\}\right| \xrightarrow{p} 0.
\]

Hence \( \hat{\pi}_n \) can be expressed as a continuous function of \( G_{\psi,n}(\psi_{0,n}, \pi) \) and \( \hat{H}_{\psi,n}(\pi) \).

Second, \( G_{\psi,n}(\psi_{0,n}, \pi) \) and \( \hat{H}_{\psi,n}(\pi) \) converge jointly because the latter has a non-random limit uniformly on \( \Pi \) (cf. Andrews and Cheng, 2012b, p. 25). Hence

\[
\left\{ \sqrt{n}(\hat{\psi}_n(\pi) - \psi_n), \hat{\pi}_n : \Pi\right\} \Rightarrow^* \{\tau(\pi, b), \pi^*(b) : \Pi\}.
\]

By the mapping theorem it therefore follows

\[
\left( \sqrt{n}(\hat{\psi}_n(\pi) - \psi_n), \hat{\pi}_n\right) \Rightarrow^* (\tau(\pi^*(b), b), \pi^*(b)).
\]

Finally, a subsequent proof requires uniform consistency

\[
\sup_{\pi \in \Pi} \left\| \hat{\psi}_n(\pi) - \psi_n \right\| \xrightarrow{p} 0. \tag{C.20}
\]
Note that
\[ \hat{\psi}_n(\pi) - \psi_n = -\hat{H}^{-1}_{\psi,n}(\pi) \frac{\partial}{\partial \psi} Q_n(\psi_{0,n}, \pi) - \left[ \beta'_{n}, 0'_{k\beta} \right], \]
where \( \sup_{\pi \in \Pi} \| \hat{H}^{-1}_{\psi,n}(\pi) - H^{-1}_{\psi}(\pi) \| \overset{P}{\to} 0 \) and \( \beta_n \to 0 \). Moreover, by the Assumption 1.b,c,d(iii) moment and envelope bounds and \( \beta_n \to 0 \):

\[
\sup_{\pi \in \Pi} \left\| \frac{\partial}{\partial \psi} Q_n(\psi_{0,n}, \pi) \right\| \leq \sup_{\pi \in \Pi} \left\| \frac{\partial}{\partial \psi} Q_n(\psi_{0,n}, \pi) - E [\epsilon_t(\psi_{0,n}, \pi) d_{\psi,t}(\pi)] \right\| + \sup_{\pi \in \Pi} \| E [\beta'_n g(x_t, \pi_0) d_{\psi,t}(\pi)] \|
= \sup_{\pi \in \Pi} \left\| \frac{1}{n} \sum_{t=1}^{n} \epsilon_t(\psi_{0,n}, \pi) d_{\psi,t}(\pi) - E [\epsilon_t(\psi_{0,n}, \pi) d_{\psi,t}(\pi)] \right\| + o_p(1)
\equiv \epsilon_n + o_p(1).\]

Finally, \( \epsilon_n \overset{P}{\to} 0 \) by the same arguments used to prove Lemmas B.2 and B.6. Therefore:

\[
\sup_{\pi \in \Pi} \left\| \hat{\psi}_n(\pi) - \psi_n \right\| = \sup_{\pi \in \Pi} \left\| -\hat{H}^{-1}_{\psi,n}(\pi) \frac{\partial}{\partial \psi} Q_n(\psi_{0,n}, \pi) - \left[ \beta'_{n}, 0'_{k\beta} \right] \right\|
\leq \sup_{\pi \in \Pi} \left\| -\hat{H}^{-1}_{\psi,n}(\pi) \epsilon_n - \left[ \beta'_{n}, 0'_{k\beta} \right] \right\| + o_p(1) \overset{P}{\to} 0.
\]

This proves (C.20).

**Claim b.** Let drift case \( C(ii, \omega_0) \) hold, and define

\[ \hat{H}_n(\omega, \pi) \equiv \frac{1}{n} \sum_{t=1}^{n} d_{\theta,t}(\omega, \pi) d_{\theta,t}(\omega, \pi)' \text{ and } H_\theta(\omega, \pi) \equiv E [d_{\theta,t}(\omega, \pi) d_{\theta,t}(\omega, \pi)']. \]

Recall \( B(\beta) \) defined in (A.2) and \( \omega(\beta) \) defined in (A.5). By the first order condition \( (\partial/\partial \theta) Q_n(\hat{\theta}_n) = 0 \) and the mean value theorem there exists \( \theta^*_n, ||\theta^*_n - \theta_n|| \leq ||\hat{\theta}_n - \theta_n|| \), such that:

\[
0 = B(\beta_n)^{-1} \sqrt{n} \frac{\partial}{\partial \theta} Q_n(\theta_n) + B(\beta_n)^{-1} \frac{\partial^2}{\partial \theta \partial \theta'} Q_n(\theta^*_n) B(\beta_n)^{-1} \times \sqrt{n} B(\beta_n) \left( \hat{\theta}_n - \theta_n \right)
= B(\beta_n)^{-1} \sqrt{n} \frac{\partial}{\partial \theta} Q_n(\theta_n) + \hat{H}_n(\omega(\beta^*_n), \pi^*_n) \sqrt{n} B(\beta_n) \left( \hat{\theta}_n - \theta_n \right).
\]

The second equality follows from the constructions of \( B(\beta), (\partial^2/\partial \theta \partial \theta') Q_n(\theta) \) and \( \hat{H}_n(\theta) \). Hence:

\[
\sqrt{n} B(\beta_n) \left( \hat{\theta}_n - \theta_n \right) = \hat{H}^{-1}_{\theta,n}(\omega(\beta^*_n), \pi^*_n) B(\beta_n)^{-1} \sqrt{n} \frac{\partial}{\partial \theta} Q_n(\theta_n). \tag{C.21}
\]
Observe that \( ||\theta_n^* - \theta_n|| \leq ||\hat{\theta}_n - \theta_n|| \), and by the argument below:

\[
\left\| \hat{\theta}_n - \theta_n \right\| \overset{p}{\to} 0. \tag{C.22}
\]

Hence \( \hat{H}_n(\omega(\beta_n^*), \pi_n^*) \overset{p}{\to} H_\theta \) by Lemma B.6 and continuity. Corollary B.4 now yields the result.

It remains to prove (C.22). Use (C.21) to yield:

\[
\left\| \sqrt{n} \mathfrak{B}(\beta_n) \left( \hat{\theta}_n - \theta_n \right) \right\| \leq \sup_{\omega \in \mathbb{R}^k: ||\omega||=1, \pi \in \Pi} \left\| \hat{H}_n^{-1}(\omega, \pi) - H_\theta^{-1}(\omega, \pi) \right\| \left\| \mathfrak{B}(\beta_n)^{-1} \sqrt{n} \frac{\partial}{\partial \theta} Q_n(\theta_n) \right\| 
\]

\[+ \sup_{\omega \in \mathbb{R}^k: ||\omega||=1, \pi \in \Pi} \left\| H_\theta^{-1}(\omega, \pi) \right\| \left\| \mathfrak{B}(\beta_n)^{-1} \sqrt{n} \frac{\partial}{\partial \theta} Q_n(\theta_n) \right\|. \]

By Lemma B.6 and the Slutsky Theorem

\[
\sup_{\omega \in \mathbb{R}^k: ||\omega||=1, \pi \in \Pi} \left\| \hat{H}_n^{-1}(\omega, \pi) - H_\theta^{-1}(\omega, \pi) \right\| \overset{p}{\to} 0,
\]

where \( \sup_{\omega \in \mathbb{R}^k: ||\omega||=1, \pi \in \Pi} \left\| H_\theta^{-1}(\omega, \pi) \right\| < \infty \) follows from the eigenvalue bounds in Lemma B.6. Moreover, by Lemma B.3 and the mapping theorem

\[\mathfrak{B}(\beta_n)^{-1} \sqrt{n} \frac{\partial}{\partial \theta} Q_n(\theta_n) = O_p(1).\]

This proves \( \sqrt{n} \mathfrak{B}(\beta_n)(\hat{\theta}_n - \theta_n) = O_p(1) \) hence (C.22). \( \text{Q.E.D.} \)

### D Identification Category Selection Type 2 P-Value

Operate under \( H_0 \). Define \( \mathcal{F}_\infty(c) \equiv P(\mathcal{T}(\lambda) \leq c) \) where \( \{\mathcal{T}(\lambda) : \lambda \in \Lambda\} \) is the asymptotic null chi-squared process under strong identification, and let \( \mathcal{F}_{\lambda,h}(c) \equiv P(\mathcal{T}_{\psi}(\lambda, h) \leq c) \) where \( \{\mathcal{T}_{\psi}(\lambda, h) : \lambda \in \Lambda\} \) is the asymptotic null process under weak identification. The case specific asymptotic p-values are

\[p_n^{\infty}(\lambda) \equiv 1 - \mathcal{F}_\infty(\mathcal{T}_n(\lambda)) = \bar{F}_{\infty}(\mathcal{T}_n(\lambda)) \quad \text{and} \quad p_n(\lambda, h) \equiv 1 - \mathcal{F}_{\lambda,h}(\mathcal{T}_n(\lambda)) = \bar{F}_{\lambda,h}(\mathcal{T}_n(\lambda)).\]

The ICS-2 p-value is computed as follows. Recall \( \mathcal{A}_n \) in (A.6). Let \( (\Delta_1, \Delta_2) \in [0, 1) \) and \( \kappa > 0 \) be user chosen numbers. Let \( s \) be a continuous function on \([0, \infty)\), such that \( s(x) \in [0, 1], s(x) \) is non-increasing in \( x \), \( s(0) = 1 \), and \( s(x) \to 0 \) as \( x \to \infty \). Then:

\[p_n^{(ICS-2)}(\lambda) = \begin{cases} 
  p_{n,1}(\lambda; \Delta_1) \text{ if } \mathcal{A}_n \leq \kappa, & p_{n,2}(\lambda, \Delta_1, \Delta_2) \text{ if } \mathcal{A}_n > \kappa
\end{cases} \]
where
\[ p_{n,1}(\lambda; \Delta_1) \equiv \max \left\{ \sup_{h \in \Theta} \left\{ p_n(\lambda, h) \right\}, p_n^\infty(\lambda) \right\} + \Delta_1 \]  
\[ p_{n,2}(\lambda, \Delta_1, \Delta_2) \equiv p_n^\infty(\lambda) + \Delta_2 + \left\{ p_{n,1}(\lambda; \Delta_1) - p_n^\infty(\lambda) - \Delta_2 \right\} s(A_n - \kappa). \]  

(D.23)

The construction allows for a smooth transition between identification cases, and allows for a non-diverging threshold. The latter necessitates the tuning parameters \((\Delta_1, \Delta_2)\) which promote a correct asymptotic size. See also Andrews and Barwick (2012) for a related method.

See Andrews and Cheng (2012a, p. 2193) for details on determining appropriate choices for \((\Delta_1, \Delta_2, \kappa)\). In theory \(\kappa > 0\) can be any value since the ICS-2 p-value \(p_n^{ICS-2}(\lambda)\) promotes a test with correct asymptotic level. Andrews and Cheng (2012a, p. 2194) and Andrews and Cheng (2013a, p. 50) choose \(\kappa\) for robust t-statistics by minimizing the False Coverage Probability [FCP] for the corresponding robust confidence set.\(^1\) The CM test statistic is not based on a parametric hypothesis, hence the FCP method does not apply. Instead, we may choose ad hoc values like \(\kappa = 1\) or \(\kappa = 1.5\), based on finite sample experiments for various models.\(^2\) Since our focus is an asymptotically valid method for computing \(p_n(\lambda, h)\), and therefore \(\{p_n^{(LE)}(\lambda), p_n^{(ICS-2)}(\lambda), p_n^{(ICS-2)}(\lambda)\}\), we do not present here a theory based alternative to minimizing the FCP in order to select \(\kappa\) for CM tests.

We choose \((\Delta_1, \Delta_2)\) to ensure the asymptotic Null Rejection Probability [NRP] under weak identification \(\sqrt{n||\beta_n||} \to [0, \infty)\) is not larger than \(\alpha\) (Andrews and Cheng, 2012a, Section 5.3). The NRP is

\[ NRP_n(\Delta_1, \Delta_2; \lambda, \kappa) \equiv P\left( p_{n,1}(\lambda; \Delta_1) \leq \alpha \cap A_n \leq \kappa \right) + P\left( p_{n,2}(\lambda; \Delta_1, \Delta_2) \leq \alpha \cap A_n > \kappa \right). \]

Note that \(A_n \overset{d}{\to} A(b)\) under weak identification, where \(A(b)\) is defined in Theorem 5.1.a. Under strong identification and regularity conditions, \(A_n \overset{p}{\to} \infty\) (Theorem 5.1.b).

Define
\[ p_1(\lambda, \tilde{h}; \Delta_1) \equiv \max \left\{ \sup_{h \in \Theta} \left\{ \mathcal{F}_{\lambda,h}(\mathcal{T}_\psi(\lambda, \tilde{h})) \right\}, \mathcal{F}_\infty(\mathcal{T}_\psi(\lambda, \tilde{h})) \right\} + \Delta_1 \]  
\[ p_2(\lambda, \tilde{h}; \Delta_1, \Delta_2) \equiv \mathcal{F}_\infty(\mathcal{T}_\psi(\lambda, \tilde{h})) + \Delta_2 + \left\{ p_1(\lambda) - \mathcal{F}_\infty(\mathcal{T}_\psi(\lambda, \tilde{h})) - \Delta_2 \right\} s(A(b) - \kappa). \]  

(D.24)

The operator \(\sup_{h \in \Theta}\) operates on the distribution function \(\mathcal{F}_{\lambda,h}\) and not its argument \(\mathcal{T}_\psi(\lambda, \tilde{h})\).

\(^1\)Consider the parametric hypothesis \(R(\theta) = 0\). The FCP of a confidence set for \(R(\theta)\) is the probability that the confidence set contains a value different from the true \(R(\theta_n)\), where \(\theta_n \equiv [\beta_n', \phi_n', \pi_n']\).

\(^2\)Andrews and Cheng (2012a,b, 2013a,b) find that a wide range of values for \(\kappa\) lead to similar results for robust Smooth Transition Autoregression model based t-tests, including \(\kappa = 1\) and \(\kappa = 1.5\), because \(\Delta_1\) and \(\Delta_2\) are computed to ensure correct asymptotic size for any chosen \(\kappa\).
This follows from the definition \( p_{n,1}(\lambda; \Delta_1) \equiv \max \{ \sup_{h \in \mathcal{H}} \{ p_n(\lambda, h) \}, p_n^\infty(\lambda) \} + \Delta_1 \), and under weak identification:

\[
\sup_{h \in \mathcal{H}} \{ p_n(\lambda, h) : \lambda \in \Lambda \} = \sup_{h \in \mathcal{H}} \{ \tilde{F}_{\lambda,h}(\mathcal{T}_n(\lambda)) : \lambda \in \Lambda \} \Rightarrow \sup_{h \in \mathcal{H}} \{ \tilde{F}_{\lambda,h}(\mathcal{T}_\psi(\lambda, \tilde{h})) : \lambda \in \Lambda \}.
\]

By Theorem 6.1 and the mapping theorem:

\[
\{ p_{n,1}(\lambda; \Delta_1) : \lambda \in \Lambda \} \Rightarrow^* \{ p_1(\lambda, \tilde{h}; \Delta_1) : \lambda \in \Lambda \}
\]

and

\[
\{ p_{n,2}(\lambda; \Delta_1, \Delta_2) : \lambda \in \Lambda \} \Rightarrow^* \{ p_2(\lambda, \tilde{h}; \Delta_1, \Delta_2) : \lambda \in \Lambda \}.
\]

Joint convergence for \( (p_{n,1}(\lambda; \Delta_1), p_{n,2}(\lambda; \Delta_1, \Delta_2), A_n) \) is straightforward to prove: see the proof of Theorem 6.2. The asymptotic NRP under weak identification is therefore:

\[
NRP(\Delta_1, \Delta_2; \lambda, \tilde{h}) \equiv P \left( \frac{p_1(\lambda, \tilde{h}; \Delta_1) < \alpha \cap A(b) \leq \kappa}{P \left( \frac{p_2(\lambda, \tilde{h}; \Delta_1, \Delta_2) < \alpha \cap A(b) > \kappa}{D.25} \right. \right).
\]

The role \((\Delta_1, \Delta_2)\) play are the same as in Andrews and Cheng (2012a, p. 2193). Let \( \tilde{b}_{sup} \) be such that

\[
\tilde{h}_{sup} \equiv [\tilde{b}_{sup}, \tilde{\gamma}_{sup}] = \arg \max_{h \in \mathcal{H}} \sup_{\tilde{h} \in \mathcal{H}} \tilde{F}_{\lambda,h}(\mathcal{T}_\psi(\lambda, \tilde{h}))
\]

and \( C \geq 0 \) is some constant, e.g. \( C = 1 \). Define

\[
\mathcal{H}_1 \equiv \{ h = [b, \gamma] : h \in \mathcal{H}, \|b\| \leq \|\tilde{b}_{sup}\| + C \}
\]

and

\[
\Delta_1 \equiv \sup_{h \in \mathcal{H}_1} \Delta_1(\tilde{h}) \text{ where } \begin{cases} \Delta_1(\tilde{h}) \geq 0 \text{ solves } NRP(\Delta_1(\tilde{h}), 0; \tilde{h}) = \alpha \\ \Delta_1(\tilde{h}) = 0 \text{ if } NRP(0, 0; \tilde{h}) < \alpha \end{cases}
\]

\[
\Delta_2 \equiv \sup_{h \in \mathcal{H}_1} \Delta_2(\tilde{h}) \text{ where } \begin{cases} \Delta_2(\tilde{h}) \geq 0 \text{ solves } NRP(\Delta_1, \Delta_2(\tilde{h}); \tilde{h}) = \alpha \\ \Delta_1(\tilde{h}) = 0 \text{ if } NRP(\Delta_1, 0; \tilde{h}) < \alpha \end{cases}
\]

If \( NRP(\Delta_1, 0; \tilde{h}) = \alpha \) does not hold for any \( \Delta_1 \), then choose any \( \Delta_1 \) that satisfies \( NRP(\Delta_1, 0; \tilde{h}) \leq \alpha \). The following lemma shows the latter is always feasible (see the proof for examples). Thus, \( NRP(\Delta_1, 0; \tilde{h}) = \alpha \) for some \( \Delta_1 \) holds when \( NRP(\Delta_1, 0; \tilde{h}) \) is strictly decreasing and continuous in \( \Delta_1 \), which generally holds in view of the construction of \( \mathcal{T}_\psi(\lambda, \tilde{h}) \). Similar derivations apply.
to \( \Delta_2 \).

**Lemma D.1.** Let \( \sqrt{n}\|\beta_n\| \to [0, \infty) \), and assume \( F_{\lambda, h}(c) \) is continuous a.e. on \([0, \infty) \). There always exists a (possibly non-unique) \( \Delta_1 \) such that \( \sup_{\bar{h} \in \mathcal{H}} \text{NRP}(\Delta_1, 0; \bar{h}) \leq \alpha \).

Define
\[
\text{AsySz}(\lambda) = \limsup_{n \to \infty} \sup_{\gamma \in \Gamma^*} P_n \left( p_n^{(1)}(\lambda) < \alpha | H_0 \right). 
\]

**Theorem D.2.** Let Assumptions 1-2, 4 and 5 hold. The ICS-2 \( p_n^{(ICS-2)}(\lambda) \) satisfies \( \text{AsySz}(\lambda) \leq \alpha \).

**Proof of Lemma D.1.** By (D.25), the asymptotic Null Rejection Probability under \( \sqrt{n}\|\beta_n\| \to [0, \infty) \) is
\[
\text{NRP}(\Delta_1, \Delta_2; \bar{h}) = P \left( p_1(\lambda, \bar{h}; \Delta_1) < \alpha \cap A(b) \leq \kappa \right) + P \left( p_2(\lambda, \bar{h}; \Delta_1, \Delta_2) < \alpha \cap A(b) > \kappa \right). 
\]
Define \( p^{(LF)}(\lambda, \bar{h}) \equiv \max \{ \sup_{h \in \mathcal{H}} \{ F_{\lambda, h}(T_{\psi}(\lambda, \bar{h})) \}, \bar{F}_n(T_{\psi}(\lambda, \bar{h})) \} \). Note that
\[
P \left( p_1(\lambda, \bar{h}; \Delta_1) < \alpha \cap A(b) \leq \kappa \right) \leq P \left( p^{(LF)}(\lambda, \bar{h}) < \alpha \cap A(b) \leq \kappa \right) \leq P \left( \sup_{h \in \mathcal{H}} \{ F_{\lambda, h}(T_{\psi}(\lambda, \bar{h})) \} < \alpha \cap A(b) \leq \kappa \right) \] (D.27)

and
\[
P \left( p_2(\lambda, \bar{h}; \Delta_1, \Delta_2) < \alpha \cap A(b) > \kappa \right) 
\] 
\[
= P \left( \bar{F}_n(T_{\psi}(\lambda, \bar{h})) + \Delta_2 + \left\{ p^{(LF)}(\lambda, \bar{h}) + \Delta_1 - \bar{F}_n(T_{\psi}(\lambda, \bar{h})) - \Delta_2 \right\} s (A(b) - \kappa) < \alpha \cap A(b) > \kappa \right) 
\] 
\[
\leq P \left( \bar{F}_n(T_{\psi}(\lambda, \bar{h})) (1 - s(A(b) - \kappa)) + p^{(LF)}(\lambda, \bar{h}) s(A(b) - \kappa) + \Delta_1 s(A(b) - \kappa) < \alpha \cap A(b) > \kappa \right). 
\]

Consider two examples:
\[
\Delta_1(\bar{h}) = \left( \max \left\{ \sup_{h \in \mathcal{H}} \{ F_{\lambda, h}(T_{\psi}(\lambda, \bar{h})) \}, \bar{F}_n(T_{\psi}(\lambda, \bar{h})) \right\} - \bar{F}_n(T_{\psi}(\lambda, \bar{h})) \right) \frac{1 - s(A(b) - \kappa)}{s(A(b) - \kappa)} \] (D.28)
\[
\Delta_1(\bar{h}) = \max \left\{ \sup_{h \in \mathcal{H}} \{ F_{\lambda, h}(T_{\psi}(\lambda, \bar{h})) \}, \bar{F}_n(T_{\psi}(\lambda, \bar{h})) \right\} \frac{1 - s(A(b) - \kappa)}{s(A(b) - \kappa)}. \] (D.29)

Use \( \Delta_1(\bar{h}) \) in (D.28) to yield
\[
P \left( p_2(\lambda, \bar{h}; \Delta_1, \Delta_2) < \alpha \cap A(b) > \kappa \right) \leq P \left( p^{(LF)}(\lambda, \bar{h}) < \alpha \cap A(b) > \kappa \right) 
\]
hence
\[
\sup_{\tilde{h} \in \tilde{\mathcal{H}}} \text{NRP}(\Delta_1(\tilde{h}), 0; \tilde{h}) \leq \sup_{\tilde{h} \in \tilde{\mathcal{H}}} P \left( \sup_{h \in \tilde{\mathcal{H}}} \left\{ F_{\lambda,h}(T_\psi(\lambda, \tilde{h})) \right\} < \alpha \cap \mathcal{A}(b) \leq \kappa \right) \\
+ \sup_{\tilde{h} \in \tilde{\mathcal{H}}} P \left( \sup_{h \in \tilde{\mathcal{H}}} \left\{ F_{\lambda,h}(T_\psi(\lambda, \tilde{h})) \right\} < \alpha \cap \mathcal{A}(b) > \kappa \right)
\]
\[
= \sup_{\tilde{h} \in \tilde{\mathcal{H}}} P \left( \sup_{h \in \tilde{\mathcal{H}}} \left\{ F_{\lambda,h}(T_\psi(\lambda, \tilde{h})) \right\} < \alpha \right) \leq \sup_{\tilde{h} \in \tilde{\mathcal{H}}} P \left( F_{\lambda,\tilde{h}}(T_\psi(\lambda, \tilde{h})) < \alpha \right) = \alpha.
\]

The final equality holds because $F_{\lambda,\tilde{h}}$ is continuous by assumption, and $T_\psi(\lambda, \tilde{h})$ is distributed $F_{\lambda,\tilde{h}}$.

Finally, note that
\[
P \left( p_2(\lambda, \tilde{h}; \Delta_1, \Delta_2) < \alpha \cap \mathcal{A}(b) > \kappa \right)
\]
\[
\leq P \left( F_{\infty}(T_\psi(\lambda, \tilde{h})) (1 - s (\mathcal{A}(b) - \kappa)) \right)
\]
\[
+ p^{(LF)}(\lambda, \tilde{h}) s (\mathcal{A}(b) - \kappa) + \Delta_1 s (\mathcal{A}(b) - \kappa) < \alpha \cap \mathcal{A}(b) > \kappa
\]
\[
\leq P \left( p^{(LF)}(\lambda, \tilde{h}) s (\mathcal{A}(b) - \kappa) + \Delta_1 s (\mathcal{A}(b) - \kappa) < \alpha \cap \mathcal{A}(b) > \kappa \right).
\]

Then using $\Delta_1(\tilde{h})$ in (D.29):
\[
P \left( p_2(\lambda, \tilde{h}; \Delta_1, \Delta_2) < \alpha \cap \mathcal{A}(b) > \kappa \right) \leq P \left( p^{(LF)}(\lambda, \tilde{h}) < \alpha \cap \mathcal{A}(b) > \kappa \right) \quad \text{(D.30)}
\]
\[
\leq P \left( F_{\lambda,\tilde{h}}(T_\psi(\lambda, \tilde{h})) < \alpha \cap \mathcal{A}(b) > \kappa \right).
\]

Combine (D.26), (D.27) and (D.30) to deduce
\[
\sup_{\tilde{h} \in \tilde{\mathcal{H}}} \text{NRP}(\Delta_1(\tilde{h}), 0; \tilde{h}) \leq \sup_{\tilde{h} \in \tilde{\mathcal{H}}} P \left( \sup_{h \in \tilde{\mathcal{H}}} \left\{ F_{\lambda,h}(T_\psi(\lambda, \tilde{h})) \right\} < \alpha \right) = \alpha.
\]

This completes the proof. \textit{QED}.

**Proof of Theorem D.2.** Recall
\[
p^{(LF)}_n(\lambda) = \max \left\{ \sup_{h \in \tilde{\mathcal{H}}} \left\{ F_{\lambda,h}(T_n(\lambda)) \right\}, F_{\infty}(T_n(\lambda)) \right\}
\]
\[
p_{n,1}(\lambda; \Delta_1) \equiv p^{(LF)}_n(\lambda) + \Delta_1
\]
Joint weak convergence

**Step 1.** Under $C(i, b)$ with $||b|| < \infty$, $A_n \overset{p}{\to} A(b)$ where $A(b)$ is defined in Theorem 5.1.a. Joint weak convergence

$$\{T_n(\lambda), A_n : \Lambda \} \Rightarrow^* \{T_\psi(\lambda, h), A(b) : \Lambda \}$$

is shown in Step 2. Therefore, by the mapping theorem and Assumption 5:

$$\{p_{n,1}(\lambda; \Delta_1), p_{n,2}(\lambda; \Delta_1, \Delta_2), A_n : \Lambda \} \Rightarrow^* \left\{ p_1(\lambda, \tilde{h}; \Delta_1), p_2(\lambda, \tilde{h}; \Delta_1, \Delta_2), A(b) : \Lambda \right\}$$

where

$$p_1(\lambda, \tilde{h}; \Delta_1) = \max\left\{ \sup_{h \in \delta} \left\{ \mathcal{F}_{\lambda,h}(T_\psi(\lambda, \tilde{h})) \right\}, \mathcal{F}_\infty(T_\psi(\lambda, \tilde{h})) \right\} + \Delta_1 \equiv p(\text{IF})_{\lambda, \tilde{h}} + \Delta_1$$

$$p_2(\lambda, \tilde{h}; \Delta_1, \Delta_2) \equiv \mathcal{F}_\infty(T_\psi(\lambda, \tilde{h})) + \Delta_2 + \left\{ p_1 - \mathcal{F}_\infty(T_\psi(\lambda, \tilde{h})) - \Delta_2 \right\} s(A(b) - \kappa).$$

The asymptotic size $\text{AsySz}(\lambda)$ is therefore

$$\limsup_{n \to \infty} \sup_{\gamma \in \Gamma^*} P_\gamma \left( p_n^{(\text{ICS}-2)}(\lambda) < \alpha | H_0 \right)$$

$$= \sup_{h \in \delta} P \left( p_1(\lambda, \tilde{h}; \Delta_1) < \alpha \cap A(b) \leq \kappa \right) + \sup_{h \in \delta} P \left( p_2(\lambda, \tilde{h}; \Delta_1, \Delta_2) < \alpha \cap A(b) > \kappa | H_0 \right)$$

$$= \sup_{h \in \delta} NRP(\Delta_1, \Delta_2; \lambda, \tilde{h}),$$

where $NRP$ is the asymptotic Null Rejection Probability defined in (D.25). The tuning parameters $(\Delta_1, \Delta_2)$ are chosen by supposition to ensure $\sup_{h \in \delta} NRP(\Delta_1, \Delta_2; \lambda, \tilde{h}) \leq \alpha$, cf. Lemma D.1.

Under $C(ii, \omega_0)$ we have $A_n \overset{p}{\to} \infty$ by Theorem 5.1.b. Hence $s(A_n - \kappa) \overset{p}{\to} 0$ since the continuous function $s(x) \to 0$ as $x \to \infty$. Now apply Theorem 4.2.b and the mapping theorem to yield $\{p_{n,2}(\lambda; \Delta_1, \Delta_2) : \Lambda \} \Rightarrow^* \{ \mathcal{F}_\infty(T(\lambda)) + \Delta_2 : \Lambda \}$. Since $T(\lambda)$ is distributed $\mathcal{F}_\infty$, it therefore follows:

$$\text{AsySz}(\lambda) = \limsup_{n \to \infty} \sup_{\gamma \in \Gamma^*} P_\gamma \left( p_n^{(\text{ICS}-2)} < \alpha | H_0 \right)$$

$$= P \left( \mathcal{F}_\infty(T(\lambda)) + \Delta_2 < \alpha | H_0 \right) \leq P \left( \mathcal{F}_\infty(T(\lambda)) < \alpha | H_0 \right) = \alpha.$$

**Step 2 (joint convergence).** It remains to prove under $C(i, b), ||b|| < \infty$:

$$\left\{ T_n(\lambda), A_n : \Lambda \} \Rightarrow^* \{ T_\psi(\lambda, h), A(b) : \Lambda \}. $$

28
Recall $S_\beta \equiv [I_{k_\beta} : 0_{k_x \times k_x}]$, and define:

$$\omega(\hat{\beta}_n(\hat{\pi}_n)) = \frac{\sqrt{n}S_\beta \hat{\psi}_n(\hat{\pi}_n)}{\sqrt{\sqrt{n}S_\beta \hat{\psi}_n(\hat{\pi}_n)}} = \frac{\sqrt{n}S_\beta (\hat{\psi}_n(\hat{\pi}_n) - \psi_n)}{\sqrt{\sqrt{n}S_\beta (\hat{\psi}_n(\hat{\pi}_n) - \psi_n)}} + \sqrt{n}\beta_n \equiv \omega_n(\hat{\pi}_n),$$

hence $\omega_n(\hat{\pi}_n)$ is a continuous function of $\sqrt{n}(\hat{\psi}_n(\hat{\pi}_n) - \psi_n)$ and $\hat{\pi}_n$. By the argument leading to (A.12) in the proof of Theorem 4.2:

$$\sup_{\lambda \in \Lambda} \left| \mathcal{T}_n(\lambda) - \frac{(\mathcal{J}_n(\hat{\pi}_n, \lambda) + \mathcal{R}(\hat{\pi}_n, \lambda))^2}{v^2(\omega_n(\hat{\pi}_n), \hat{\pi}_n, \lambda)} \right| \rightarrow^p 0.$$

Recall $\{\mathcal{T}_n(\lambda) : \Lambda\} \Rightarrow^* \{\mathcal{T}_\psi(\lambda, h) : \Lambda\}$ by Theorem 4.2.

By the proof of Theorem 5.1.a and the mapping theorem, $||\bar{\Sigma}_n - \bar{\Sigma}(\pi^*(b), b)|| \rightarrow^p 0$, where $\bar{\Sigma}(\pi, b) \equiv \Sigma(\omega^*(\pi, b), \pi) = \Sigma(||\beta_0||, \omega^*(\pi, b), \zeta_0, \pi)$, and $\Sigma(||\beta||, \omega, \zeta, \pi) = \Sigma(\theta^+) \equiv \mathcal{H}_\theta(\theta^+)^{-1}\mathcal{V}(\theta^+)\mathcal{H}_\theta(\theta^+)^{-1}$.

Therefore

$$\mathcal{A}_n = \left( \frac{1}{p+1} n^{\hat{\beta}_n' \Sigma^{-1}_{\beta, \beta, n} \hat{\beta}_n'} \right)^{1/2}$$

$$= \left( \frac{1}{p+1} (S_\beta \sqrt{n} (\hat{\psi}_n - \psi_n) + \sqrt{n}\beta_n)' \Sigma^{-1}_{\beta, \beta}(\hat{\pi}_n, b) \left( S_\beta \sqrt{n} (\hat{\psi}_n - \psi_n) + \sqrt{n}\beta_n \right) \right)^{1/2} + o_p(1),$$

where $\Sigma_{\beta, \beta}(\pi, b)$ is the upper $(p + 1) \times (p + 1)$ block of $\bar{\Sigma}(\pi, b)$. Further $\mathcal{A}_n \rightarrow^d \mathcal{A}(b)$ by Theorem 5.1.a.

Therefore $\{\mathcal{T}_n(\lambda), \mathcal{A}_n : \Lambda\} \Rightarrow^* \{\mathcal{T}_\psi(\lambda, h), \mathcal{A}(b) : \Lambda\}$ if we prove joint weak convergence for $(\mathcal{J}_n(\pi, \lambda), \sqrt{n}(\hat{\psi}_n(\pi) - \psi_n), \hat{\pi}_n)$ on $\Pi \times \Lambda$. By the proof of Theorem 4.1.a, $\sqrt{n}(\hat{\psi}_n(\pi) - \psi_n)$ and $\hat{\pi}_n$ are continuous functions of $\mathcal{G}_{\psi,n}(\psi_{0,n}, \pi)$ and $\mathcal{H}_{\psi,n}(\pi)$, and $\hat{\mathcal{H}}_{\psi,n}(\pi)$ has a constant limit in probability uniformly on $\Pi$. Joint weak convergence for $(\mathcal{J}_n(\pi, \lambda), \sqrt{n}(\hat{\psi}_n(\pi) - \psi_n), \hat{\pi}_n)$ therefore follows from joint weak convergence for $(\mathcal{J}_n(\pi, \lambda), \mathcal{G}_{\psi,n}(\psi_{0,n}, \pi))$, which is shown in Step 3 in the proof of Theorem 4.2. $\mathcal{QED}$.

### E Robust Critical Values

In this appendix we present bootstrapped identification category robust critical values. The idea is based on unobserved robust Least Favorable, and type 1 and 2 identification category selection [ICS] critical values presented in Andrews and Cheng (2012a).
E.1 Least Favorable and ICS Critical Values

We remind the reader of our notation conventions. As in Andrews and Cheng (2012a, 2013a), technical results are derived under two overlapping cases which align with Categories I-III in Table 1, below: (i) $\beta_n \to \beta_0 = 0$ and $\sqrt{n} \beta_n \to (\mathbb{R} \cup \{\pm \infty\})^{p+1}$; and (ii) $\beta_0 \in \mathbb{R}^{p+1}$ and $\beta_n / ||\beta_n|| \to \omega_0 \in \mathbb{R}^{p+1}$ with $||\omega_0|| = 1$. Let $\{T_\psi(\lambda, b) : \lambda \in \Lambda\}$ denote the null limit process of $T_n(\lambda)$ under weak identification $\sqrt{n} \beta_n \to b$ with $||b|| < \infty$ (see Theorem 5.2). Recall that $\phi_0$ indexes all remaining (nuisance) parameters such that the distribution of $W_t \equiv [y_t, y_{t-1}, \ldots, y_{t-p}]'$ is determined by:

$$
\gamma_0 \equiv (\theta_0, \phi_0) \in \Gamma^* \equiv \{\theta \in \Theta^*, \phi \in \Phi^*(\theta)\}. \tag{E.31}
$$

Assume $\Phi^*(\theta) \subset \Phi^* \forall \theta \in \Theta^*$, where $\Phi^*$ is a compact metric space with some metric that induces weak convergence of the bivariate distributions of $(W_t, W_{t+h})$ for all $t$ and $h \geq 1$.

Define the parametric set that characterizes data generating processes under weak identification $\beta_n \to \beta_0 = 0$, and $\sqrt{n} \beta_n \to b$ with $||b|| < \infty$:

$$
h \equiv (\gamma_0, b) \in \mathcal{H} \equiv \{h : \gamma_0 \in \Gamma^*, \text{ and } ||b|| < \infty, \text{ with } \beta_0 = 0\}. \tag{E.32}
$$

Now let $\{T_\psi(\lambda, h) : \lambda \in \Lambda\}$ denote the non-standard null limit process under weak identification. Under strong identification the null limit law is $\chi^2(1)$. Let $c_{1-\alpha}(\lambda, h)$ and $\chi^2_{1-\alpha}$ respectively be the $1 - \alpha$ quantiles for $T_\psi(\lambda, h)$ and $\chi^2(1)$. All subsequent critical values are functions of $c_{1-\alpha}(\lambda, h)$, hence in Appendix E.3 we discuss how to compute $c_{1-\alpha}(\lambda, h)$.

The following summarizes ideas developed in Andrews and Cheng (2012a, Section 5).

E.1.1 Least Favorable Critical Value

The Least favorable [LF] critical value is

$$
c^{(LF)}_{1-\alpha}(\lambda) \equiv \max \left\{ \sup_{h \in \mathcal{H}} \{c_{1-\alpha}(\lambda, h)\}, \chi^2_{1-\alpha} \right\}. \tag{E.33}
$$


A better critical value in terms of power uses the fact that \((\hat{\zeta}_0, \beta_n)\) are consistently estimated by \((\hat{\zeta}_n, \hat{\beta}_n)\) under any degree of (non)identification. The plug-in LF critical value \(c^{(LF)}_1(\lambda)\) uses \(\widehat{\mathcal{H}} \equiv \{h \in \mathcal{H} : \theta = [\hat{\zeta}_n', \hat{\beta}_n', \pi']\}\) in place of \(\mathcal{H}\).

In the present environment the null hypothesis is tested by using a sample version of \(E[\epsilon_t F(\lambda' W(x_t))]\). Thus, so-called parametric null imposed critical values in Andrews and Cheng (2012a) for t-, Quasi-Likelihood Ratio and Wald statistics do not play a role here.

**E.1.2 Identification Category Selection Type 1**

The LF critical value does not exploit data related information that may point toward a particular identification case. The ICS procedure uses the sample to choose between \(\sqrt{n} \beta_n \rightarrow b\) when \(||b|| < \infty\) (weak and non-identification) and \(||b|| = \infty\) (semi-strong and strong identification).

Recall the statistic \(A_n\) in (A.6). Now let \(\{\kappa_n\}\) be a sequence of positive constants, with \(\kappa_n \rightarrow \infty\) and \(\kappa_n = o(n^{1/2})\). The case \(||b|| < \infty\) is selected when \(A_n \leq \kappa_n\), else \(||b|| = \infty\) is selected. Now define the type 1 ICS [ICS-1] critical value:

\[
c^{(ICS-1)}_{1-\alpha,n}(\lambda) = \begin{cases} 
  c^{(LF)}_{1-\alpha}(\lambda) & \text{if } A_n \leq \kappa_n \\
  \chi^2_{1-\alpha} & \text{if } A_n > \kappa_n 
\end{cases}
\]

See the remark following Theorem 6.1, and Andrews and Cheng (2012a, p. 2191), for intuition on \(c^{(ICS-1)}_{1-\alpha,n}(\lambda)\). Briefly: only when \(\sqrt{n}||\beta_n|| \rightarrow \infty\) faster than \(\kappa_n \rightarrow \infty\) will the chi-squared based critical value be chosen asymptotically with probability approaching one since then \(A_n/\kappa_n \uparrow \infty\). Thus, a high bar must be passed in order to select the strong identification case. In every other case the LF value is chosen, which is always asymptotically correct.

**E.1.3 Identification Category Selection Type 2**

Let \(s : [0, \infty) \rightarrow [0, 1]\) be a continuous function, \(s(x)\) is non-increasing in \(x\), \(s(0) = 1\), and \(s(x) \rightarrow 0\) as \(x \rightarrow \infty\). An example is \(s(x) = \exp\{-cx\}\) for some \(c > 0\). Let \((\Delta_1, \Delta_2) \geq 0\) and \(\kappa > 0\) be user selected numbers. Define

\[
c_1(\lambda) = c^{(LF)}_{1-\alpha}(\lambda) + \Delta_1
\]
\[
c_2(\lambda) = \chi^2_{1-\alpha} + \Delta_2 + (c^{(LF)}_{1-\alpha}(\lambda) - \chi^2_{1-\alpha} + \Delta_1 - \Delta_2)s(A_n - \kappa).
\]

The type 2 ICS [ICS-2] critical value is

\[
c^{(ICS-2)}_{1-\alpha,n}(\lambda) = \begin{cases} 
  c_1(\lambda) & \text{if } A_n \leq \kappa \\
  c_2(\lambda) & \text{if } A_n > \kappa 
\end{cases}
\]
The construction allows for a smooth transition between identification cases, and allows for a non-diverging threshold. The latter necessitates the tuning parameters \((\Delta_1, \Delta_2)\) which promote a correct asymptotic size. See also Andrews and Barwick (2012) for a related method.

See Andrews and Cheng (2012a, p. 2193) for details on determining appropriate choices for \((\Delta_1, \Delta_2, \kappa)\), and see Appendix D above.

### E.2 Asymptotics for Robust Critical Values

Let \(c_{1-\alpha,n}^{(i)}(\lambda)\) denote the LF, ICS-1 or ICS-2 plug-in robust critical value. Conditions leading to critical value asymptotics follow, and are presented in Andrews and Cheng (2012a, Section 5) and Andrews and Cheng (2013a, Section 5.5).

**Assumption 7** (critical value). If \(c_{1-\alpha,n}^{(i)}(\lambda)\) is (i) LF, (ii) ICS-1, or (iii) ICS-2, then assume respectively that Andrews and Cheng’s (2012a) Assumption (i) LF, (ii) K and V3, or (iii) Rob2 holds.

Let \(F_{\gamma}\) be the distribution function of \(W_t\) under some \(\gamma \in \Gamma^*\), where \(\Gamma^*\) is the true parameter space in (E.31). Let \(P_{\gamma}\) denote probability under \(F_{\gamma}\). For any critical value \(c_{1-\alpha,n}^{(i)}(\lambda)\) and each \(\lambda\) the asymptotic size of the test is the maximum rejection probability over \(\gamma\) such that the null is true:

\[
\text{AsySz}(\lambda) = \lim_{n \to \infty} \sup_{\gamma \in \Gamma^*} P_{\gamma}(T_n(\lambda) > c_{1-\alpha,n}^{(i)}(\lambda)|H_0).
\]

Proofs are presented in Appendix E.4.

**Theorem E.1.** Under Assumptions 1-2, 4 and 7 and \(H_0\), the LF, ICS-1 and ICS-2 \(c_{1-\alpha,n}^{(i)}(\lambda)\) satisfy \(\text{AsySz}(\lambda) = \alpha\).

### E.3 Computation of \(c_{1-\alpha,n}^{(i)}(\lambda)\)

Steps 1-4 of the wild bootstrap procedure outlined in Section 6.2 of the main paper carries over verbatim. We repeat them here for ease of reference for a proof below.

**Step 1: Compute components \(\mathcal{H}_\psi(\pi), \mathcal{D}_\psi(\pi), \) etc.**

Recall \(d_{\psi,t}(\pi) \equiv [g(x_t, \pi)' , x_t']\), \(d_{\theta,t}(\omega, \pi) \equiv [d_{\psi,t}(\pi)', \omega' x_t'(\partial/\partial \pi) g(x_t, \pi)]'\) and \(K_{\psi,t}(\pi, \lambda) \equiv F(\lambda' W(x_t)) - b_{\psi,\lambda}(\pi)' \mathcal{H}^{-1}_{\psi}(\pi) d_{\psi,t}(\pi)\). Define \(\epsilon_t(\psi, \pi) \equiv y_t - \zeta' x_t - \beta' g(x_t, \pi)\) and estimators

\[
\mathcal{\hat{H}}_{\psi,n}(\pi) \equiv \frac{1}{n} \sum_{t=1}^{n} d_{\psi,t}(\pi) d_{\psi,t}(\pi)' \quad \text{and} \quad \mathcal{\hat{H}}_n(\omega, \pi) = \frac{1}{n} \sum_{t=1}^{n} d_{\theta,t}(\omega, \pi) d_{\theta,t}(\omega, \pi)'.
\]
\[ \hat{D}_{\psi,n}(\pi, \pi_0) \equiv -\frac{1}{n} \sum_{t=1}^{n} d_{\psi,t}(\pi)g(x_t, \pi_0)' \text{ and } \hat{\mathcal{C}}_{\psi,n,t}(\pi, \lambda) \equiv F(\lambda' \mathcal{W}(x_t)) - \hat{b}_{\psi,n}(\pi, \lambda)' \hat{H}_{\psi,n}^{-1}(\pi) d_{\psi,t}(\pi) \]

\[ \hat{b}_{\psi,n}(\pi, \lambda) \equiv \frac{1}{n} \sum_{t=1}^{n} F(\lambda' \mathcal{W}(x_t)) d_{\psi,t}(\pi) \text{ and } \hat{b}_{\theta,n}(\omega, \pi, \lambda) \equiv \frac{1}{n} \sum_{t=1}^{n} F(\lambda' \mathcal{W}(x_t)) d_{\theta,t}(\omega, \pi). \]

**Step 2: Draw from \( \pi^*(b) \)**

By Assumption 2 \( \pi^*(b) \equiv \arg \inf_{\pi \in \Pi} \xi(\pi, b) \equiv -\arg \inf_{\pi \in \Pi} \{ S_{b} \mathcal{H}_{\psi,1}^{-1}(\pi)(\mathcal{G}_{\psi}(\pi) + D_{\psi}(\pi)b) \} \). Under weak identification, Lemma B.1 yields that \( \{ \mathcal{G}_{\psi}(\pi) : \pi \in \Pi \} \) is the weak limit of

\[ \mathcal{G}_{\psi,n}(\psi_0, \pi) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left\{ \epsilon_t(\psi_0, \pi) d_{\psi,t}(\pi) - E[\epsilon_t(\psi_0, \pi) d_{\psi,t}(\pi)] \right\} \tag{E.33} \]

\[ = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \epsilon_t d_{\psi,t}(\pi) + \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left\{ d_{\psi,t}(\pi)g(x_t, \pi_0)' - E[d_{\psi,t}(\pi)g(x_t, \pi_0)'] \right\} \times b. \]

By the argument used to prove Lemma B.2,

\[ \sup_{\pi \in \Pi} \left\{ \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left\{ d_{\psi,t}(\pi)g(x_t, \pi_0)' - E[d_{\psi,t}(\pi)g(x_t, \pi_0)'] \right\} \right\} \overset{p}{\to} 0. \]

Hence, \( E[\epsilon_t | x_t] = 0 \) a.s. and \( E[\epsilon_t^2 | x_t] = \sigma_0^2 \in (0, \infty) \) a.s. under \( H_0 \), the covariance for \( \mathcal{G}_{\psi}(\pi) \) is

\[ E[\epsilon_t^2 d_{\psi,t}(\pi)d_{\psi,t}(\pi)'] = E[\epsilon_t^2] \times E[d_{\psi,t}(\pi)d_{\psi,t}(\pi)'] = \sigma_0^2 \times \mathcal{H}_{\psi}(\pi, \pi), \]

say. Thus \( \mathcal{H}_{\psi,1/2}(\pi)\mathcal{G}_{\psi}(\pi) \) is distributed \( N(0, \sigma_0^2) \) with kernel \( \sigma_0^2 \mathcal{H}_{\psi,1/2}(\pi) \times \mathcal{H}_{\psi}(\pi, \pi) \times \mathcal{H}_{\psi,1/2}(\pi) \).

Next, let \( \{ z_t \}_{t=1}^{n} \) be a sequence of independent draws from \( N(0, 1) \), and define \( \mathcal{G}_{\psi,n}(\pi) \equiv 1/\sqrt{n} \sum_{t=1}^{n} z_t \hat{H}_{\psi,n}^{-1/2}(\pi) d_{\psi,t}(\pi) \). By \( \sigma_n \overset{p}{\to} \sigma_0 \) and the proof of Theorem 6.2, \( \{ \sigma_n \mathcal{G}_{\psi,n}(\pi) : \pi \in \Pi \} \]
\[ \overset{p}{\Rightarrow} \{ \mathcal{H}_{\psi,1/2}(\pi)\mathcal{G}_{\psi}(\pi) : \pi \in \Pi \}, \] where \( \Rightarrow \) denotes weak convergence in probability defined in Gine and Zinn (1990, Section 3).\(^3\) Thus, \( \hat{\sigma}_n \mathcal{G}_{\psi,n}(\pi) \) is a draw from \( \{ \mathcal{H}_{\psi,1/2}(\pi)\mathcal{G}_{\psi}(\pi) : \pi \in \Pi \} \) with probability approaching one as \( n \to \infty \).

Now use \( \{ \hat{G}_{\psi,n}(\pi), \hat{\mathcal{C}}_{\psi,n}(\pi), \hat{D}_{\psi,n}(\pi, \pi_0) \} \) to compute

\[ \hat{\xi}_{\psi,n}(\pi, \pi_0, b) = -\frac{1}{2} \left\{ \hat{\sigma}_n \hat{G}_{\psi,n}(\pi) + \hat{\mathcal{C}}_{\psi,n}(\pi, \pi_0) \times b \right\} ' \left\{ \hat{\sigma}_n \hat{G}_{\psi,n}(\pi) + \hat{\mathcal{C}}_{\psi,n}(\pi, \pi_0) \times b \right\} . \]

\(^3\)Gine and Zinn (1990) work under weak convergence in \( l_\infty \) as in Hoffman-Jørgensen (1991), which is the same rubric of weak convergence that we use. Thus, for example, \( \{ \sigma \hat{G}_{\psi,n}(\pi) : \pi \in \Pi \} \Rightarrow \{ \mathcal{H}_{\psi,1/2}(\pi)\mathcal{G}_{\psi}(\pi) : \pi \in \Pi \} \) if and only if \( \{ \sigma \hat{G}_{\psi,n}(\pi) : \pi \in \Pi \} \Rightarrow \{ \mathcal{H}_{\psi,1/2}(\pi)\mathcal{G}_{\psi}(\pi) : \pi \in \Pi \} \) asymptotically with probability approaching one with respect to the sample draw.
The bootstrapped \( \pi^*(b) \) is therefore:

\[
\hat{\pi}^*_n(\pi_0, b) = \arg \min_{\pi \in \Pi} \left\{ \hat{\xi}^*_{\psi, n}(\pi, \pi_0, b) \right\}.
\] (E.34)

**Step 3: Draw from \( \mathcal{Z}_\psi(\pi, \lambda) \)**

Write \( \epsilon_t(\psi, \pi) \equiv y_t - \zeta^* x_t - \beta_n' g(x_t, \pi) \). By Lemma B.9.a, under the null \( \mathcal{Z}_\psi(\pi, \lambda) \) is the zero mean Gaussian limit process of \( 1/\sqrt{n} \sum_{t=1}^n \epsilon_t \mathcal{K}_\psi(t, \pi, \lambda) \), where \( \mathcal{K}_\psi(t, \pi, \lambda) \equiv F(\lambda' \mathcal{W}(x_t)) - b_\psi(\pi, \lambda)' \mathcal{H}^{-1}_\psi(\pi) \times d_\psi(t, \pi) \). Use the Step 2 draws \( \{z_t\}_{t=1}^n \) to define

\[
\hat{\mathcal{Z}}^*_n(\pi, \lambda) \equiv \frac{1}{\sqrt{n}} \sum_{t=1}^n z_t \left( F(\lambda' \mathcal{W}(x_t)) - \hat{b}_\psi(n, \pi, \lambda)' \hat{\mathcal{H}}^{-1}_\psi(\pi)d_\psi(t, \pi) \right).
\] (E.35)

Then \( \{\hat{\sigma}_n \hat{\mathcal{Z}}^*_n(\pi, \lambda) : \pi \in \Pi, \lambda \in \Lambda \} \Rightarrow \{ \mathcal{Z}_\psi(\pi, \lambda) : \pi \in \Pi, \lambda \in \Lambda \} \), hence \( \hat{\sigma}_n \hat{\mathcal{Z}}^*_n(\pi, \lambda) \) is the bootstrap draw from \( \mathcal{Z}_\psi(\pi, \lambda) \).

**Step 4: \( \tau_\beta(\cdot), \hat{\mathcal{X}}_\psi(\cdot), \hat{v}_\theta^2(\cdot), \mathcal{T}_\psi(\cdot) \)**

We now have all the required components for computing the following key quantities (recall \( \mathcal{S}_\beta \equiv [I_{k_\beta} : 0_{k_\times k_{2\times}}] \)):

\[
\hat{\tau}_{\beta,n}(\pi_0, b) \equiv -\mathcal{S}_\beta \hat{\mathcal{H}}^{-1}_{\psi,n}(\hat{\pi}^*_n(\pi_0, b)) \left\{ \hat{\sigma}_n \hat{\mathcal{G}}^*_{\psi,n}(\hat{\pi}^*_n(\pi_0, b)) + \hat{\mathcal{D}}_{\psi,n}(\hat{\pi}^*_n(\pi_0, b), \pi_0) \times b \right\}
\] (E.36)

\[
\hat{\omega}^*_n(\pi_0, b) = \frac{\hat{\tau}_{\beta,n}(\pi_0, b)}{\|\hat{\tau}_{\beta,n}(\pi_0, b)\|}
\]

\[
\hat{\mathcal{X}}^*_{\psi,n}(\pi, \lambda, \pi_0, b)
\]

\[
\equiv \hat{\sigma}_n \hat{\mathcal{Z}}^*_{\psi,n}(\pi, \lambda) + \hat{b}_\psi(n, \pi, \lambda)' \left( \hat{\mathcal{H}}^{-1}_{\psi,n}(\pi) \hat{\mathcal{D}}_{\psi,n}(\pi, \pi_0) \times b + [b, 0_{k_\beta}']' \right)
\]

\[
+ \hat{b}_\psi(n, \pi, \lambda)' \hat{\mathcal{H}}^{-1}_{\psi,n}(\pi) \frac{1}{n} \sum_{t=1}^n d_\psi(t, \pi) \{ g(x_t, \pi_0) - g(x_t, \pi) \}' b
\]

\[
+ \frac{1}{n} \sum_{t=1}^n \left\{ F(\lambda' \mathcal{W}(x_t)) - \hat{b}_\psi(n, \pi, \lambda)' \hat{\mathcal{H}}^{-1}_n(\lambda)d_\psi(t, \pi) \right\} \{ g(x_t, \pi_0) - g(x_t, \pi) \}' b
\]

\[
\hat{v}_n^2(\omega, \pi, \lambda) \equiv \frac{1}{n} \sum_{t=1}^n \epsilon_t^2(\hat{\psi}_n, \pi) \left\{ F(\lambda' \mathcal{W}(x_t)) - \hat{b}_{\theta,n}(\omega, \pi, \lambda)' \hat{\mathcal{H}}^{-1}_n(\omega, \pi)d_\theta(t, \omega, \pi) \right\}^2
\]

\[
\hat{v}_n^2(\pi, \lambda, b) \equiv v_n^2(\hat{\omega}^*_n(\pi_0, b), \pi, \lambda).
\]
The bootstrap draw from $\mathcal{T}_\psi(\pi^*(b), \lambda, b)$ is $\hat{T}_{\psi,n}^*(\lambda, h) = \hat{T}_{\psi,n}^*(\pi, \pi_0, b) = \hat{T}_{\psi,n}(\hat{\pi}_n^*(\pi_0, b), \lambda, \pi_0, b)$ where

$$
\hat{T}_{\psi,n}^*(\pi, \lambda, \pi_0, b) \equiv \left( \frac{\hat{\xi}_{\psi,n}(\pi, \lambda, \pi_0, b)}{\hat{v}_n(\pi, \lambda, b)} \right)^2.
$$

(E.37)

Notice $h = (\pi_0, b)$ are nuisance parameters that cannot be consistently estimated under weak identification $\sqrt{n}\|\beta_n\| \to [0, \infty)$.

**Step 5**

Repeat Steps 1-4 $M$ times resulting in a sequence of independent draws $\{\hat{T}_{\psi,n,j}(\lambda, h)\}_{j=1}^M$. Define order statistics $\hat{T}_{\psi,n,[1]}(\lambda, h) \leq \hat{T}_{\psi,n,[2]}(\lambda, h) \leq \cdots$. The critical value approximation is $\hat{c}_{1-\alpha,n,M}(\lambda, h) \equiv \hat{T}_{\psi,n,[1-(1-\alpha)M]}(\lambda, h)$, which is consistent for the asymptotic critical value $c_{1-\alpha}(\lambda, h)$.

**Theorem E.2.** Let the true value $\sigma^2 \equiv E[\varepsilon_t^2] \in \mathcal{S}^*$, where the true parameter space $\mathcal{S}^*$ is a compact subset of $(0, \infty)$. Let $M = M_n \to \infty$ as $n \to \infty$. Under Assumptions 1-2, 4 and 7, $\hat{c}_{1-\alpha,n,M_n}(\lambda, h) \xrightarrow{p} c_{1-\alpha}(\lambda, h)$ for each $h \in \mathcal{H}$ and $\lambda \in \Lambda$.

**E.4 Proofs**

**Proof of Theorem E.1.**

**Step 1 (LF).** The proof for the LF critical value $c^{(LF)}_{1-\alpha} = \max\{\sup_{h \in \mathcal{H}}\{c_{1-\alpha}(\lambda, h)\}, \chi^2_{1-\alpha}\}$ is identical to arguments in Andrews and Cheng (2012b, Appendix B: proof of Theorem 5.1). We verify the conditions of Lemma 2.1 in Andrews and Cheng (2012a) below. An application of their Lemma 2.1 to the asymptotic size for $\mathcal{T}_\psi(\lambda, h)$, and Theorem 4.2.a, yields

$$
\text{AsySz}(\lambda) = \max\left\{\sup_{h \in \mathcal{H}} P\left(\mathcal{T}_\psi(\lambda, h) > c^{(LF)}_{1-\alpha}\right), P\left(\mathcal{T}(\lambda) > c^{(LF)}_{1-\alpha}\right)\right\}.
$$

If $c^{(LF)}_{1-\alpha} = \chi^2_{1-\alpha}$ then by the definition of $c_{1-\alpha}(\lambda, h)$:

$$
\sup_{h \in \mathcal{H}} P\left(\mathcal{T}_\psi(\lambda, h) > c^{(LF)}_{1-\alpha}\right) \leq \sup_{h \in \mathcal{H}} P\left(\mathcal{T}_\psi(\lambda, h) > c_{1-\alpha}(\lambda, h)\right) = \alpha,
$$

hence $\text{AsySz}(\lambda)$ is:

$$
\max\left\{\sup_{h \in \mathcal{H}} P\left(\mathcal{T}_\psi(\lambda, h) > \chi^2_{1-\alpha}\right), P\left(\mathcal{T}(\lambda) > \chi^2_{1-\alpha}\right)\right\} = \max\left\{\sup_{h \in \mathcal{H}} P\left(\mathcal{T}_\psi(\lambda, h) > \chi^2_{1-\alpha}\right), \alpha\right\} = \alpha.
$$
Conversely, if $c_{1-\alpha}^{(LF)} = \sup_{h \in \Theta} \{c_{1-\alpha}(\lambda, h)\}$ then

$$\sup_{h \in \Theta} P\left( T_\psi(\lambda, h) > c_{1-\alpha}^{(LF)} \right) = \sup_{h \in \Theta} P\left( T_\psi(\lambda, h) > \sup_{h \in \Theta} \{c_{1-\alpha}(\lambda, h)\} \right) = \alpha,$$

and $P(T(\lambda) > c_{1-\alpha}^{(LF)}) \leq \alpha$ hence again $AsySz(\lambda) = \alpha$.

It remains to verify the conditions of Lemma 2.1 in Andrews and Cheng (2012a). We must show their Assumption ACP holds, parts (i)-(iv). Recall $\{\gamma_n\}$ is a sequence of true parameters $\gamma_n \equiv (\theta_n, \phi_0)$ under local drift which fully determine the joint distribution of the data $[y_t, y_{t-1}, ..., y_{t-p}]'$. The limiting true value is $\gamma_0 \equiv (\theta_0, \phi_0)$. By Theorem 4.2, $P_{\gamma_n}(T_n(\lambda) > c_{1-\alpha}^{(LF)}) \to P(T(\lambda) > c_{1-\alpha}^{(LF)})$ under $C(i, b)$ with $||b|| < \infty$, and $P_{\gamma_n}(T_n(\lambda) > c_{1-\alpha}^{(LF)}) \to P(T(\lambda) > c_{1-\alpha}^{(LF)})$ under $C(ii, \omega_0)$. Hence Assumption ACP.i.i,ii,iii hold. Assumption ACP.iv holds under true parameter space Assumption 1.e, because the latter is identically Assumption STAR4 in Andrews and Cheng (2013a). See also Andrews and Cheng (2013b, Section 15.7).

**Step 2 (ICS1).** Theorem 5.1 implies the ICS statistic satisfies $A_n = O_p(1)$ under $C(i, b)$ with $||b|| < \infty$. Under $C(ii, \omega_0)$ we have $A_n \law \infty$, and if $\beta_0 \neq 0$ then $\kappa_n^{-1}A_n \law \infty$ where by supposition $\kappa_n \to \infty$ and $\kappa_n = o(\sqrt{n})$. Now invoke Theorem 4.2 to deduce $P_{\gamma_n}(T_n(\lambda) > c_{1-\alpha,n}^{(ICS^{-1})}(\lambda)) \to P(T(\lambda) > c_{1-\alpha,n}^{(ICS^{-1})}(\lambda))$ under $C(i, b)$ with $||b|| < \infty$, and $P_{\gamma_n}(T_n(\lambda) > c_{1-\alpha}^{(ICS^{-1})}) \to P(T(\lambda) > \chi_1^{2})$ under $C(ii, \omega_0)$ if $\beta_0 \neq 0$. Hence Assumption ACP.i,ii,iii in Andrews and Cheng (2012a) hold. Their Assumption ACP.iv holds by Step 1. Arguments in Andrews and Cheng (2012b, p. 56-58) now carry over to prove the ICS-1 and ICS-2 claims. \textit{QED}.

**Proof of Theorem E.2.** By Step 1 in the proof of Theorem 6.2:

$$\left\{ \hat{T}_{\psi,n}^*(\lambda, h) : \lambda \in \Lambda \right\} \law \left\{ \left( \frac{\delta_{\psi}(\pi^*(b), \lambda, b)}{v(\pi^*(b), \lambda, b)} \right)^2 : \lambda \in \Lambda \right\} = \left\{ T_\psi(\lambda, h) : \lambda \in \Lambda \right\}, \quad (E.38)$$

the Theorem 5.2 null limit process under weak identification.

Define quantile functions

$$F_{\alpha}^{-1}(u) \equiv \inf \left\{ c \geq 0 : \Pr(T_\psi \leq c) \geq u \right\},$$

$$F_{\alpha/n}^{-1}(u) \equiv \inf \left\{ c \geq 0 : \Pr(T_n \leq c) \geq u \right\},$$

$$F_{\lambda,h}^{-1}(u) \equiv \inf \left\{ c \geq 0 : \Pr(T_\psi(\lambda, h) \leq c) \geq u \right\},$$

By Theorem 5.2.a, $\{T_n(\lambda) : \Lambda \} \law \{T_\psi(\lambda, h) : \Lambda \}$ under $H_0$ and $C(i, b)$ with $||b|| < \infty$. Weak convergence implies convergence in finite dimensional distribution. By the construction of distribution convergence it therefore follows that $F_{\alpha,n}^{-1}(u) \to F_{\lambda,h}^{-1}(u)$.
Now operate conditionally on the sample $\mathfrak{W}_n$. By weak convergence in probability (E.38), \( \{ \hat{T}^{*}_{\psi,n,j}(\lambda, h) \}_{j=1}^{M} \) is a sequence of iid draws from \( \{ T_{\psi}(\lambda, h) : \Lambda \} \), asymptotically with probability approaching one with respect to the draw $\mathfrak{W}_n \equiv \{(y_t, x_t)\}_{t=1}^{n}$. Therefore $T_n(\lambda)$ under $C(i, b)$ with \(|b| < \infty\), and $\hat{T}^{*}_{\psi,n,1}(\lambda, h)$ have the same weak limits in probability under $H_0$. Since $T_n(\lambda)$, and $\hat{T}^{*}_{\psi,n,j}(\lambda, h)$ conditionally on $\mathfrak{W}_n$ have the same weak limits in probability under $H_0$, it follows that (see Gine and Zinn, 1990, Section 3, eqs (3.4) and (3.5))

$$\sup_{c \geq 0} \left| P(\hat{T}^{*}_{\psi,n,j}(\lambda, h) \leq c|\mathfrak{W}_n) - F_{n,\lambda}(c) \right| \xrightarrow{p} 0 \forall \lambda \in \Lambda$$

Therefore, by construction of convergence of probability measures (see, e.g., Chapt. 21 in van der Vaart, 1998):

$$\sup_{u \in [0,1]} \left| \hat{F}_{n,\lambda}^{-1}(u|\mathfrak{W}_n) - F_{n,\lambda}^{-1}(u) \right| \xrightarrow{p} 0 \forall \lambda \in \Lambda.$$ Moreover, by independence and $M_n \to \infty$, the bootstrapped critical value $\hat{c}_{1-\alpha,n,M_n}(\lambda, h) \equiv \hat{T}^{*}_{\psi,n,[(1-\alpha),M_n]}(\lambda, h)$ is a central order statistic of a (conditionally) iid random variable, hence pointwise on $\Lambda$:

$$\left| \hat{c}_{1-\alpha,n,M_n}(\lambda, h) - \hat{F}_{n,\lambda}^{-1}(1 - \alpha|\mathfrak{W}_n) \right| \xrightarrow{p} 0$$

See, e.g., Galambos (1987), for a classic treatment of order statistics. Now combine

$$\left| \hat{c}_{1-\alpha,n,M_n}(\lambda, h) - \hat{F}_{n,\lambda}^{-1}(1 - \alpha|\mathfrak{W}_n) \right| \xrightarrow{p} 0$$

$$\left| \hat{F}_{n,\lambda}^{-1}(1 - \alpha|\mathfrak{W}_n) - F_{n,\lambda}^{-1}(1 - \alpha) \right| \xrightarrow{p} 0$$

$$F_{n,\lambda}^{-1}(1 - \alpha) \to F_{\lambda,h}^{-1}(1 - \alpha)$$

to yield

$$\left| \hat{c}_{1-\alpha,n,M_n}(\lambda, h) - F_{\lambda,h}^{-1}(1 - \alpha) \right| \xrightarrow{p} 0.$$ By definition $c_{1-\alpha}(\lambda, h) = F_{\lambda,h}^{-1}(1 - \alpha)$ hence the proof is complete. QED.

**F Bibliography**


<table>
<thead>
<tr>
<th></th>
<th>$H_0$: LSTAR</th>
<th>$H_1$-weak</th>
<th>$H_1$-strong</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1%</td>
<td>5%</td>
<td>10%</td>
</tr>
<tr>
<td>Strong Identification: $\beta_n = .3$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>supremum</td>
<td>.025</td>
<td>.094</td>
<td>.163</td>
</tr>
<tr>
<td>average</td>
<td>.025</td>
<td>.078</td>
<td>.135</td>
</tr>
<tr>
<td>random</td>
<td>.011</td>
<td>.052</td>
<td>.096</td>
</tr>
<tr>
<td>random LF</td>
<td>.007</td>
<td>.015</td>
<td>.038</td>
</tr>
<tr>
<td>random ICS-1</td>
<td>.013</td>
<td>.050</td>
<td>.089</td>
</tr>
<tr>
<td>PVOT*</td>
<td>.015</td>
<td>.065</td>
<td>.124</td>
</tr>
<tr>
<td>PVOT LF</td>
<td>.007</td>
<td>.014</td>
<td>.052</td>
</tr>
<tr>
<td>PVOT ICS-1</td>
<td>.007</td>
<td>.043</td>
<td>.073</td>
</tr>
<tr>
<td>Weak Identification: $\beta_n = 3/\sqrt{n}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>supremum</td>
<td>.064</td>
<td>.155</td>
<td>.239</td>
</tr>
<tr>
<td>average</td>
<td>.057</td>
<td>.146</td>
<td>.219</td>
</tr>
<tr>
<td>random</td>
<td>.027</td>
<td>.083</td>
<td>.175</td>
</tr>
<tr>
<td>random LF</td>
<td>.012</td>
<td>.042</td>
<td>.093</td>
</tr>
<tr>
<td>random ICS-1</td>
<td>.012</td>
<td>.046</td>
<td>.104</td>
</tr>
<tr>
<td>PVOT</td>
<td>.038</td>
<td>.127</td>
<td>.196</td>
</tr>
<tr>
<td>PVOT LF</td>
<td>.015</td>
<td>.049</td>
<td>.108</td>
</tr>
<tr>
<td>PVOT ICS-1</td>
<td>.014</td>
<td>.049</td>
<td>.107</td>
</tr>
<tr>
<td>Non-Identification: $\beta_n = \beta_0 = 0$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>supremum</td>
<td>.066</td>
<td>.164</td>
<td>.249</td>
</tr>
<tr>
<td>average</td>
<td>.062</td>
<td>.148</td>
<td>.226</td>
</tr>
<tr>
<td>random</td>
<td>.044</td>
<td>.107</td>
<td>.186</td>
</tr>
<tr>
<td>random LF</td>
<td>.013</td>
<td>.046</td>
<td>.115</td>
</tr>
<tr>
<td>random ICS-1</td>
<td>.013</td>
<td>.047</td>
<td>.116</td>
</tr>
<tr>
<td>PVOT</td>
<td>.049</td>
<td>.134</td>
<td>.190</td>
</tr>
<tr>
<td>PVOT LF</td>
<td>.015</td>
<td>.061</td>
<td>.117</td>
</tr>
<tr>
<td>PVOT ICS-1</td>
<td>.015</td>
<td>.057</td>
<td>.116</td>
</tr>
</tbody>
</table>

Numerical values are rejection frequency at the given level. LSTAR is Logistic STAR. Empirical power is not size-adjusted. \textit{supremum} and \textit{average} tests are based on a wild bootstrapped p-value. \textit{random}: $T_n(\lambda)$ with randomized $\lambda$ on [1,5]. \textit{PVOT}: p-value occupation time test. PVOT uses the chi-squared distribution, LF is the least favorable p-value, and ICS-1 is the type I identification category selection p-value with threshold $\kappa_n = \ln(\ln(n))$. 
Table 3: STAR Test Rejection Frequencies: Sample Size $n = 250, \sigma = 1$

<table>
<thead>
<tr>
<th></th>
<th>$H_0$: LSTAR</th>
<th>$H_1$-weak</th>
<th>$H_1$-strong</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1% 5% 10%</td>
<td>1% 5% 10%</td>
<td>1% 5% 10%</td>
</tr>
<tr>
<td>Strong Identification: $\beta_n = .3$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>supremum</td>
<td>.018 .088 .163</td>
<td>.359 .468 .551</td>
<td>.953 .984 .990</td>
</tr>
<tr>
<td>average</td>
<td>.014 .077 .133</td>
<td>.262 .387 .468</td>
<td>.873 .949 .975</td>
</tr>
<tr>
<td>random</td>
<td>.014 .064 .126</td>
<td>.165 .299 .396</td>
<td>.793 .912 .952</td>
</tr>
<tr>
<td>random LF</td>
<td>.001 .010 .025</td>
<td>.067 .235 .368</td>
<td>.688 .888 .936</td>
</tr>
<tr>
<td>random ICS-1</td>
<td>.008 .031 .077</td>
<td>.076 .244 .375</td>
<td>.762 .902 .947</td>
</tr>
<tr>
<td>PVOT</td>
<td>.016 .067 .125</td>
<td>.328 .437 .517</td>
<td>.952 .983 .991</td>
</tr>
<tr>
<td>PVOT LF</td>
<td>.004 .020 .041</td>
<td>.132 .348 .417</td>
<td>.938 .972 .976</td>
</tr>
<tr>
<td>PVOT ICS-1</td>
<td>.011 .051 .108</td>
<td>.147 .370 .433</td>
<td>.947 .978 .985</td>
</tr>
<tr>
<td>Weak Identification: $\beta_n = .3/\sqrt{n}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>supremum</td>
<td>.051 .139 .224</td>
<td>.764 .922 .957</td>
<td>.992 1.00 1.00</td>
</tr>
<tr>
<td>average</td>
<td>.046 .118 .215</td>
<td>.539 .779 .853</td>
<td>.969 .992 .998</td>
</tr>
<tr>
<td>random</td>
<td>.027 .086 .169</td>
<td>.451 .695 .785</td>
<td>.911 .979 .993</td>
</tr>
<tr>
<td>random LF</td>
<td>.018 .060 .097</td>
<td>.180 .481 .641</td>
<td>.851 .961 .980</td>
</tr>
<tr>
<td>random ICS-1</td>
<td>.018 .058 .098</td>
<td>.298 .633 .770</td>
<td>.926 .975 .991</td>
</tr>
<tr>
<td>PVOT</td>
<td>.051 .122 .201</td>
<td>.740 .894 .934</td>
<td>1.00 1.00 1.00</td>
</tr>
<tr>
<td>PVOT LF</td>
<td>.014 .061 .110</td>
<td>.380 .708 .805</td>
<td>.990 1.00 1.00</td>
</tr>
<tr>
<td>PVOT ICS-1</td>
<td>.015 .060 .111</td>
<td>.618 .848 .878</td>
<td>.999 1.00 1.00</td>
</tr>
<tr>
<td>Non-Identification: $\beta_n = \beta_0 = 0$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>supremum</td>
<td>.061 .152 .223</td>
<td>.751 .922 .956</td>
<td>1.00 1.00 1.00</td>
</tr>
<tr>
<td>average</td>
<td>.054 .145 .200</td>
<td>.526 .765 .849</td>
<td>.975 .996 .999</td>
</tr>
<tr>
<td>random</td>
<td>.036 .123 .184</td>
<td>.417 .696 .803</td>
<td>.025 .976 .988</td>
</tr>
<tr>
<td>random LF</td>
<td>.008 .047 .108</td>
<td>.205 .504 .655</td>
<td>.838 .955 .973</td>
</tr>
<tr>
<td>random ICS-1</td>
<td>.008 .049 .109</td>
<td>.411 .653 .770</td>
<td>.923 .977 .989</td>
</tr>
<tr>
<td>PVOT</td>
<td>.036 .145 .211</td>
<td>.732 .885 .930</td>
<td>1.00 1.00 1.00</td>
</tr>
<tr>
<td>PVOT LF</td>
<td>.010 .058 .114</td>
<td>.373 .717 .806</td>
<td>.990 1.00 1.00</td>
</tr>
<tr>
<td>PVOT ICS-1</td>
<td>.010 .059 .116</td>
<td>.682 .853 .898</td>
<td>1.00 1.00 1.00</td>
</tr>
</tbody>
</table>

Numerical values are rejection frequency at the given level. LSTAR is Logistic STAR. Empirical power is not size-adjusted. supremum and average tests are based on a wild bootstrapped p-value. random: $T_n(\lambda)$ with randomized $\lambda$ on [1,5]. PVOT: p-value occupation time test. PVOT uses the chi-squared distribution, LF is the least favorable p-value, and ICS-1 is the type 1 identification category selection p-value with threshold $\kappa_n = \ln(\ln(n))$. 

40
<table>
<thead>
<tr>
<th></th>
<th>$H_0$: LSTAR</th>
<th>$H_1$-weak</th>
<th>$H_1$-strong</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1% 5% 10%</td>
<td>1% 5% 10%</td>
<td>1% 5% 10%</td>
</tr>
<tr>
<td>Strong Identification: $\beta_n = .3$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>supremum</td>
<td>.029 .069 .153</td>
<td>.441 .590 .676</td>
<td>.997 .999 .999</td>
</tr>
<tr>
<td>average</td>
<td>.022 .055 .120</td>
<td>.382 .546 .624</td>
<td>.988 .996 .997</td>
</tr>
<tr>
<td>random</td>
<td>.008 .049 .098</td>
<td>.328 .488 .598</td>
<td>.976 .999 .996</td>
</tr>
<tr>
<td>random LF</td>
<td>.001 .018 .042</td>
<td>.227 .450 .565</td>
<td>.967 .989 .998</td>
</tr>
<tr>
<td>random ICS-1</td>
<td>.009 .046 .096</td>
<td>.230 .449 .565</td>
<td>.974 .990 .998</td>
</tr>
<tr>
<td>PVOT</td>
<td>.014 .055 .115</td>
<td>.423 .568 .655</td>
<td>.996 .999 .999</td>
</tr>
<tr>
<td>PVOT LF</td>
<td>.002 .023 .051</td>
<td>.311 .509 .618</td>
<td>.995 .998 1.00</td>
</tr>
<tr>
<td>PVOT ICS-1</td>
<td>.013 .058 .106</td>
<td>.314 .510 .618</td>
<td>.995 .998 1.00</td>
</tr>
<tr>
<td>Weak Identification: $\beta_n = .3/\sqrt{n}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>supremum</td>
<td>.044 .134 .184</td>
<td>.984 .998 1.00</td>
<td>1.00 1.00 1.00</td>
</tr>
<tr>
<td>average</td>
<td>.029 .125 .176</td>
<td>.883 .968 /989</td>
<td>1.00 1.00 1.00</td>
</tr>
<tr>
<td>random</td>
<td>.032 .096 .162</td>
<td>.817 .929 .970</td>
<td>.995 .998 .998</td>
</tr>
<tr>
<td>random LF</td>
<td>.009 .051 .108</td>
<td>.519 .835 .914</td>
<td>.984 .996 .998</td>
</tr>
<tr>
<td>random ICS-1</td>
<td>.009 .051 .120</td>
<td>.785 .921 .954</td>
<td>.990 .998 1.00</td>
</tr>
<tr>
<td>PVOT</td>
<td>.050 .118 .194</td>
<td>.981 .995 1.00</td>
<td>1.00 1.00 1.00</td>
</tr>
<tr>
<td>PVOT LF</td>
<td>.012 .053 .109</td>
<td>.823 .965 .975</td>
<td>1.00 1.00 1.00</td>
</tr>
<tr>
<td>PVOT ICS-1</td>
<td>.012 .054 .109</td>
<td>.958 .987 .993</td>
<td>1.00 1.00 1.00</td>
</tr>
<tr>
<td>Non-Identification: $\beta_n = \beta_0 = 0$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>supremum</td>
<td>.051 .151 .196</td>
<td>.981 .998 .998</td>
<td>1.00 1.00 1.00</td>
</tr>
<tr>
<td>average</td>
<td>.043 .136 .189</td>
<td>.886 .968 .984</td>
<td>1.00 1.00 1.00</td>
</tr>
<tr>
<td>random</td>
<td>.047 .111 .177</td>
<td>.826 .938 .967</td>
<td>.997 1.00 1.00</td>
</tr>
<tr>
<td>random LF</td>
<td>.006 .058 .110</td>
<td>.549 .859 .926</td>
<td>1.00 1.00 1.00</td>
</tr>
<tr>
<td>random ICS-1</td>
<td>.006 .058 .109</td>
<td>.827 .940 .973</td>
<td>1.00 1.00 1.00</td>
</tr>
<tr>
<td>PVOT</td>
<td>.061 .148 .208</td>
<td>.977 .993 .996</td>
<td>1.00 1.00 1.00</td>
</tr>
<tr>
<td>PVOT LF</td>
<td>.014 .058 .108</td>
<td>.853 .970 .989</td>
<td>1.00 1.00 1.00</td>
</tr>
<tr>
<td>PVOT ICS-1</td>
<td>.013 .057 .107</td>
<td>.978 .996 .998</td>
<td>1.00 1.00 1.00</td>
</tr>
</tbody>
</table>

Numerical values are rejection frequency at the given level. LSTAR is Logistic STAR. Empirical power is not size-adjusted. *supremum* and *average* tests are based on a wild bootstrapped p-value. *random*: $T_n(\lambda)$ with randomized $\lambda$ on $[1,5]$. *PVOT*: p-value occupation time test. PVOT uses the chi-squared distribution, LF is the least favorable p-value, and ICS-1 is the type 1 identification category selection p-value with threshold $\kappa_n = \ln(\ln(n))$. 