Econometric Theory, **37**, 2021, 409–463.
doi:10.1017/S0266466620000201

WEAK-IDENTIFICATION ROBUST WILD BOOTSTRAP APPLIED TO A CONSISTENT MODEL SPECIFICATION TEST*

JONATHAN B. HILL
University of North Carolina at Chapel Hill

We present a new robust bootstrap method for a test when there is a nuisance parameter under the alternative, and some parameters are possibly weakly or nonidentified. We focus on a Bierens (1990, *Econometrica* 58, 1443–1458)-type conditional moment test of omitted nonlinearity for convenience. Existing methods include the supremum p-value which promotes a conservative test that is generally not consistent, and test statistic transforms like the supremum and average for which bootstrap methods are not valid under weak identification. We propose a new wild bootstrap method for p-value computation by targeting specific identification cases. We then combine bootstrapped p-values across polar identification cases to form an asymptotically valid p-value approximation that is robust to any identification case. Our wild bootstrap procedure does not require knowledge of the covariance structure of the bootstrapped processes, whereas Andrews and Cheng's (2012a, *Econometrica* 80, 2153–2211; 2013, *Journal of Econometrics* 173, 36–56; 2014, *Econometric Theory* 30, 287–333) simulation approach generally does. Our method allows for robust bootstrap critical value computation as well. Our bootstrap method (like conventional ones) does not lead to a consistent p-value approximation for test statistic functions like the supremum and average. Therefore, we smooth over the robust bootstrapped p-value as the basis for several tests which achieve the correct asymptotic level, and are consistent, for any degree of identification. They also achieve uniform size control. A simulation study reveals possibly large empirical size distortions in nonrobust tests when weak or nonidentification arises. One of our smoothed p-value tests, however, dominates all other tests by delivering accurate empirical size and comparatively high power.

1. INTRODUCTION

We present a new bootstrap procedure for nonstandard tests where some regression model parameters may be weakly or nonidentified. We focus ideas at the

*This paper was previously circulated under the title "Inference When There is a Nuisance Parameter under the Alternative and Some Parameters are Possibly Weakly Identified." We thank two referees and Co-Editor Michael Jansson for helpful comments and suggestions. Address correspondence to Jonathan B. Hill, Department of Economics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA; e-mail: jbhill@email.unc.edu.

© The Author(s) 2020. Published by Cambridge University Press.

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

expense of greater generality by working with a regression model that has additive nonlinearity:

$$y_t = \zeta_0' x_t + \beta_0' g(x_t, \pi_0) + \epsilon_t = f(\theta_0, x_t) + \epsilon_t \text{ where } x_t \in \mathbb{R}^{k_x} \text{ and } \theta \equiv \left[\zeta', \beta', \pi'\right]'. \quad \textbf{(1)}$$

The variable y_t is a scalar, $x_t \in \mathbb{R}^{k_x}$ are covariates with a constant term and 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 Π are compact subsets of $\mathbb{R}^{k_\beta}$, $\mathbb{R}^{k_x}$, and $\mathbb{R}^{k_\pi}$, respectively, for finite $(k_\beta, k_\pi) \geq 1$.x_t includes a constant term and at least one stochastic regressor, and let $E[\epsilon_t] = 0$ and $E[\epsilon_t^2] \in (0, \infty)$ for some unique $\theta_0 \in \Theta \equiv \mathcal{Z} \times \mathcal{B} \times \Pi$. We want to test $H_0 : E[y_t|x_t] = f(\theta_0, x_t)$ a.s. for some unique θ_0 against a general alternative $H_1 : \sup_{\theta \in \Theta} P(E[y_t|x_t] = f(\theta, x_t)) < 1$. We assume that the (pseudo) true value θ_0 minimizes a standard criterion function (see, e.g., Kullback and Leibler, 1951; Sawa, 1978; White, 1982). In order to test H_0, we work with the Bierens (1990)-type conditional moment (CM) test of omitted nonlinearity for convenience.

Under H_1, it is known that $E[\epsilon_t F(\lambda' x_t)] \neq 0$ for a large class of weight functions $F : \mathbb{R} \to \mathbb{R}$, and $\forall \lambda \in \Lambda / S_\Lambda$ where Λ is any compact subset of $\mathbb{R}^{k_x}$ and $S_\Lambda \subset \Lambda$ has measure zero (S_Λ depends on F and the underlying distribution). Examples of F include the exponential (Bierens, 1982, 1990; de Jong, 1996), logistic (White, 1989), and the covering class of nonpolynomial real analytic functions (Stinchcombe and White, 1998). A CM test operates on a normalized sample version of $E[\epsilon_t F(\lambda' x_t)]$, cf. Newey (1985) and Tauchen (1985), and therefore has the nuisance parameter λ under H_1. As an example, consider testing whether y_t is governed by a Logistic Smooth Transition Autoregressive (LSTAR) process with a single transition function. Under the alternative, the true process may be LSTAR with multiple transition functions, or Smooth Transition Autoregression (STAR) with a different transition function (e.g., exponential, normal), or may not be in the STAR class at all (e.g., Self Exciting Threshold Autoregression, cf. Tong and Lim, 1980). STAR model estimation generally involves the possibility of weakly or nonidentified parameters that is routinely assumed away in the STAR literature (cf. Terasvirta, 1994). See Example 1 in Section 3 for further details.

If $\beta_0 = 0$, then π_0 is not identified. In fact, if n is the sample size and there is local drift $\beta_0 = \beta_n \to 0$ with $\sqrt{n}||\beta_n|| \to [0, \infty)$, then estimators for θ_0 have nonstandard limit distributions, and estimators of π_0 have a random probability limit. See Andrews and Cheng (2012a, 2013, 2014) for a broad literature review, and for results on estimation and classic inference generally under the assumption of model correctness $E[\epsilon_t|x_t] = 0$ a.s. See also Cheng (2015). We assume throughout that π_0 is identified when $\beta_0 \neq 0$. Otherwise, an approach similar to Cheng (2015) would be appropriate, leading to more intense asymptotics. In the weak identification literature in which a regression model forms the basis of study, correct model specification in the sense that ϵ_t is i.i.d., a martingale difference or $E[\epsilon_t|x_t] = 0$ a.s. is typically assumed. Thus, H_0 above is assumed to be true. Consider, for example, Sargan (1983), Phillips (1990), Choi and Phillips (1992), Stock and Wright (2000), Andrews, Moreira, and Stock (2006), Andrews and Cheng (2012a, Ex. 1, Sect. 6; 2013, Ex. 1 and 2, Sect. 7; 2014, Ex. 2), Cheng

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

(2015), and McCloskey (2017, Ex. 1) amongst others. A broad literature exists on weak identification related to weak instruments (e.g., Dufour, 1997; Stock and Wright, 2000; Moreira, 2003; Andrews, Moreira, and Stock, 2006). This is not our primary focus since the source of weak identification is a specific feature of the regression model.[1]

Conversely, in the omitted nonlinearity test literature, strong identification is universally assumed or implied. This translates here to assuming $\beta_0 \neq 0$, or simply testing whether $y_t = \zeta' x_t + \epsilon_t$ is the correct specification. This literature is equally massive: see, for example, Bierens (1982, 1990), White (1989), Hong and White (1995), Hansen (1996), Bierens and Ploberger (1997), de Jong (1997), Stinchcombe and White (1998), Dette (1999), Li (1999), Whang (2000), Delgado, Dominguez, and Lavergne (2006), Hill (2008), Davidson and Halunga (2014), and Li, Li, and Liu (2016).

Our contributions are twofold. First, we deliver a first-time bridge between these literatures: a consistent model specification test of H_0 that is robust to the full sweep of identification cases (cf. Andrews and Cheng, 2012a). Thus, since we test for correct model specification $E[y_t|x_t] = f(\theta_0, x_t)$ a.s. we do not assume it a priori, contrary to many offerings in the weak identification literature. We must, however, make some assumptions on the model error ϵ_t in order to identify the (possibly pseudo-true) model parameters. In a similar vein, Andrews and Cheng (2014, Ex. 1) only impose a weak orthogonality condition because a regressor may be endogenous, but only treat i.i.d. data in a linear model. We allow for a nonlinear time series setting and only require a weak orthogonality condition (under the alternative).

Second, we provide a bootstrap procedure that is robust to the degree of identification. This topic has apparently been ignored to date. Our method broadly applies to other tests, including t-, Wald, Lagrange Multiplier, and Quasi-Likelihood Ratio (QLR) tests, as well as other model specification tests including nonparametric tests, although we restrict attention to a CM statistic for brevity.

The presence of λ prompts a test statistic transform detailed below. This promotes a nonstandard asymptotic theory and therefore requires a bootstrap method (e.g., Hansen, 1996). But the possibility of weak identification alone leads to nonstandard asymptotics. Thus, even if a nonparametric model specification test is explored which bypasses a nuisance parameter, including Härdle and Mammen (1993), Hong and White (1995), and Zheng (1996), a nonstandard approach is required to handle allowing for any degree of identification. This paper proposes a new bootstrap method suitable for the nuisance parametric approach, that is robust to weak identification. The method is general, and can therefore be extended in principle to any other model specification test approach.

Let $\mathcal{T}_n(\lambda) \geq 0$ be the proposed CM test statistic. In the setting of (1) and mild regularity conditions, $\mathcal{T}_n(\lambda)$ has a chi-squared limit law when $\sqrt{n}||\beta_n|| \to \infty$,

[1] Many treatments in the weak identification literature do not focus on a regression model, but work with unconditional moment conditions, or on a (nonstochastic) parametric function. See Dufour (1997) and Caner (2010). See also Elliott, Muller, and Watson (2015).

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

and otherwise has a nonstandard limit. This represents polar cases of semi-strong or strong identification, and weak or nonidentification (cf. Andrews and Cheng, 2012a). Andrews and Cheng (2012a, 2013, 2014) propose a robust critical value for t, Wald, and QLR statistics, where simulated data are used to approximate the limit distribution.

In the following, unless confusion cannot be avoided, we say "weak identification" to mean non- or weak cases, and "strong identification" to mean semi-strong or strong cases.

The simulation approach of Andrews and Cheng (2012a, 2013, 2014) requires knowledge of the covariance kernel of the simulated stochastic process. This may be intractable when the weakly identified parameter is nonscalar. Furthermore, simulating an asymptotic distribution presumes the latter well approximates the small sample distribution. This may fail to be true when there exists conditional heteroskedasticity, when error tails are leptokurtic, when the parameter dimension is large, and/or when the sample size is small. The typical solution is a bootstrap or subsampling method. In the case of testing H_0, a natural method is the wild (or multiplier) bootstrap as in Hansen (1996), cf. Wu (1986) and Liu (1988). A major advantage of the wild bootstrap premise over the simulation method is that knowledge of the covariance kernel of the bootstrapped process is not required. Bootstrap methods applied to $\mathcal{T}_n(\lambda)$ that do not take into consideration the possibility of weak identification, however, are asymptotically invalid because the weak limit of $\mathcal{T}_n(\lambda)$ exhibits discontinuities with respect to β_n, in which case uniform asymptotics fail. See Bickel and Freedman (1981), Romano (1989), Sheehy and Wellner (1992), and Andrews and Guggenberger (2010). See Gine and Zinn (1990) for discussion on types of uniformity in bootstrap environments.

We solve the problem of nonuniformity by targeting the wild bootstrap to identification category specific first-order expansions of $\mathcal{T}_n(\lambda)$ under the null. Once bootstrapped p-values are computed for polar identification cases, we combine them as in Andrews and Cheng (2012a) using their notions of *Least Favorable* (LF) and *Identification Category Selection* (ICS) constructions (they develop critical values). The result is an asymptotically valid p-value approximation $\hat{p}_n(\lambda)$, irrespective of the degree of identification.

The Bonferroni-based size correction approach of McCloskey (2017) is a plausible alternative to the LF and ICS methods of Andrews and Cheng (2012a). The theory there is presented for a test of a fixed parameter value, where other (model-based nuisance) parameters are also present and cause the test statistic limit distribution discontinuity. In McCloskey's (2017, eqn. (1)) example, the nuisance parameter is part of the data generating process. In our setting, we test whether a chosen model is correct, where a nuisance parameter arises that is not part of the data generating process, and is due solely to the construction of a test statistic. Discontinuity of the limit distribution is not caused by the nuisance parameter, but by a parameter subset from the model. We leave for another venue a consideration of generalizing the Bonferroni-based size correction to our setting.

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

In order to handle the nuisance parameter, we randomize λ, use the classic sup-transform $\sup_{\lambda\in\Lambda}\hat{p}_n(\lambda)$ (see, e.g., Lehmann, 1994, Chap. 3.1), and use the p-Value Occupation Time (PVOT) $\hat{\mathcal{P}}_n(\alpha)\equiv\int_\Lambda I(\hat{p}_n(\lambda)<\alpha)d\lambda$, where $I(A)=1$ if A is true, and $\alpha\in(0,1)$ is the nominal level (Hill, 2018). Randomizing λ sacrifices power (e.g., White, 1989). $\sup_{\lambda\in\Lambda}\hat{p}_n(\lambda)$ by construction promotes a conservative test that is generally not consistent since a Bierens (1990)-type CM test is not known to be consistent for all λ (Bierens, 1990; Stinchcombe and White, 1998). We also present conditions under which our tests achieve uniform size control for any degree of identification.

The challenge of constructing valid tests in the presence of nuisance parameters under H_1 dates at least to Chernoff and Zacks (1964) and Davies (1977, 1987). Nuisance parameters that are not identified under H_1 are either chosen at random (e.g., White, 1989); or $\mathcal{T}_n(\lambda)$ is smoothed over Λ, resulting in a nonstandard limit distribution (e.g., Chernoff and Zacks, 1964; Davies, 1977; Andrews and Ploberger, 1994); or a computed p-value like $\hat{p}_n(\lambda)$ is smoothed. Examples of transforms are the average $\int_\Lambda\mathcal{T}_n(\lambda)\mu(d\lambda)$ and supremum $\sup_{\lambda\in\Lambda}\mathcal{T}_n(\lambda)$, where $\mu(\lambda)$ is a measure on Λ that is absolutely continuous with respect to Lebesgue measure (Chernoff and Zacks, 1964; Davies, 1977; Andrews and Ploberger, 1994). Bierens and Ploberger (1997) integrate the squared numerator from a conventional CM statistic $\mathcal{T}_n(\lambda)$, resulting in the Integrated Conditional Moment (ICM) test, cf. Bierens (1982).

Unless strong identification is assumed, then $\int_\Lambda\mathcal{T}_n(\lambda)\mu(d\lambda)$, $\sup_{\lambda\in\Lambda}\mathcal{T}_n(\lambda)$ and the ICM cannot be consistently bootstrapped by conventional methods or our method. The intuition is simple. We can write $\mathcal{T}_n(\lambda)=\mathcal{Z}_n^2(\lambda)$ for some sample process $\{\mathcal{Z}_n^2(\lambda)\}$, for example, Bierens (1990, eqn. (18)). Under strong identification and fairly general assumptions $\{\mathcal{Z}_n(\lambda):\lambda\in\Lambda\}$ converges to a Gaussian process $\{\mathcal{Z}(\lambda):\lambda\in\Lambda\}$ with covariance kernel $E[\mathcal{Z}(\lambda)\mathcal{Z}(\tilde{\lambda})]$ that generally depends on θ_0 and $E[\epsilon_t^2]$. See Theorem 4.2. In order to use $\mathcal{Z}_n(\lambda)$ to obtain bootstrap draws from the process $\{\mathcal{Z}(\lambda):\lambda\in\Lambda\}$, we therefore need consistent estimators for θ_0 and $E[\epsilon_t^2]$. That is impossible if π_0 is truly only weakly identified (cf. Andrews and Cheng, 2012a) . The same problem applies to Bierens and Ploberger's (1997) ICM test.

Our setting is decidedly different from Hansen's (1996) who tests $\beta_0=0$ and treats π_0 as an unidentified nuisance parameter under the null. We do not require $F(\lambda' x_t)$ to be part of the true data generating process under H_0, and we estimate all parameters θ_0 allowing for weak identification. Moreover, Hansen (1996) delivers a valid bootstrap method for test statistic transforms like $\int_\Lambda\mathcal{T}_n(\lambda)\mu(d\lambda)$ and $\sup_{\lambda\in\Lambda}\mathcal{T}_n(\lambda)$ under strong identification. If any identification category is allowed, then neither his nor our bootstrap methods are valid for such *test statistic* transforms, and we are unaware of any bootstrap method that is valid. The PVOT *p-value* transform, however, does lend itself to weak identification robust inference (cf. Hill, 2018).

We work with p-values due to their convenience of interpretation: one p-value can be used to test H_0 at any desired level of significance, although

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

our bootstrap method can also be used for robust critical value approximations $\hat{c}_{1-\alpha,n}(\lambda)$ (see Hill, 2020, Appendix E). An unavoidable difference in theory, however, is $\hat{c}_{1-\alpha,n}(\lambda)$ leads to an asymptotically correctly *sized* test, while $\hat{p}_n(\lambda)$ only promotes a test with correct asymptotic *level*. [2] Ultimately, this is due to weak identification and the way parameters enter $\hat{p}_n(\lambda)$: see Theorem 6.1 and its proof. In simulation experiments not reported here, however, robust critical and p-values perform essentially identically.

Our tests are consistent irrespective of the choice of Λ, although for a given Λ power in small samples is naturally amplified in certain directions away from the null. These issues are well known and not dealt with in this paper.

A simulation experiment reveals tests based on $\mathcal{T}_n(\lambda^*)$ with randomly selected λ^*, $\sup_{\lambda\in\Lambda}\mathcal{T}_n(\lambda)$, $\int_\Lambda \mathcal{T}_n(\lambda)\mu(d\lambda)$, and the PVOT with a conventional wild bootstrapped p-value $p_n(\lambda)$ are all strongly over-sized under weak-identification. Somewhat ironically, the conservative test based on $\sup_{\lambda\in\Lambda} p_n(\lambda)$ counters the large size distortion under weak identification, but results in low power.[3] The test based on our robust p-value $\hat{p}_n(\lambda^*)$, however, achieves the correct level, but has comparatively low power, while $\sup_{\lambda\in\Lambda}\hat{p}_n(\lambda)$ is conservative with low power. The PVOT test with $\hat{p}_n(\lambda)$ in simulation experiments has the correct size, and under weak identification achieves the highest power.

Our approach is parametric, while in the nonparametric literature weak identification robust methods are increasingly popular. Nevertheless, robust bootstrap procedures have not apparently been treated. See, for example, Andrews and Mikusheva (2016), Cox (2016), Han and McCloskey (2016), and McCloskey (2017) and the references provided there.

The remainder of the paper is organized as follows. Section 2 presents the CM statistic and its transforms. Assumptions and main results are presented in Sections 3 and 4. In Sections 5 and 6, we present robust p-values and develop a method for bootstrapping the robust p-values. A simulation study is contained in Section 7 and concluding remarks follow in Section 8. Proofs are given in Appendix A.

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; $||\cdot||$ 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. $0_{a\times b}$ is an $a \times b$ dimensional matrix of zeros. a.e. denotes *almost everywhere*. $\Rightarrow^*$ denotes weak convergence on l_∞, the space of bounded functions with sup-norm topology, in the sense of Hoffman-Jørgensen (1984, 1991), cf. Dudley (1978) and Pollard (1984, 1990).

[2] Let α be the desired significance level, and let $AsySz$ be the asymptotic size of a test. The asymptotic *level* of the test is α if $AsySz \leq \alpha$.

[3] $\sup_{\lambda\in\Lambda} p_n(\lambda)$ is not robust to identification category: its conservativeness merely tempers the degree of size distortion under weak identification.

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

2. TEST STATISTIC CONSTRUCTION

Let $\{\beta_n\}$ be the drifting sequence such that $\lim_{n\to\infty}\beta_n = \beta_0$. As in Andrews and Cheng (2012a, 2013), technical results are derived under two overlapping cases which align with the following three categories: I.a. $\beta_n = \beta_0 = 0 \forall n \geq 1 (\pi_0$ is *unidentified*); I.b. $n^{1/2}\beta_n \to b \in \mathbb{R}/0$ hence $\beta_0 = 0$ (π_0 is *weakly identified*); II. $n^{1/2}||\beta_n|| \to \infty$ hence $\beta_0 \in \mathbb{R}$ (π_0 is *semi-strongly identified*); and III. $\beta_n \to \beta_0 \neq 0$ (π_0 is *strongly identified*).

The two key over-lapping cases for all asymptotic results are denoted as (see Andrews and Cheng, 2012a, eqn. (2.7)):

$\mathcal{C}(i,b)$. $\beta_n \to \beta_0 = 0$ and $\sqrt{n}\beta_n \to b$ where $b \in (\mathbb{R} \cup \{\pm\infty\})^{k_\beta}$

$\mathcal{C}(ii,\omega_0)$. $\beta_n \to \beta_0$ where $\beta_0 \gtreqless 0$, $\sqrt{n}\,\|\beta_n\| \to \infty$, and $\beta_n/\|\beta_n\| \to \omega_0$ where $\|\omega_0\| = 1$.

Case $\mathcal{C}(i,b)$ contains sequences β_n close to zero, and when $||b|| < \infty$ then π_0 is either weakly or nonidentified. Case $\mathcal{C}(ii,\omega_0)$ contains sequences β_n farther from zero, covering semi-strong ($\beta_0 = 0$ and $\sqrt{n}||\beta_n|| \to \infty$) and strong ($\beta_0 \neq 0$) identification for π_0. Notice b and ω_0 represent two different limits: $\sqrt{n}\beta_n \to b$ versus $\beta_n/||\beta_n|| \to \omega_0$.

Let $F : \mathbb{R} \to \mathbb{R}$ be a real analytic and nonpolynomial function, and $\mathcal{W} : \mathbb{R}^{k_x} \to \mathbb{R}^{k_x}$ is a one-to-one and bounded function. Under H_0, $E[\epsilon_t|x_t] = 0$ a.s., hence $E[\epsilon_t F(\lambda'\mathcal{W}(x_t))] = 0$. Under $H_1, E[\epsilon_t F(\lambda'\mathcal{W}(x_t))] \neq 0 \forall \lambda \in \Lambda/\mathcal{S}_\Lambda$ where $\mathcal{S}_\Lambda$ has Lebesgue measure zero. See Lem. 1 in Bierens (1990) for i.i.d. data and exponential $F(\cdot)$, see Stinchcombe and White (1998) for broad theory treating the analytic class, and see de Jong (1996) and Hill (2008) for the time series case.[4]

We first require an estimation setting on a chosen estimation parameter space Θ. Let y_t exist on the probability measure space $(\Omega, \mathcal{P}, \mathcal{F})$, where $\mathcal{F} \equiv \sigma(\cup_{t\in\mathbb{Z}}\mathcal{F}_t)$ and $\mathcal{F}_t \equiv \sigma(y_\tau : \tau \leq t)$. Assume Θ 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 β, and Π are compact subsets. $\mathcal{Z}(\beta)$ depends on β because parameter restrictions may be imposed to ensure a stationary solution. Define the parameter subset and space

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

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 Θ, it contains $\theta_0 \equiv [\beta_0', \zeta_0', \pi_0']'$, and $0 \in \mathcal{B}^*$. The dependence of $\mathcal{Z}^*(\beta)$ on β ensures Θ^* contains points consistent with stationarity and moment and memory properties imposed under Assumption 1. The spaces $\Theta^* \subset \Theta$ are assumed different to ensure the true value θ_0 does not lie on the boundary of Θ for convenience of focus.

The sample is $\{(y_t, x_t)\}_{t=1}^n$. We work with least squares to reduce notation, but an extension to a broad class of extremum estimators is straightforward.

[4] de Jong (1996) also allows for an infinite dimensional conditioning set, for example, $y_{t-1}, y_{t-2}, \ldots$ in an ARMA model. This leads to more nuanced results for identifying whether a regression model is mis-specified.

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

Define $\epsilon_t(\theta) \equiv y_t - \zeta' x_t - \beta' g(x_t, \pi)$, and define the least-squares criterion and estimator:

$$Q_n(\theta) = Q_n(\psi, \pi) \equiv \frac{1}{2}\frac{1}{n}\sum_{t=1}^{n} \epsilon_t^2(\theta) \text{ and } \hat{\theta}_n \equiv \arg\inf_{\theta \in \Theta} Q_n(\theta).$$

The criterion ensures we can express $\hat{\theta}_n$ as a concentrated estimator $\hat{\theta}_n = [\hat{\psi}_n'(\hat{\pi}_n), \hat{\pi}_n']'$, where

$$\hat{\psi}_n(\pi) = \arg\inf_{\psi \in \Psi} Q_n(\psi, \pi) \text{ and } \hat{\pi}_n = \arg\inf_{\pi \in \Pi} Q_n(\hat{\psi}_n(\pi), \pi).$$

Under weak identification, a suitably normalized $\hat{\theta}_n$ has a nonstandard limit distribution, and must be partitioned into $[\sqrt{n}(\hat{\psi}_n(\hat{\pi}_n) - \psi_n)', \hat{\pi}_n']'$ since $\hat{\pi}_n$ has a stochastic probability limit, cf. Andrews and Cheng (2012a). Thus, we cannot work with Bierens' (1990) original test statistic, nor the environments of White (1989), Bierens and Ploberger (1997), and many others cited in Section 1, since these rely on a first-order expansion of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ in order to characterize a suitable normalizing scale, implicitly ignoring weak identification (see, e.g., Bierens, 1990, p. 1446).

The robust test statistic is constructed as follows. Define

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

$$\widehat{\mathcal{H}}_n = \frac{1}{n}\sum_{t=1}^{n} d_{\theta,t}(\omega(\hat{\beta}_n), \hat{\pi}_n) d_{\theta,t}(\omega(\hat{\beta}_n), \hat{\pi}_n)'$$

$$\text{where } \omega(\beta) \equiv \begin{cases} \beta / \|\beta\| & \text{if } \beta \neq 0 \\ 1_{k_\beta} / \left\|1_{k_\beta}\right\| & \text{if } \beta = 0 \end{cases} \tag{2}$$

$$\hat{\mathfrak{b}}_{\theta,n}(\omega, \pi, \lambda) \equiv \frac{1}{n}\sum_{t=1}^{n} F\left(\lambda' \mathcal{W}(x_t)\right) d_{\theta,t}(\omega, \pi)$$

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

See Andrews and Cheng (2012a, p. 2175) and Andrews and Cheng (2013, p. 40) for discussions on rescaling with $||\beta||$ to avoid a singular Hessian matrix under semi-strong identification, that is, $\beta_0 = 0$ and $\sqrt{n}||\beta_n|| \to \infty$. The CM statistic we use is:

$$\mathcal{T}_n(\lambda) \equiv \left(\frac{1}{\hat{v}_n(\hat{\theta}_n, \lambda)} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \epsilon_t(\hat{\theta}_n) F\left(\lambda' \mathcal{W}(x_t)\right)\right)^2.$$

As discussed in Section 1, transforms like $\sup_{\lambda \in \Lambda} \mathcal{T}_n(\lambda)$, $\int_\Lambda \mathcal{T}_n(\lambda)\mu(d\lambda)$, and the ICM statistic $\int_\Lambda \{1/\sqrt{n}\sum_{t=1}^{n} \epsilon_t(\hat{\theta}_n) \times F(\lambda' \mathcal{W}(x_t))\}^2 \mu(d\lambda)$ cannot be consistently bootstrapped by our or apparently any other bootstrap method. Therefore, we focus

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

on p-value smoothing since we *can* consistently bootstrap a p-value approximation for $\mathcal{T}_n(\lambda)$ for a given λ (see Sections 5 and 6). Let $\hat{p}_n(\lambda)$ be a bootstrapped p-value. Along with $\hat{p}_n(\lambda^*)$ with randomly selected λ^*, and $\sup_{\lambda\in\Lambda}\hat{p}_n(\lambda)$, we use Hill's (2018) PVOT:

$$\hat{\mathcal{P}}_n(\alpha) \equiv \int_\Lambda I\left(\hat{p}_n(\lambda) < \alpha\right) d\lambda \text{ where } \int_\Lambda d\lambda = 1 \text{ is assumed.}$$

If $\int_\Lambda d\lambda \neq 1$, then we use $\int_\Lambda I(\hat{p}_n(\lambda) < \alpha)d\lambda / \int_\Lambda d\lambda$. Hill (2018) shows under general conditions that are verified here that $\lim_{n\to\infty} P(\hat{\mathcal{P}}_n(\alpha) < \alpha) \leq \alpha$ such that the PVOT test has correct asymptotic level.[5] We prove below that the PVOT statistic with the identification robust p-value also achieves uniform size control.

3. ASSUMPTIONS

Recall $\psi_0 = [\beta_0', \zeta_0']' \in \mathbb{R}^{k_\psi}$ where $k_\psi = k_\beta + k_x$, and $\pi_0 \in \mathbb{R}^{k_\pi}$. The following matrix is used to standardize the criterion gradient process below, ensuring a nondegenerate limit under weak identification (e.g., Andrews and Cheng, 2012a, Sect. 3):

$$\mathfrak{B}(\beta) = \begin{bmatrix} I_{k_\psi} & 0_{k_\psi \times k_\pi} \\ 0_{k_\pi \times k_\psi} & \|\beta\| \times I_{k_\pi} \end{bmatrix}. \tag{3}$$

The definition of $\mathfrak{B}(\beta)$ assumes β is a vector; in the scalar case replace $||\beta||$ with β (Andrews and Cheng, 2012a, see the discussion following eqn. (3.11)). Now write $\epsilon_t(\theta) = \epsilon_t(\psi, \pi)$, and define gradient processes:

$$\begin{aligned} \mathcal{G}_{\psi,n}(\theta) &= \sqrt{n}\left\{\frac{\partial}{\partial\psi}Q_n(\theta) - E\left[\frac{\partial}{\partial\psi}Q_n(\theta)\right]\right\} \\ &= -\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\{\epsilon_t(\theta)d_{\psi,t}(\pi) - E\left[\epsilon_t(\theta)d_{\psi,t}(\pi)\right]\right\} \end{aligned} \tag{4}$$

$$\begin{aligned} \mathcal{G}_{\theta,n}(\theta) &= \mathfrak{B}(\beta_n)^{-1}\sqrt{n}\left\{\frac{\partial}{\partial\theta}Q_n(\theta) - E\left[\frac{\partial}{\partial\theta}Q_n(\theta)\right]\right\} \\ &= -\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\{\epsilon_t(\theta)d_{\theta,t}(\omega(\beta),\pi) - E\left[\epsilon_t(\theta)d_{\theta,t}(\omega(\beta),\pi)\right]\right\}. \end{aligned}$$

In order to make $\psi \equiv [\beta', \zeta']'$ explicit, we write interchangeably

$\mathcal{G}_{\psi,n}(\psi,\pi) = \mathcal{G}_{\psi,n}(\theta)$, etc.

[5] Simulation experiments here and in Hill (2018) reveal sharp size for a PVOT tests of functional form, Generalized Autoregressive Conditional Heteroscedasticity (GARCH) effects and a one time structural break, suggesting $\lim_{n\to\infty} P(\hat{\mathcal{P}}_n(\alpha) < \alpha) = \alpha$ likely holds in these and similar cases.

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

Define:

$$\mathfrak{b}_\psi(\pi,\lambda) = E\left[F\left(\lambda'\mathcal{W}(x_t)\right)d_{\psi,t}(\pi)\right] \quad (5)$$

$$\mathfrak{b}_\theta(\omega,\pi,\lambda) \equiv E\left[F\left(\lambda'\mathcal{W}(x_t)\right)d_{\theta,t}(\omega,\pi)\right] \text{ and } \mathfrak{b}_\theta(\lambda) \equiv E\left[F\left(\lambda'\mathcal{W}(x_t)\right)d_{\theta,t}\right]$$

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

$$\text{and } \mathcal{H}_\theta \equiv \mathcal{H}_\theta(\omega_0,\pi_0)$$

$$\mathcal{K}_{\psi,t}(\pi,\lambda) \equiv F\left(\lambda'W(x_t)\right) - \mathfrak{b}_\psi(\pi,\lambda)'\mathcal{H}_\psi^{-1}(\pi)d_{\psi,t}(\pi)$$

$$\mathcal{K}_{\theta,t}(\lambda) \equiv F\left(\lambda'\mathcal{W}(x_t)\right) - \mathfrak{b}_\theta(\lambda)'\mathcal{H}_\theta^{-1}d_{\theta,t}(\beta_n/\|\beta_n\|,\pi_0)$$

$$\text{and } \mathcal{K}_{\theta,t}(\lambda;a,m) \equiv \sum\nolimits_{i=1}^{m}\alpha_i\mathcal{K}_{\theta,t}(\lambda_i).$$

Assumption 1. (Data generating process, test weight).

a. Identification:
   (i) Under H_0, $E[\epsilon_t|x_t] = 0$ a.s. and $E[\epsilon_t^2|x_t] = \sigma_0^2$ a.s., a finite positive constant.
   (ii) Under $\mathcal{C}(i,b)$: $E[(y_t - \zeta_0'x_t)d_{\psi,t}(\pi)] = 0$ for unique $\psi_0 = [0_{k_\beta}', \zeta_0']'$ in the interior of Ψ^*. Under $\mathcal{C}(ii,\omega_0)$: $E[\epsilon_t(\theta_0) \times d_{\theta,t}(\omega_0,\pi_0)] = 0$ for unique $\theta_0 = [\beta_0', \zeta_0', \pi_0']'$ in the interior of $\Theta^* = \Psi^* \times \Pi^*$.
b. Memory and Moments: $\{\epsilon_t, x_t\}$ are L_p-bounded for some $p > 6$, strictly stationary, and β-mixing with mixing coefficients $\beta_l = O(l^{-qp(q-p)-\iota})$ for some $q > p$ and tiny $\iota > 0$.
c. Response $g(x,\pi)$ and Test Weight $F(\lambda'\mathcal{W}(x))$:
   (i) $g(\cdot,\pi)$ is Borel measurable for each π; $g(\cdot,\pi)$ is twice continuously differentiable in $\pi \in \mathbb{R}^{k_\pi}$; $g(x_t,\pi)$ is a nondegenerate random variable for each $\pi \in \Pi$.
   (ii) $F : \mathbb{R} \to \mathbb{R}$ is analytic, nonpolynomial, and $\mathcal{W}$ is one-to-one and bounded.
   (iii) $E[\sup_{\pi\in\Pi}|(\partial/\partial\pi)^i g(x_t,\pi)|^6] < \infty$ and $E[\sup_{\lambda\in\Lambda}|(\partial/\partial\lambda)^j F(\lambda'\mathcal{W}(x_t))|^6] < \infty$ for $i = 0,1,2$ and $j = 0,1$.

Remark 1. Condition (a) imposes $E[\epsilon_t|x_t] = 0$ a.s. under the null. It otherwise requires $\zeta_0'x_t + \beta_0'g(x_t,\pi_0)$ to be a pseudo true representation for y_t in the sense of being the minimum mean squared error predictor. Notice under weak identification $\mathcal{C}(i,b)$, we only need to consider $\zeta_0'x_t$ as a best predictor in some sense. The only place where $E[\epsilon_t^2|x_t] = \sigma_0^2$ a.s. is used is to simplify the construction of critical values and p-values in practice. In principle, the assumption can be replaced with $E[\epsilon_t^2|x_t] = \sigma_t^2(\cdot)$, a known parametric function. a(ii) identifies parameter values under either hypothesis.

Remark 2. β-mixing under (b) allows us to exploit a probability inequality due to Eberlein (1984) and arguments in Arcones and Yu (1994) in order to prove a stochastic equicontinuity condition.

Remark 3. (c.i) ensures measurability. By assuming $g(x_t,\pi)$ is a nondegenerate random variable for each $\pi \in \Pi$, it also focuses the identification issue to π

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

alone (cf. Andrews and Cheng, 2012a, STAR(iv)).[6] The envelope bounds in (c.iii) are used to prove consistency of criterion derivatives and a sample variance. If $g(x,\pi) = xh(x,\pi)$ and h and F are exponential, logistic, normal, or trigonometric then we need only assume $E|x_t|^6 < \infty$.[7] See, for example, the proofs of Lemmas A.2 and A.5.

Remark 4. We require additional technical details on the existence of certain long-run variances, and the true and estimation parameter spaces. Since these are lengthy and standard, we place them in Assumptions 1.d–f in Appendix A.1.

Remark 5. It is understood that the true parameter space Θ^* contains points consistent with the moment and memory properties of (a)–(e).

Example 1. (STAR Model). Model (1) contains the Smooth Transition Autoregression class where $x_t = [1, y_{t-1}, \ldots, y_{t-p}]'$, $g(x_t,\pi) = x_t h(z_t,\pi)$, $z_t = y_{t-d}$ for some $1 \leq d \leq p$, and h is a scalar function (typically exponential, logistic, or the normal distribution function: see Chan and Tong, 1986; Luukkonen, Saikkonen, and Terasvirta, 1988; Granger and Terasvirta 1993; Terasvirta, 1994; Hill, 2008).

We work with a STAR(p) model with one logistic transition function:

$$y_t = \zeta_0' x_t + \beta_0' x_t \frac{1}{1+\exp\{-\pi_0 y_{t-1}\}} + \epsilon_t = f(\theta_0, x_t) + \epsilon_t \text{ where } \pi_0 > 0. \tag{6}$$

In the STAR literature, both $E[\epsilon_t|x_t] = 0$ a.s. and $\beta_0 \neq 0$ are simply assumed (e.g., Luukkonen et al., 1988; Terasvirta, 1994; Hill, 2008; Andrews and Cheng, 2013). However, neither $E[\epsilon_t|x_t] = 0$ a.s. nor $\beta_0 \neq 0$ may be true in practice. If $\beta_0 = 0$ then π_0 is not identified: this occurs when a STAR model is estimated, but the true data generating process is a linear autoregression.[8]

Furthermore, if $E[\epsilon_t|x_t] = 0$ a.s. is false, then by a(ii), we still require $f(\theta_0, x_t)$ to be pseudo-true in the sense that ϵ_t satisfies the orthogonality conditions $E[\epsilon_t x_t] = 0$ and $E[\epsilon_t x_t (1+\exp\{-\pi_0 y_{t-1}\})^{-1}] = 0$, and additionally under $\mathcal{C}(ii,\omega_0)$:

$$E\left[\epsilon_t \beta_0' x_t \frac{y_{t-1}\exp\{-\pi_0 y_{t-1}\}}{(1+\exp\{-\pi_0 y_{t-1}\})^2}\right] = 0. \tag{7}$$

Under these conditions, $E[(y_t - f(\theta, x_t))^2]$ is minimized at some unique $\theta_0 = [\zeta_0', \beta_0', \pi_0']' \in \Theta$ and compact Θ. In this paper, under $H_1 : \sup_{\theta\in\Theta} P(E[y_t|x_t] = f(\theta, x_t)) < 1$, we are agnostic about what the true data generating

[6] Consider a Logistic STAR model $y_t = \zeta_1 + \zeta_2 y_{t-1} + \beta(1+\exp\{\pi_1(y_{t-d} - \pi_2)\})^{-1} + \epsilon_t$ with two covariates $x_t = [1, y_{t-1}]'$, $\pi_1 > 0$ (see, e.g., Terasvirta, 1994), and assume $E[\epsilon_t^2] > 0$. If $\Pi^* = \{\pi \in \mathbb{R}^2 : \pi_1 \geq \varepsilon\}$ for some $\varepsilon > 0$ then $g(x_t, \pi)$ is a nondegenerate random variable on Π. Otherwise $g(x,\pi) = 1/2 \forall x$ at $\pi_1 = 0$, the model reduces to a linear structure, and β cannot be identified. We rule out such cases in order to focus the identification issue to just π.

[7] $F(u) = 1/(1-u)$ on $(-\infty, 1)$ is analytic and nonpolynomial. However, $F(\lambda' \mathcal{W}(x_t))$ need not satisfy the envelope bound depending on $\mathcal{W}(\cdot)$ and properties of x_t.

[8] Obviously a pretest for omitted STAR effects in a linear model may be performed (e.g., Hill, 2008), but rejecting the linear AR null hypothesis may be an error, and estimating (6) can have an unidentified parameter.

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

process is. Examples are an LSTAR with *two* transition functions:

$$y_t = \zeta_0' x_t + \beta_{1,0}' x_t \frac{1}{1+\exp\{-\pi_{1,0} y_{t-1}\}} + \beta_{2,0}' x_t \frac{1}{1+\exp\{-\pi_{2,0} y_{t-1}\}} + \epsilon_t;$$

or ESTAR $y_t = \zeta_0' x_t + \beta_0' x_t \exp\{-\pi_0 y_{t-1}^2\} + \epsilon_t$; or y_t may be governed by some other class of processes, like a Self-Exciting Threshold Autoregression $y_t = \zeta_0' x_t + \beta_0' x_t I(y_{t-1} > c) + \epsilon_t$, etc.

Next, we discuss the possible limit process for $\hat{\pi}_n$ under weak identification. Define the value of ψ for the nonidentification case:

$$\psi_{0,n} \equiv \left[0_{k_\beta}', \zeta_0'\right]'.$$

By Lemma A.1, $\mathcal{G}_{\psi,n}(\psi_{0,n},\pi) = -1/\sqrt{n}\sum_{t=1}^n \{\epsilon_t(\psi_{0,n}) d_{\psi,t}(\pi) - E[\epsilon_t(\psi_{0,n}) d_{\psi,t}(\pi)]\}$ satisfies $\{\mathcal{G}_{\psi,n}(\psi_{0,n},\pi) : \pi \in \Pi\} \Rightarrow^* \{\mathcal{G}_\psi(\pi) : \pi \in \Pi\}$, a zero-mean Gaussian process. Define:

$$\mathcal{D}_\psi(\pi) \equiv -\frac{\partial}{\partial \beta_0'} E\left[\epsilon_t(\theta) d_{\psi,t}(\pi)\right] = -E\left[d_{\psi,t}(\pi) g(x_t,\pi_0)'\right], \tag{8}$$

and processes $\{\xi_\psi(\pi,\cdot) : \pi \in \Pi\}$ and $\{\tau_\beta(\pi,\cdot) : \pi \in \Pi\}$:

$$\xi_\psi(\pi,b) \equiv -\frac{1}{2}\left\{\mathcal{G}_\psi(\pi) + \mathcal{D}_\psi(\pi) b\right\}' \mathcal{H}_\psi^{-1}(\pi) \left\{\mathcal{G}_\psi(\pi) + \mathcal{D}_\psi(\pi) b\right\} \tag{9}$$

$$\tau_\beta(\pi,b) \equiv -\mathcal{S}_\beta \mathcal{H}_\psi^{-1}(\pi) \left\{\mathcal{G}_\psi(\pi) + \mathcal{D}_\psi(\pi) b\right\} \text{ where } \mathcal{S}_\beta \equiv \left[I_{k_\beta} : 0_{k_x \times k_x}\right].$$

Notice $\mathcal{S}_\beta$ is the β selection matrix. $\mathcal{H}_\psi(\pi)$ is positive definite uniformly on Π by Assumption 1.d(iii). The following identifies the limit process of $\pi^*(b)$ of $\hat{\pi}_n$ under weak identification, and ensures $\tau_\beta(\pi^*(b),b)$ is not degenerate (cf. Andrews and Cheng, (2012a): Assumption C6 and Andrews and Cheng, (2013): STAR3). See Andrews and Cheng (2013, Assumption C6′) for primitive conditions that ensure the next assumption when β is a scalar.

Assumption 2. (Identification of π). 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 $\mathfrak{A}(b)$ with $P(\mathfrak{A}(b)) = 1$ is minimized over Π at a unique point $\pi^*(b)$ that may depend on the sample path. (*b*) $P(\tau_\beta(\pi^*(b),b) = 0) = 0$.

The test statistic scale $\hat{v}_n^2(\hat{\theta}_n,\lambda)$ can have a degenerate probability limit for some λ under strong identification cases, depending on the variation of x_t (see Bierens, 1990, p. 1447). Indeed, because we include a constant term, at $\lambda = 0$ with least squares $1/\sqrt{n}\sum_{t=1}^n \epsilon_t(\hat{\theta}_n) F(0'\mathcal{W}(x_t)) = 0$, hence $\hat{v}_n^2(\hat{\theta}_n,0) \xrightarrow{p} 0$. The following rules this out on Λ-a.e. by effectively ensuring the parameters (α,β,ζ,π) in the augmented model $y_t = \zeta' x_t + \beta' g(x_t,\pi) + \alpha\mu(x_t) + u_t$ can be locally identified for some Borel measurable $\mu : \mathbb{R}^{k_x} \to \mathbb{R}$. See Bierens (1990, p. 1447) for discussion, and see Lemma A.11 in Appendix B. Further details in the context of a STAR model are provided online in the Supplementary Material (Appendix F).

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

Assumption 3. (Nondegenerate scale on Λ-a.e.)

a. Let $\mathcal{C}(i,b)$ with $||b|| < \infty$ hold. Then $P(E[\inf_{\pi\in\Pi}\{\epsilon_t^2(\psi_0,\pi)\}|x_t] > 0) = 1$. 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 $\{\omega \in \mathbb{R}^{k_\beta} : \omega'\omega = 1\} \times \Pi$.
b. Let $\mathcal{C}(ii,\omega_0)$ hold. 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 $E[\kappa_t\kappa_t']$.

Remark 6. By Lem. 2 in Bierens (1990), $E[\epsilon_t^2|x_t] = \sigma_0^2 > 0$ a.s. under Assumption 1.a(i), and Assumption 3.b, imply under strong identification $E[\epsilon_t^2\{F(\lambda'\mathcal{W}(x_t)) - \mathfrak{b}_\theta(\lambda)'\mathcal{H}_\theta^{-1}d_{\theta,t}\}^2] > 0$ on Λ-a.e. since $F(\lambda'\mathcal{W}(x_t)) - \mathfrak{b}_\theta(\lambda)'\mathcal{H}_\theta^{-1}d_{\theta,t} \neq 0$ a.s. on Λ-a.e. In Lemma A.11 in Appendix B, we prove Assumption 3.a plays a similar role under weak identification. This suffices to ensure $\hat{v}_n^2(\hat{\theta}_n,\lambda) > 0$ asymptotically with probability approaching one, on Λ-a.e.

Unfortunately, Assumption 3 does not rule out $\hat{v}_n^2(\hat{\theta}_n,\lambda) \xrightarrow{p} 0$ *everywhere* on Λ. In order to avoid deviant cases, we make the following assumption which is mild in view of Lemma B.12 (see also Bierens, 1990, p. 1449). We discuss the requirement in the Supplementary Material (Appendix F).

Define the augmented parameter set $\theta^+ \equiv [||\beta||, \omega(\beta)', \zeta', \pi']'$ which is useful under weak identification cases when β is a vector. Let $\theta^+ \in \Theta^+ \equiv \{\theta^+ \in \mathbb{R}^{k_\beta+k_x+k_\pi+1} : \theta^+ = [||\beta||, \omega(\beta), \zeta, \pi]' : \beta \in \mathcal{B}, \zeta \in \mathcal{Z}(\beta), \pi \in \Pi\}$. Define $\epsilon_t(\theta^+) \equiv y_t - \zeta' x_t - ||\beta||\omega(\beta)'g(x_t,\pi)$ and

$$v^2(\theta^+,\lambda) = E\left[\epsilon_t^2(\theta^+)\left\{F\left(\lambda'\mathcal{W}(x_t)\right) - \mathfrak{b}_\theta(\omega,\pi,\lambda)'\mathcal{H}_\theta^{-1}(\omega,\pi)d_{\theta,t}(\omega,\pi)\right\}^2\right]$$

$$\hat{v}_n^2(\theta^+,\lambda) = \frac{1}{n}\sum_{t=1}^n \epsilon_t^2(\theta^+)\left\{F\left(\lambda'\mathcal{W}(x_t)\right) - \hat{\mathfrak{b}}_{\theta,n}(\omega,\pi,\lambda)'\widehat{\mathcal{H}}_n^{-1}d_{\theta,t}(\omega,\pi)\right\}^2.$$

Since $\sup_{\theta^+\in\Theta^+,\lambda\in\Lambda} ||\hat{v}_n^2(\theta^+,\lambda) - v^2(\theta^+,\lambda)|| \xrightarrow{p} 0$ by Lemma B.11, it suffices to bound $v^2(\theta^+,\lambda)$. In the scalar case $\beta \in \mathcal{B} \subset \mathbb{R}$ we use $\epsilon_t(\theta) \equiv y_t - \zeta' x_t - \beta' g(x_t,\pi)$ and

$$v^2(\theta,\lambda) = E\left[\epsilon_t^2(\theta)\left\{F\left(\lambda'\mathcal{W}(x_t)\right) - \mathfrak{b}_\theta(\omega(\beta),\pi,\lambda)'\mathcal{H}_\theta^{-1}(\omega(\beta),\pi)d_{\theta,t}(\omega(\beta),\pi)\right\}^2\right]$$

$$\hat{v}_n^2(\theta,\lambda) = \frac{1}{n}\sum_{t=1}^n \epsilon_t^2(\theta)\left\{F\left(\lambda'\mathcal{W}(x_t)\right) - \hat{\mathfrak{b}}_{\theta,n}(\omega(\beta),\pi,\lambda)'\widehat{\mathcal{H}}_n^{-1}d_{\theta,t}(\omega(\beta),\pi)\right\}^2.$$

Since $\sup_{\theta^+\in\Theta^+,\lambda\in\Lambda} ||\hat{v}_n^2(\theta^+,\lambda) - v^2(\theta^+,\lambda)|| \xrightarrow{p} 0$ and $\sup_{\theta\in\Theta^+,\lambda\in\Lambda} ||\hat{v}_n^2(\theta,\lambda) - v^2(\theta,\lambda)|| \xrightarrow{p} 0$ by Lemma A.10, it suffices to bound $v^2(\theta^+,\lambda)$ and $v^2(\theta,\lambda)$.

See Hill (2020, Appendix F) for discussion of the following assumption.

Assumption 4. (Nondegnerate scale).

a. Let β be a scalar. Let $\inf_{\pi\in\Pi} v^2((\beta_0,\zeta_0,\pi),\lambda) > 0 \forall \lambda \in \Lambda$ under identification case $\mathcal{C}(i,b)$ with $|b| < \infty$, and under $\mathcal{C}(ii,\omega_0)$ let $v^2(\theta_0,\lambda) > 0 \forall \lambda \in \Lambda$.

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

b. Let β be a vector. Let $\inf_{\omega\in\mathbb{R}^{k_\beta}:\omega'\omega=1,\pi\in\Pi} v^2((||\beta_0||,\omega,\zeta_0,\pi),\lambda) > 0 \forall \lambda \in \Lambda$ under identification case $\mathcal{C}(i,b)$ with $||b|| < \infty$, and under $\mathcal{C}(ii,\omega_0)$ let $v^2(\theta_0^+,\lambda) > 0 \forall \lambda \in \Lambda$.

4. TEST STATISTIC LIMIT THEORY

We begin by deriving the weak limit of $\hat{\theta}_n$, since it strongly influences the limit properties of $\mathcal{T}_n(\lambda)$. Recall $\hat{\theta}_n = [\hat{\psi}_n(\hat{\pi}_n)', \hat{\pi}_n']'$ where $\hat{\psi}_n(\pi) = \arg\inf_{\psi\in\Psi} Q_n(\psi,\pi)$ and $\hat{\pi}_n = \arg\inf_{\pi\in\Pi} Q_n(\hat{\psi}_n(\pi),\pi)$.

The limit process of a suitably normalized $\hat{\theta}_n$ requires the following constructions. Recall $\psi_{0,n} \equiv [0_{k_\beta}', \zeta_0']$. We need two Gaussian processes $(\mathcal{G}_\psi(\pi), \mathcal{G}_\theta)$ defined as the weak limits of $\mathcal{G}_{\psi,n}(\psi_{0,n},\pi)$ and $\sqrt{n}\mathfrak{B}(\beta_n)^{-1}(\partial/\partial\theta)Q_n(\theta_n)$, cf. Lemma B.1 and Corollary B.4 in Appendix B. Define

$$\tau(\pi,b) \equiv -\mathcal{H}_\psi^{-1}(\pi)\left\{\mathcal{G}_\psi(\psi_{0,n},\pi) + \mathcal{D}_\psi(\pi)b\right\} - \left[b', 0_{k_\beta}'\right]'.$$

Denote the true value $\psi_n \equiv [\beta_n', \zeta_0']'$ under drifting sequence $\{\beta_n\}$.

The following allows for model mis-specification in the sense of Assumption 1.a(ii). This effectively generalizes the low level extensions of Andrews and Cheng (2012a, Thm. 3.1 and 3.2) developed in Andrews and Cheng (2012a, Ex. 2) for the Autoregressive Moving Average (ARMA) model with an i.i.d. error, and Andrews and Cheng (2013, Sect. 6) for the STAR model with an mds error. The proof is similar to arguments in Andrews and Cheng (2012a): see Hill (2020, Appendix C).

THEOREM 4.1. *Let Assumptions 1 and 2 hold. a. Under drift case* $\mathcal{C}(i,b)$ *with* $||b|| < \infty$, $(\sqrt{n}(\hat{\psi}_n(\hat{\pi}_n) - \psi_n), \hat{\pi}_n) \xrightarrow{d} (\tau(\pi^*(b),b), \pi^*(b))$. *b. Under drift case* $\mathcal{C}(ii,\omega_0)$, $\sqrt{n}\mathfrak{B}(\hat{\beta}_n)(\hat{\theta}_n - \theta_n) \xrightarrow{d} -\mathcal{H}_\theta^{-1}\mathcal{G}_\theta$.

Remark 7. Under any degree of (non)identification, $\hat{\sigma}_n^2 \equiv 1/n\sum_{t=1}^n (y_t - f(\hat{\theta}_n, x_t))^2 \xrightarrow{p} \sigma_0^2$ is easily verified. The only issue is case $\mathcal{C}(i,b)$, but $\beta_n \to 0$ ensures the nonstandard asymptotic properties of $\hat{\pi}_n$ are irrelevant asymptotically.

Now turn to the test statistic $\mathcal{T}_n(\lambda) = \{1/\sqrt{n}\sum_{t=1}^n \epsilon_t(\hat{\theta}_n)F(\lambda'\mathcal{W}(x_t))/\hat{v}_n(\hat{\theta}_n,\lambda)\}^2$. The required limit processes under the null are constructed as follows. First, consider the weak identification case $\mathcal{C}(i,b)$ with $||b|| < \infty$. 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, we have weak convergence:

$$\left\{\frac{1}{\sqrt{n}}\sum_{t=1}^n \epsilon_t\left(F\left(\lambda'\mathcal{W}(x_t)\right) - \mathfrak{b}_\psi(\pi,\lambda)'\mathcal{H}_\psi^{-1}(\pi)d_{\psi,t}(\pi)\right) : \Pi,\Lambda\right\}$$
$$\Rightarrow^* \left\{\mathfrak{Z}_\psi(\pi,\lambda) : \Pi,\Lambda\right\},$$

where $\{\mathfrak{Z}_\psi(\pi,\lambda) : \pi \in \Pi, \lambda \in \Lambda\}$ is a zero-mean Gaussian process with covariance kernel $\sigma_0^2 E[\mathcal{K}_{\psi,t}(\pi,\lambda) \times \mathcal{K}_{\psi,t}(\tilde{\pi},\tilde{\lambda})]$. The numerator of the test statistic,

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

$(1/\sqrt{n}\sum_{t=1}^{n}\epsilon_t(\hat{\theta}_n)F(\lambda'\mathcal{W}(x_t)))^2$, therefore converges under H_0 to $\mathfrak{T}^2_\psi(\pi^*(b),\lambda,b)$, where

$$\begin{aligned}\mathfrak{T}_\psi(\pi,\lambda,b) \equiv \mathfrak{Z}_\psi(\pi,\lambda) + \mathfrak{b}_\psi(\pi,\lambda)'\left\{\mathcal{H}_\psi^{-1}(\pi)\mathcal{D}_\psi(\pi)b + \left[b,0'_{k_\beta}\right]'\right\} \\ + \mathfrak{b}_\psi(\pi,\lambda)'\mathcal{H}_\psi^{-1}(\pi)E\left[d_{\psi,t}(\pi)\{g(x_t,\pi_0)-g(x_t,\pi)\}'\right]b \\ + E\left[\mathcal{K}_{\psi,t}(\pi,\lambda)\{g(x_t,\pi_0)-g(x_t,\pi)\}'\right]b.\end{aligned} \tag{10}$$

The form reflects that (*i*) we only expand around the (possibly drifting) true ψ_n because $\hat{\pi}_n$ has a nonstandard limit law, hence $\mathfrak{Z}_\psi(\pi,\lambda)$; (*ii*) weak identification with $b\neq 0$ adds asymptotic bias in $\sqrt{n}(\hat{\psi}_n(\hat{\pi}_n)-\psi_n)$ in the first-order expansion, hence $\mathfrak{b}_\psi(\pi,\lambda)'\{\mathcal{H}_\psi^{-1}(\pi)\mathcal{D}_\psi(\pi)b+[b,0'_{k_\beta}]'\}$; and (*iii*) bias subsequently arises through $\epsilon_t(\psi_n,\hat{\pi}_n)$ which does not have a zero mean in general, hence the remaining two terms. See the proof of Theorem 4.2 for details.

The scale $\hat{v}_n^2(\hat{\theta}_n,\lambda)$ limit under weak identification is constructed from:

$$\begin{aligned}v^2(\omega,\pi,\lambda) \equiv E\Big[\epsilon_t^2(\psi_0,\pi)\big\{F\left(\lambda'\mathcal{W}(x_t)\right) \\ -\mathfrak{b}_\theta(\omega,\pi,\lambda)'\mathcal{H}_\theta^{-1}(\omega,\pi)d_{\theta,t}(\omega,\pi)\big\}^2\Big]\end{aligned} \tag{11}$$

$$\bar{v}^2(\pi,\lambda,b) \equiv v^2(\omega^*(\pi,b),\pi,\lambda) \text{ where } \omega^*(\pi,b) \equiv \tau_\beta(\pi,b)/\left\|\tau_\beta(\pi,b)\right\|.$$

The null limit process of the test statistic under weak identification is therefore:

$$\mathcal{T}_\psi(\pi,\lambda,b) \equiv \frac{\mathfrak{T}^2_\psi(\pi,\lambda,b)}{\bar{v}^2(\pi,\lambda,b)} \text{ and } \mathcal{T}_\psi(\lambda,b) \equiv \mathcal{T}_\psi(\pi^*(b),\lambda,b).$$

Now consider strong identification $\mathcal{C}(ii,\omega_0)$. By Lemma B.9.b, we have the weak limit under the null:

$$\left\{\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\epsilon_t\left(F\left(\lambda'\mathcal{W}(x_t)\right)-\mathfrak{b}_\theta(\lambda)'\mathcal{H}_\theta^{-1}d_{\theta,t}\right):\lambda\in\Lambda\right\} \Rightarrow^* \{\mathfrak{Z}_\theta(\lambda):\lambda\in\Lambda\},$$

where $\{\mathfrak{Z}_\theta(\lambda):\lambda\in\Lambda\}$ is a zero-mean Gaussian process with variance $v^2(\lambda)\equiv E[\epsilon_t^2\{F\left(\lambda'\mathcal{W}(x_t)\right)-\mathfrak{b}_\theta(\lambda)'\mathcal{H}_\theta^{-1}d_{\theta,t}\}^2]$. The limit process $\mathcal{T}(\lambda)\equiv\mathfrak{Z}^2_\theta(\lambda)/v^2(\lambda)$ is therefore chi-squared with one degree of freedom, as in Bierens (1990), cf. de Jong (1996) and Hill (2008).

THEOREM 4.2. *Let Assumptions 1, 2 and 4, and H_0, hold. a. If drift case $\mathcal{C}(i,b)$ holds with $||b||<\infty$, then $\{\mathcal{T}_n(\lambda):\lambda\in\Lambda\}\Rightarrow^*\{\mathcal{T}_\psi(\lambda,b):\lambda\in\Lambda\}$. Furthermore, $\inf_{\pi\in\Pi}\bar{v}^2(\pi,\lambda,b)>0\forall\lambda\in\Lambda$, $\mathcal{T}_\psi(\lambda,b)\geq 0$ a.s., and $\sup_{\lambda\in\Lambda}\{\mathcal{T}_\psi(\lambda,b)\}<\infty$ a.s. b. If drift case $\mathcal{C}(ii,\omega_0)$ holds then $\{\mathcal{T}_n(\lambda):\lambda\in\Lambda\}\Rightarrow^*\{\mathcal{T}(\lambda):\lambda\in\Lambda\}$, a chi-squared process with one degree of freedom, with a version that has almost surely uniformly continuous sample paths. Furthermore, $v^2(\lambda)>0\forall\lambda\in\Lambda$, $\mathcal{T}(\lambda)\geq 0$ a.s., $\sup_{\lambda\in\Lambda}\{\mathcal{T}(\lambda)\}<\infty$ a.s., and $\mathcal{T}(\lambda)$ has an absolutely continuous distribution for each λ.*

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

The CM test is consistent on Λ-a.e. under any identification category. Recall the alternative is nonlocal to null: $H_1 : \sup_{\theta\in\Theta} P(E[y_t|x_t] = f(\theta, x_t)) < 1$.

THEOREM 4.3. *Let Assumptions 1, 2, and 4 hold, and let drift case $\mathcal{C}(i,b)$ with $||b|| < \infty$, or $\mathcal{C}(ii,\omega_0)$, apply. Under H_1, $\mathcal{T}_n(\lambda) \xrightarrow{p} \infty$ for all $\lambda \in \Lambda/S$ where $S \subset \Lambda$ has Lebesgue measure zero.*

5. ROBUST *p*-VALUE CONSTRUCTIONS

We now develop identification category robust p-values, while their bootstrapped approximations are handled in Section 6. Critical value computation is presented in Hill (2020, Appendix E).

Recall the Theorem 4.2.a null limit process $\{\mathcal{T}_\psi(\lambda,b) : \lambda \in \Lambda\}$ of $\mathcal{T}_n(\lambda)$ under weak identification $\sqrt{n}\beta_n \to b$ with $||b|| < \infty$. Note $\theta = [\zeta', \beta', \pi']'$ may not fully parameterize the distribution of $W_t \equiv [y_t, x_t']'$, hence $\mathcal{T}_\psi(\lambda,b)$ may not reveal *all* distribution based nuisance parameters. Let ϕ_0 index all remaining (nuisance) parameters such that the distribution of W_t is determined by $\gamma_0 \equiv (\theta_0, \phi_0)$ where (see Andrews and Cheng, 2012a, p. 2161–2162):

$$\gamma_0 \equiv (\theta_0, \phi_0) \in \Gamma^* \equiv \left\{\theta \in \Theta^*, \phi \in \Phi^*(\theta)\right\}. \tag{12}$$

In our nonlinear regression setting (1), ϕ_0 is a possibly infinite dimensional parameter that indexes all remaining characteristics of the error distribution not represented in (1).

Assume $\Phi^*(\theta) \subset \Phi^* \forall \theta \in \Theta^*$, where Φ^* is a compact metric space with a metric that induces weak convergence for $\{W_t, W_{t+m}\}$ with respect to drifting $\gamma \to \gamma_0$ (Andrews and Cheng, 2012a, eq. (2.3)).[9] The space $\Phi^*(\theta)$ generally depends on $\theta \in \Theta^*$ because we implicitly assume Γ^* indexes only those distributions that satisfy the maintained assumptions, which ensures uniform asymptotics. [10] Under drift $\theta_n = [\zeta_0', \beta_n', \pi_0']'$, the parameter set becomes $\gamma_n \equiv (\theta_n, \phi_0) \to \gamma_0$. We only let β exhibit drift to ease notation, and since that parameter governs identification cases for π.

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

$$h \equiv (\gamma_0, b) \in \mathfrak{H} \equiv \left\{h : \gamma_0 \in \Gamma^*, \text{ and } \|b\| < \infty, \text{ with } \beta_0 = 0\right\}. \tag{13}$$

[9] Let $d_\phi(\cdot,\cdot)$ be the metric on Φ^*. Under these assumptions Γ^* is a metric space with metric $d_\Gamma(\gamma_1,\gamma_2) \equiv ||\theta_1 - \theta_2|| + d_\phi(\phi_1,\phi_1)$. Now let P_γ denote the joint probability function of (W_t, W_{t+m}) induced by the measure $\mathcal{P}$ under γ. The metric $d_\Gamma(\gamma_1,\gamma_2)$ therefore satisfies the following weak convergence: if $\gamma \to \gamma_0$ then $P_\gamma(W_t \le a, W_{t+m} \le b) \to P_{\gamma_0}(W_t \le a, W_{t+m} \le b) \forall a,b \in \mathbb{R}^{p+1}$, $\forall t$, $\forall m \ge 1$. A key implicit use of these ideas arises under local drift $\gamma_n \to \gamma_0$, in which case only β matters under our assumptions, hence $d_\Gamma(\gamma_n,\gamma_0) \equiv ||\beta_n - \beta_0||$. In the sequel, we therefore do not explicitly state such weak convergence in order to avoid redundancy. See also Andrews and Cheng (2012a, footnote 21).

[10] See Andrews and Cheng (2012b, p. 119) for an example.

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

5.1. *p*-Values for $\mathcal{T}_n(\lambda)$

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

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

The following summarizes and extends ideas developed in Andrews and Cheng, 2012a, Section. 5). TheLF p-value is defined as $p_n^{(LF)}(\lambda) \equiv \max\{\sup_{h\in\mathfrak{H}}\{p_n(\lambda,h)\}, p_n^\infty(\lambda)\}$. A better p-value in terms of power uses the fact that $(\zeta_0, \beta_n, \sigma_0^2)$ are consistently estimated by $(\hat{\zeta}_n, \hat{\beta}_n, \hat{\sigma}_n^2)$ under any degree of (non)identification. The *plug-in* LF p-value $\hat{p}_n^{(LF)}(\lambda)$ uses $\widehat{\mathfrak{H}} \equiv \{h \in \mathfrak{H} : \theta = [\hat{\zeta}_n', \hat{\beta}_n', \pi']', \sigma^2 = \hat{\sigma}_n^2\}$ in place of $\mathfrak{H}$.[11]

The LF p-value does not exploit information that may point toward a particular identification case. The ICS procedure uses the sample to choose between $\sqrt{n}\beta_n \to b$ when $||b|| < \infty$ (weak and nonidentification) and $||b|| = \infty$ (semi-strong and strong identification). The statistic used to determine whether b is finite is

$$\mathcal{A}_n \equiv \left(\frac{1}{k_\beta} n \hat{\beta}_n' \hat{\Sigma}_{\beta,\beta,n}^{-1} \hat{\beta}_n\right)^{1/2}, \tag{14}$$

where $\hat{\Sigma}_{\beta,\beta,n}$ is the upper $k_\beta \times k_\beta$ block of $\hat{\Sigma}_n \equiv \widehat{\mathcal{H}}_n^{-1} \hat{\mathcal{V}}_n \widehat{\mathcal{H}}_n^{-1}$, and

$$\hat{\mathcal{V}}_n \equiv \frac{1}{n}\sum_{t=1}^n \epsilon_t^2(\hat{\theta}_n) d_{\theta,t}(\omega(\hat{\beta}_n), \hat{\pi}_n) d_{\theta,t}(\omega(\hat{\beta}_n), \hat{\pi}_n)'. \tag{15}$$

Now let $\{\kappa_n\}$ be a sequence of positive constants, with $\kappa_n \to \infty$ and $\kappa_n = o(\sqrt{n})$. The case $||b|| < \infty$ is selected when $\mathcal{A}_n \leq \kappa_n$, else $||b|| = \infty$ is selected. The type 1 ICS (ICS-1) p-value is:

$$p_n^{(ICS-1)}(\lambda) = \left\{ \begin{array}{ll} p_n^{(LF)}(\lambda) \text{ if } \mathcal{A}_n \leq \kappa_n, & p_n^\infty(\lambda) \text{ if } \mathcal{A}_n > \kappa_n \end{array} \right. .$$

A plug-in version is similar. Only when $\sqrt{n}||\beta_n|| \to \infty$ faster than $\kappa_n \to \infty$ will the chi-squared based p-value be chosen asymptotically with probability approaching one since then $\mathcal{A}_n/\kappa_n \xrightarrow{p} \infty$. 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.

The type 2 ICS (ICS-2) p-value involves a subtler comparison for category selection, cf. Andrews and Cheng (2012a, Sect. 5.3). Since our simulation study focuses on LF and ICS-1 p-values due to the added computational complexity

[11] The null hypothesis is tested by using a sample version of $E[\epsilon_t F(\lambda' \mathcal{W}(x_t))]$. Thus, so-called parametric *null imposed* p-values, similar to null imposed critical values in Andrews and Cheng (2012a) for t-, quasi-likelihood ratio, and Wald statistics, do not play a role here.

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

of ICS-2 p-values, and ICS-1 works well, we relegate ICS-2 details to the Supplementary Material (Appendix D).

A limit theory for the ICS-1 p-value requires the limit distribution of $\mathcal{A}_n$. It is easily derived along the lines of Theorems 4.1 and 4.2. Recall the augmented parameter set $\theta^+ \equiv [||\beta||, \omega(\beta)', \zeta', \pi']'$. Define $\mathcal{H}_\theta(\theta^+) = E[d_{\theta,t}(\omega(\beta),\pi)d_{\theta,t}(\omega(\beta),\pi)']$ and $\mathcal{V}(\theta^+) = E[\epsilon_t^2(\theta^+)d_{\theta,t}(\omega(\beta),\pi) \times d_{\theta,t}(\omega(\beta),\pi)']$, and

$$\Sigma(\theta^+) \equiv \mathcal{H}_\theta(\theta^+)^{-1}\mathcal{V}(\theta^+)\mathcal{H}_\theta(\theta^+)^{-1},\ \Sigma(\omega,\pi) \equiv \Sigma(\|\beta_0\|,\omega,\zeta_0,\pi) \tag{16}$$
$$\bar{\Sigma}(\pi,b) = \left[\bar{\Sigma}_{i,j}(\pi,b)\right]_{i,j=1}^{k_\theta} \equiv \Sigma\left(\omega^*(\pi,b),\pi\right).$$

THEOREM 5.1. *Let Assumptions 1 and 2, and H_0, hold. a. Under drift case $\mathcal{C}(i,b)$ with $||b|| < \infty$, $\mathcal{A}_n \xrightarrow{d} \mathcal{A}(b) \equiv \{\tau_\beta(\pi^*(b),b)' \times \bar{\Sigma}_{\beta,\beta}^{-1}(\pi^*(b),b) \times \tau_\beta(\pi^*(b),b)/k_\beta\}^{1/2}$, where $\bar{\Sigma}_{\beta,\beta}(\pi,b) \equiv [\bar{\Sigma}_{i,j}(\pi,b)]_{i,j=1}^{k_\beta}$. b. Let $\{\kappa_n\}$ be a sequence of positive constants, $\kappa_n \to \infty$ and $\kappa_n = o(\sqrt{n})$. Under drift case $\mathcal{C}(ii,\omega_0)$ we have $\mathcal{A}_n \xrightarrow{p} \infty$. If $\sqrt{n}||\beta_n||/\kappa_n = O(1)$ then $\kappa_n^{-1}\mathcal{A}_n \xrightarrow{p} [0,\infty)$. If $\sqrt{n}||\beta_n||/\kappa_n \to \infty$ then $\kappa_n^{-1}\mathcal{A}_n \xrightarrow{p} \infty$, for example when $\beta_0 \neq 0$ for any sequence $\{\kappa_n\}$ defined above.*

Remark 8. Intuitively, $\kappa_n^{-1}\mathcal{A}_n \xrightarrow{p} \infty$ (and therefore the ICS-1 p-value is based on the chi-squared distribution) only when there is strong evidence in favor of strong identification. If $\sqrt{n}||\beta_n|| \to \infty$ too slowly, in this case $\sqrt{n}||\beta_n||/\kappa_n = O(1)$, then the LF value is selected, which leads always to asymptotically correct inference.

6. *p*-VALUE ASYMPTOTICS AND COMPUTATION

Let $p_n^{(\cdot)}(\lambda)$ be the LF or ICS-1 p-value. We first prove that $p_n^{(\cdot)}(\lambda)$ leads to a test with correct asymptotic level. Analogous results carry over to plug-in versions. We then show how to bootstrap the key component $p_n(\lambda,h) \equiv 1 - \mathcal{F}_{\lambda,h}(\mathcal{T}_n(\lambda))$ where $\mathcal{F}_{\lambda,h}(c) \equiv P(\mathcal{T}_\psi(\lambda,h) \leq c)$, with accompanying limit theory. Once we have an asymptotically valid approximation for $p_n(\lambda,h)$, a robust p-value follows as in Section 5. The same method leads to robust critical value approximations (see Hill, 2020, Appendix E).

6.1. Asympotics $p_n^{(\cdot)}(\lambda)$

Technical arguments are made feasible when $\mathcal{F}_{\lambda,h}(c) \equiv P(\mathcal{T}_\psi(\lambda,h) \leq c)$ is continuous, because $p_n^{(\cdot)}(\lambda)$ contains $\mathcal{F}_{\lambda,h}(\mathcal{T}_n(\lambda))$ which is evaluated by a weak limit theory and the continuous mapping theorem. See Hill (2020, Appendix F) for discussion.

Assumption 5. (p-Value). *a.$\mathcal{F}_{\lambda,h}(\cdot)$ is continuous a.e. on $[0,\infty)$, $\forall h \in \mathfrak{H}$.*[12] *b.* The ICS-1 threshold sequence $\{\kappa_n\}$ satisfies $\kappa_n \to \infty$ and $\kappa_n = o(\sqrt{n})$.

[12]The bulk of Assumptions LF and V3 in Andrews and Cheng (2012a) ensure distribution continuity at nonrandom critical value points for each α. Since these must hold for *any* nominal level α, Assumption 5 is not restrictive by comparison.

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

Let F_γ be the distribution function of $W_t = [y_t, x_t']'$ under some $\gamma \in \Gamma^*$, where Γ^* is the true parameter space in (12). Let P_γ denote probability under F_γ. For any p-value $p_n^{(\cdot)}(\lambda)$ and each λ, the asymptotic size of the test is the asymptotic maximum rejection probability over γ such that the null is true: $AsySz(\lambda) = \limsup_{n\to\infty} \sup_{\gamma\in\Gamma^*} P_\gamma(p_n^{(\cdot)}(\lambda) < \alpha | H_0)$. Uniform size control over λ is captured by $AsySz \equiv \sup_{\lambda\in\Lambda} AsySz(\lambda)$.

THEOREM 6.1. *Let Assumptions 1, 2, 4, and 5 hold.*

a. LF and ICS-1 $p_n^{(\cdot)}(\lambda)$ satisfy $AsySz \le \alpha$.

b. Let $H_1 : \sup_{\theta\in\Theta} P(E[y_t|x_t] = f(\theta, x_t)) < 1$ be true. Then $p_n^{(\cdot)}(\lambda) \xrightarrow{p} 0$ for all $\lambda \in \Lambda/S$ where $S \subset \Lambda$ has Lebesgue measure zero.

Remark 9. Under Assumptions K, LF, and V3 in Andrews and Cheng (2012a), the robust ICS-1 critical value leads to a correctly sized test. The assumptions primarily concern continuity of $\mathcal{F}_{\lambda,h}$ at the critical value for a given level α. In turn, these allow for $AsySz$ to be reduced by their Lemma 2.1, a key step toward proving $AsySz = \alpha$. The robust p-value requires that $\mathcal{F}_{\lambda,h}(c) \equiv P(\mathcal{T}_\psi(\lambda, h) \le c)$ be continuous everywhere, and force us to exploit probability bounds rather than their Lemma 2.1. Hence, we can only prove that $p_n^{(\cdot)}(\lambda)$ yields a correct uniform asymptotic $levelAsySz \le \alpha$. That seems irrelevant in small sample experiments since correct size appears to be achieved: see Section 7.

6.2. Computation of $p_n^{(\cdot)}(\lambda)$

We propose a wild bootstrap method for computing $p_n(\lambda, h) \equiv 1 - \mathcal{F}_{\lambda,h}(\mathcal{T}_n(\lambda))$. A similar method applies to bootstrapping the chi-squared based p-value $p_n^\infty(\lambda) \equiv 1 - \mathcal{F}_\infty(\mathcal{T}_n(\lambda))$ under strong identification (see, e.g., Hansen, 1996) . This may be an attractive option in small samples where the chi-squared distribution may not well approximate the small sample distribution of $\mathcal{T}_n(\lambda)$ under strong identification.

Operate under H_0, and under weak identification $\sqrt{n}\beta_n \to b$ and $||b|| < \infty$. Set for the sake of brevity $\beta_n = b/\sqrt{n}$ where $||b|| < \infty$ indexes the true value β_n. We first give the steps for computing $p_n(\lambda, h)$, and then prove its validity. In all that follows, independence is conditional on the sample.

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

. Define $\epsilon_t(\psi, \pi) \equiv y_t - \zeta' x_t - \beta' g(x_t, \pi)$ and estimators:

$$\widehat{\mathcal{H}}_{\psi,n}(\pi) \equiv \frac{1}{n}\sum_{t=1}^n d_{\psi,t}(\pi) d_{\psi,t}(\pi)' \text{ and } \widehat{\mathcal{H}}_n(\omega, \pi) = \frac{1}{n}\sum_{t=1}^n d_{\theta,t}(\omega, \pi) d_{\theta,t}(\omega, \pi)'$$

$$\hat{\mathcal{D}}_{\psi,n}(\pi, \pi_0) \equiv -\frac{1}{n}\sum_{t=1}^n d_{\psi,t}(\pi) g(x_t, \pi_0)' \text{ and } \hat{\mathcal{K}}_{\psi,n,t}(\pi, \lambda)$$

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

$$\equiv F\left(\lambda'\mathcal{W}(x_t)\right) - \hat{\mathfrak{b}}_{\psi,n}(\pi,\lambda)'\widehat{\mathcal{H}}_{\psi,n}^{-1}(\pi)d_{\psi,t}(\pi)$$

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

$$\equiv \frac{1}{n}\sum_{t=1}^{n} F\left(\lambda'\mathcal{W}(x_t)\right) d_{\theta,t}(\omega,\pi).$$

6.2.1. *Step 2: Draw from* $\pi^*(b)$. By Assumption 2, $\pi^*(b) \equiv \arg\inf_{\pi\in\Pi}\xi_\psi(\pi,b) \equiv -\arg\inf_{\pi\in\Pi}\{\mathcal{S}_\beta\mathcal{H}_\psi^{-1}(\pi)(\mathcal{G}_\psi(\pi)+\mathcal{D}_\psi(\pi)b)\}$. Under weak identification, Lemma B.1 yields that $\{\mathcal{G}_\psi(\pi) : \pi \in\Pi\}$ is the weak limit of

$$\mathcal{G}_{\psi,n}(\psi_{0,n},\pi) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\{\epsilon_t(\psi_{0,n},\pi)d_{\psi,t}(\pi) - E\left[\epsilon_t(\psi_{0,n},\pi)d_{\psi,t}(\pi)\right]\right\} \tag{17}$$

$$= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\epsilon_t d_{\psi,t}(\pi)$$

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

By the argument used to prove Lemma B.2, $\sup_{\pi\in\Pi}|1/n\sum_{t=1}^{n}\{d_{\psi,t}(\pi)g(x_t,\pi_0)' - E[d_{\psi,t}(\pi)g(x_t,\pi_0)']\}| \xrightarrow{p} 0$. Hence, by $E[\epsilon_t|x_t] = 0$ 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

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

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

Next, let $\{z_t\}_{t=1}^n$ be a sequence of independent draws from $N(0,1)$, and define $\widehat{\mathcal{G}}_{\psi,n}^*(\pi) \equiv 1/\sqrt{n}\sum_{t=1}^{n} z_t\widehat{\mathcal{H}}_{\psi,n}^{-1/2}(\pi)d_{\psi,t}(\pi)$. By $\hat{\sigma}_n \xrightarrow{p} \sigma_0$ and the proof of Theorem 6.2, below, $\{\hat{\sigma}_n\widehat{\mathcal{G}}_{\psi,n}^*(\pi) : \pi\in\Pi\} \Rightarrow^p \{\mathcal{H}_\psi^{-1/2}(\pi)\mathcal{G}_\psi(\pi) : \pi\in\Pi\}$, where $\Rightarrow^p$ denotes weak convergence in probability defined in Gine and Zinn (1990, Section 3).[13] Thus, $\hat{\sigma}_n\widehat{\mathcal{G}}_{\psi,n}^*(\pi)$ 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 $\{\widehat{\mathcal{G}}_{\psi,n}^*(\pi), \widehat{\mathcal{H}}_{\psi,n}(\pi), \hat{\mathcal{D}}_{\psi,n}(\pi,\pi_0)\}$ to compute

$$\hat{\xi}_{\psi,n}^*(\pi,\pi_0,b) = -\frac{1}{2}\left\{\hat{\sigma}_n\widehat{\mathcal{G}}_{\psi,n}^*(\pi) + \widehat{\mathcal{H}}_{\psi,n}^{-1/2}(\pi)\hat{\mathcal{D}}_{\psi,n}(\pi,\pi_0)\times b\right\}'$$

$$\left\{\hat{\sigma}_n\widehat{\mathcal{G}}_{\psi,n}^*(\pi) + \widehat{\mathcal{H}}_{\psi,n}^{-1/2}(\pi)\hat{\mathcal{D}}_{\psi,n}(\pi,\pi_0)\times b\right\}.$$

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

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

The bootstrapped $\pi^*(b)$ is therefore:

$$\hat{\pi}_n^*(\pi_0, b) = \arg\min_{\pi\in\Pi} \left\{ \hat{\xi}_{\psi,n}^*(\pi, \pi_0, b) \right\}. \tag{18}$$

6.2.1. *Step 3: Draw from* $\mathfrak{Z}_\psi(\pi,\lambda)$. By Lemma A.8.a, under the null $\mathfrak{Z}_\psi(\pi,\lambda)$ is the zero-mean Gaussian limit process of $1/\sqrt{n}\sum_{t=1}^n \epsilon_t \mathcal{K}_{\psi,t}(\pi,\lambda)$. Use the Step 2 draws $\{z_t\}_{t=1}^n$ to define

$$\hat{\mathfrak{Z}}_{\psi,n}^*(\pi,\lambda) \equiv \frac{1}{\sqrt{n}} \sum_{t=1}^n z_t \left(F\left(\lambda'\mathcal{W}(x_t)\right) - \hat{\mathfrak{b}}_{\psi,n}(\pi,\lambda)' \widehat{\mathcal{H}}_{\psi,n}^{-1}(\pi) d_{\psi,t}(\pi) \right). \tag{19}$$

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

6.2.1. *Step 4:* $\tau_\beta(\cdot), \hat{\mathfrak{T}}_\psi(\cdot), \bar{v}^2(\cdot), \mathcal{T}_{,\psi}(\cdot)$. We now have all the required components for computing the following key quantities (recall $\mathcal{S}_\beta \equiv [I_{k_\beta} : 0_{k_x \times k_x}]$):

$$\hat{\tau}_{\beta,n}^*(\pi_0, b) \equiv -\mathcal{S}_\beta \widehat{\mathcal{H}}_{\psi,n}^{-1}(\hat{\pi}_n^*(\pi_0,b)) \left\{ \hat{\sigma}_n \widehat{\mathcal{G}}_{\psi,n}^*(\hat{\pi}_n^*(\pi_0,b)) + \hat{\mathcal{D}}_{\psi,n}(\hat{\pi}_n^*(\pi_0,b), \pi_0) \times b \right\} \tag{20}$$

$$\hat{\omega}_n^*(\pi_0, b) \equiv \frac{\hat{\tau}_{\beta,n}^*(\pi_0,b)}{\left\| \hat{\tau}_{\beta,n}^*(\pi_0,b) \right\|}$$

$$\begin{aligned}
&\hat{\mathfrak{T}}_{\psi,n}^*(\pi,\lambda,\pi_0,b) \\
&\quad \equiv \hat{\sigma}_n \hat{\mathfrak{Z}}_{\psi,n}^*(\pi,\lambda) + \hat{\mathfrak{b}}_{\psi,n}(\hat{\psi}_n,\pi,\lambda)' \left(\widehat{\mathcal{H}}_{\psi,n}^{-1}(\pi) \hat{\mathcal{D}}_{\psi,n}(\pi,\pi_0) \times b + \left[b, 0_{k_\beta}' \right]' \right) \\
&\qquad + \hat{\mathfrak{b}}_{\psi,n}(\hat{\psi}_n,\pi,\lambda)' \widehat{\mathcal{H}}_{\psi,n}^{-1}(\pi) \frac{1}{n} \sum_{t=1}^n d_{\psi,t}(\pi) \{ g(x_t,\pi_0) - g(x_t,\pi) \}' b \\
&\qquad + \frac{1}{n} \sum_{t=1}^n \left\{ F\left(\lambda'\mathcal{W}(x_t)\right) - \hat{\mathfrak{b}}_{\psi,n}(\pi,\lambda)' \widehat{\mathcal{H}}_{\psi,n}^{-1}(\pi) d_{\psi,t}(\pi) \right\} \\
&\qquad \{ g(x_t,\pi_0) - g(x_t,\pi) \}' b
\end{aligned}$$

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

$$\widehat{\bar{v}}_n^2(\pi,\lambda,b) \equiv \hat{v}_n^2(\hat{\omega}_n^*(\pi_0,b),\pi,\lambda).$$

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

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

$$\hat{\mathcal{T}}^*_{\psi,n}(\pi,\lambda,\pi_0,b) \equiv \left(\frac{\hat{\mathfrak{T}}^*_{\psi,n}(\pi,\lambda,\pi_0,b)}{\hat{\bar{v}}_n(\pi,\lambda,b)}\right)^2. \tag{21}$$

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

6.2.1. *Step 5.* Repeat Steps 1–4 $\mathcal{M}$ times resulting in a sequence of independent draws $\{\hat{\mathcal{T}}^*_{\psi,n,j}(\lambda,h)\}_{j=1}^{\mathcal{M}}$. The p-value approximation is simply:

$$\hat{p}^*_{n,\mathcal{M}}(\lambda,h) \equiv \frac{1}{\mathcal{M}}\sum_{j=1}^{\mathcal{M}} I\left(\hat{\mathcal{T}}^*_{\psi,n,j}(\lambda,h) > \mathcal{T}_n(\lambda)\right).$$

Let $\hat{p}^{(\cdot)}_{n,\mathcal{M}}(\lambda)$ be the LF or ICS-1 p-value computed with $\hat{p}^*_{n,\mathcal{M}}(\lambda,h)$, and the corresponding asymptotic size $AsySz^*(\lambda) \equiv \limsup_{n\to\infty}\sup_{\gamma\in\Gamma^*} P_\gamma(\hat{p}^{(\cdot)}_{n,\mathcal{M}}(\lambda) < \alpha|H_0)$ and $AsySz^* \equiv \sup_{\lambda\in\Lambda} AsySz^*(\lambda)$. $\hat{p}^*_{n,\mathcal{M}}(\lambda,h)$ is consistent for the asymptotic p-value under weak identification $p_n(\lambda,h) \equiv 1 - \mathcal{F}_{\lambda,h}(\mathcal{T}_n(\lambda))$, and the resulting test achieves the correct uniform asymptotic level, $AsySz^* \leq \alpha$.

In order to demonstrate $AsySz^* \leq \alpha$ we need to verify uniform convergence $\sup_{\lambda\in\Lambda}|\hat{p}^*_{n,\mathcal{M}_n}(\lambda,h) - p_n(\lambda,h)| \xrightarrow{p} 0$. Due to the nonsmooth structure of $\hat{p}^*_{n,\mathcal{M}_n}(\lambda,h)$ and how it enters $AsySz^*$, we need additional structure on key processes. We exploit properties of the Vapnik–Červonenkis *subgraph* class of functions, denoted $\mathcal{V}(\mathcal{C})$. The $\mathcal{V}(\mathcal{C})$ class is large: it contains indicator, monotonic and continuous functions; and $\mathcal{V}(\mathcal{C})$ mappings of $\mathcal{V}(\mathcal{C})$ functions are in $\mathcal{V}(\mathcal{C})$, including linear combinations, minima, maxima, products and indicator transforms. See, for example, van der Vaart and Wellner (1996, Chap. 2.6) for a compendium of $\mathcal{V}(\mathcal{C})$ properties.[14] See Vapnik and Červonenkis (1971), Dudley (1978, Section 7), van der Vaart and Wellner (1996, Sect. 2), and see Pollard (1984, Chap. II.4) for the closely related *polynomial discrimination* class.

Write $F_{n,\lambda}(c) \equiv P(\mathcal{T}_n(\lambda) \leq c)$ and $F^*_{n,\lambda,h}(c) \equiv P(\hat{\mathcal{T}}^*_{\psi,n,1}(\lambda,h) \leq c|\mathfrak{W}_n)$ where $\mathfrak{W}_n \equiv \{(y_t,x_t)\}_{t=1}^n$.

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 $\{F^*_{n,\lambda,h}(c) : \lambda \in \Lambda, c \in [0,\infty)\}$ belong to the $\mathcal{V}(\mathcal{C})$ class.

Remark 10. The popularly used logistic and exponential weight functions $F(\cdot)$ are in $\mathcal{V}(\mathcal{C})$ because they are continuous. Under Assumption 5, we impose

[14] We exploit the facts that an indicator function of a $\mathcal{V}(\mathcal{C})$ index function is in $\mathcal{V}(\mathcal{C})$, and a continuous function evaluated at a $\mathcal{V}(\mathcal{C})$ function is in $\mathcal{V}(\mathcal{C})$.

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

continuity on the distribution function $F_{n,\lambda}(\cdot)$, but we need more structure here to handle uniform asymptotics over λ for the bootstrapped p-value. The index functions $F_{n,\lambda}(c)$ and $F^*_{n,\lambda,h}(c)$ need to be in $\mathcal{V}(\mathcal{C})$ both in terms of the argument c *and* the index λ since they are evaluated at the test statistics $\{\hat{\mathcal{T}}^*_{\psi,n,j}(\lambda,h), \mathcal{T}_n(\lambda)\}$.

THEOREM 6.2. *Let $\mathcal{M} = \mathcal{M}_n \to \infty$ as $n \to \infty$, and let Assumptions 1, 2, 4, and 5 hold. a.* $|\hat{p}^*_{n,\mathcal{M}}(\lambda,h) - p_n(\lambda,h)| \xrightarrow{p} 0$. *b. If additionally Assumption 6 holds, then* $\sup_{\lambda\in\Lambda} |\hat{p}^*_{n,\mathcal{M}}(\lambda,h) - p_n(\lambda,h)| \xrightarrow{p} 0$ *and* $AsySz^* \le \alpha$.

Finally, we consider a theory for the PVOT test. Define the LF or ICS-1 PVOT $\hat{\mathcal{P}}_{n,\mathcal{M}}(\alpha) \equiv \int_\Lambda I(\hat{p}^{(\cdot)}_{n,\mathcal{M}}(\lambda) < \alpha)d\lambda$. The test rejects H_0 when $\hat{\mathcal{P}}^{(\cdot)}_{n,\mathcal{M}}(\alpha) > \alpha$. The (nonuniform) asymptotic level of the test is α, and the test is consistent.

THEOREM 6.3. *Let $\mathcal{M} = \mathcal{M}_n \to \infty$ as $n \to \infty$, and let Assumptions 1, 2, 4, and 5 hold. Under H_0,* $\lim_{n\to\infty} P(\hat{\mathcal{P}}_{n,\mathcal{M}}(\alpha) > \alpha) \le \alpha$. *Conversely,* $P(\hat{\mathcal{P}}_{n,\mathcal{M}}(\alpha) > \alpha) \to 1$ *under* $H_1 : \sup_{\theta\in\Theta} P(E[y_t|x_t] = f(\theta,x_t)) < 1$.

The (uniform) asymptotic size of the PVOT test is $AsySz(pvot) = \limsup_{n\to\infty} \sup_{\gamma\in\Gamma^*} P_\gamma(\hat{\mathcal{P}}_{n,\mathcal{M}}(\alpha) > \alpha|H_0)$. Under the additional structure of Assumption 6, $AsySz(pvot) \le \alpha$. Hence, the PVOT test controls for size uniformly.

THEOREM 6.4. *Let $\mathcal{M} = \mathcal{M}_n \to \infty$ as $n \to \infty$, and let Assumptions 1, 2, 4, 5, and 6 hold. Then $AsySz(pvot) \le \alpha$.*

7. MONTE CARLO STUDY

We now perform a simulation study in order to assess how well the proposed bootstrap method works.

7.1. Set up

Throughout ϵ_t is i.i.d. $N(0,1)$ distributed, 10,000 samples are generated, and sample sizes are $n \in \{100, 250, 500\}$. The wild bootstrap used for robust p-value computation, and for the supremum and average tests, uses 500 bootstrap samples to reduce computation time.

The data generating process is

$$y_t = \zeta_0 y_{t-1} + \beta_n y_{t-1} \frac{1}{1+\exp\{-10\,(y_{t-1}-\pi_0)\}} + \varpi_0 \frac{1}{1+y^2_{t-1}} + \epsilon_t.$$

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

If $\varpi_0 = 0$, then y_t is a Logistic STAR model and the null hypothesis is true. We use a fixed value 10 for the speed of transition to reduce computation complexity (see also Andrews and Cheng, 2013).[15]

We use $\zeta_0 = .6, \pi_0 = 0$ and $\varpi_0 \in \{0, .03, .3\}$. The latter allows us to inspect power against weak and strong degrees of deviation from a STAR null hypothesis. The key parameter for identification cases takes values $\beta_n \in \{.3, .3/\sqrt{n}, 0\}$, representing strong identification, weak identification with $\sqrt{n}\beta_n = b = .3$ and $\beta_n \to \beta_0 = 0$, and nonidentification with $\beta_n = \beta_0 = 0$. Other values for (ζ_0, β_n) lead to similar results.

Let $\iota = 10^{-10}$. The true parameter spaces are $\mathcal{B}^* = [-1 + 2\iota, 1 - 2\iota]$, $\mathcal{Z}^*(\beta) = [-1 - \beta + \iota < \zeta < 1 - \beta - \iota]$, and $\Pi^* = [-1, 1]$. The estimation spaces are $\mathcal{B} = [-1+\iota, 1-\iota]$, $\mathcal{Z}(\beta) = [-1 - \beta < \zeta < 1 - \beta]$, and $\Pi = [-2, 2]$. Thus $|\zeta + \beta| < 1$ on $\Theta \equiv \mathcal{B} \times \mathcal{Z}(\beta) \times \Pi$.

The estimated model is an LSTAR:

$$y_t = \zeta_0 y_{t-1} + \beta_0 y_{t-1} \frac{1}{1 + \exp\{-10\,(y_{t-1} - \pi_0)\}} + \epsilon_t.$$

We draw 100 start values θ from the uniform distribution on Θ and estimate $\theta_0 = [\zeta_0, \beta_0, \pi_0]'$ by least squares, resulting in 100 estimates $\{\hat{\theta}_{n,i}\}_{i=1}^{100}$. The final estimate $\hat{\theta}_n$ minimizes the least-squares criterion over $\{\hat{\theta}_{n,i}\}_{i=1}^{100}$.[16] The CM weight is logistic $F(u) = 1/(1 + \exp\{u\})$, and $F(\lambda' \mathcal{W}(x_t))$ uses the bounded one-to-one transform $\mathcal{W}(x) = \text{atan}(x)$ as in Bierens (1990, p. 1445, 1453). The parameter space is $\Lambda = [1, 5]$. We use a discretization Λ_n with endpoints $\{1, 5\}$, and equal increments with n elements (e.g., $\Lambda_{100} = \{1, 1.04, 1.08, \ldots, 5\}$).

Eleven tests are performed. The first five are not robust to weak identification: (*i*) uniformly randomly chosen λ^* from Λ_n, compute $\mathcal{T}_n(\lambda^*)$ and use $\chi^2(1)$ for p-value computation; (*ii*) $\sup_{\lambda \in \Lambda_n} p_n(\lambda)$; (*iii*) $\sup_{\lambda \in \Lambda_n} \mathcal{T}_n(\lambda)$ and (*iv*) $\int_{\Lambda_n} \mathcal{T}_n(\lambda) \mu(d\lambda)$ where μ is the uniform measure on Λ_n, and p-values are computed by wild bootstrap; and (*v*) the PVOT test using Λ_n, and a p-value computed from the $\chi^2(1)$ distribution [PVOT-χ^2].

The final six tests are robust. We compute $\mathcal{T}_n(\lambda^*)$ using (*vi*) the plug-in LF and (*vii*) plug-in ICS-1 p-values [$\mathcal{T}_n(\lambda^*)$-LF, $\mathcal{T}_n(\lambda^*)$ -ICS]; $\sup_{\lambda \in \Lambda_n} p_n(\lambda)$ using (*viii*) the plug-in LF and (*ix*) plug-in ICS-1 p-values [supp_n-LF, supp_n -ICS]; and PVOT using (*x*) the plug-in LF and (*xi*) plug-in ICS-1 p-values (PVOT-LF, PVOT-ICS). The bootstrap procedure in Section 6.2 is used to approximate the p-value under weak identification $p_n(\lambda, h)$ with $\hat{p}^*_{n,\mathcal{M}}(\lambda, h)$. Then $\hat{p}^*_{n,\mathcal{M}}(\lambda, h)$ is used to compute the plug-in LF and plug-in ICS-1 p-values from Section 5. Using ICS-2 would naturally lead to an improved p-value, but the computational cost is too great at this time.

[15] In empirical work when the transition function is, for example, $(1 + \exp\{-\pi_{0,1}(y_{t-d} - \pi_{0,2})\})^{-1}$, often the transition speed $\pi_{0,1}$ or threshold $\pi_{0,2}$ are fixed to ease computation, for example, Lundbergh and Terasvirta (2006, p. 592) and Gonzalez-Rivera (1998, Definition. 1).

[16] An analytic gradient is used for optimization. The criterion tolerance for ceasing iterations is $1e^{-8}$, and the maximum number of allowed iterations is 20,000.

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

TABLE 1. STAR test rejection frequencies: Sample size $n = 100$

	H_0: LSTAR			H_1-weak			H_1-strong		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
Strong identification: $\beta_n = .3$									
sup $\mathcal{T}_n$	.025	.094	.163	.147	.280	.365	.757	.872	.907
aver $\mathcal{T}_n$	.025	.078	.135	.087	.209	.289	.552	.726	.804
rand $\mathcal{T}_n$	.011	.052	.096	.053	.143	.232	.446	.635	.732
rand LF	**.007**	**.015**	**.038**	**.013**	**.066**	**.141**	**.442**	**.553**	**.661**
rand ICS-1	**.013**	**.050**	**.089**	**.028**	**.089**	**.170**	**.379**	**.593**	**.692**
sup p_n	.009	.039	.068	.036	.118	.209	.378	.554	.656
sup p_n LF	**.006**	**.009**	**.032**	**.012**	**.057**	**.120**	**.262**	**.457**	**.572**
sup p_n ICS-1	**.006**	**.036**	**.061**	**.020**	**.081**	**.138**	**.310**	**.506**	**.617**
PVOT	.015	.065	.124	.101	.257	.335	.727	.859	.883
PVOT LF	**.007**	**.014**	**.052**	**.026**	**.121**	**.208**	**.552**	**.781**	**.817**
PVOT ICS-1	**.007**	**.043**	**.073**	**.042**	**.153**	**.237**	**.622**	**.815**	**.842**
Weak identification: $\beta_n = .3/\sqrt{n}$									
sup $\mathcal{T}_n$	.064	.155	.239	.337	.574	.681	.929	.978	.993
aver $\mathcal{T}_n$	.057	.146	.219	.215	.430	.554	.739	.888	.932
rand $\mathcal{T}_n$	.027	.083	.175	.164	.343	.474	.604	.810	.870
rand LF	**.012**	**.042**	**.093**	**.060**	**.161**	**.308**	**.467**	**.685**	**.794**
rand ICS-1	**.012**	**.046**	**.104**	**.116**	**.261**	**.382**	**.545**	**.749**	**.841**
sup p_n	.019	.087	.145	.107	.253	.411	.493	.700	.785
sup p_n LF	**.001**	**.061**	**.084**	**.036**	**.124**	**.230**	**.351**	**.598**	**.698**
sup p_n ICS-1	**.001**	**.065**	**.085**	**.088**	**.193**	**.335**	**.454**	**.663**	**.756**
PVOT	.038	.127	.196	.328	.542	.591	.893	.968	.950
PVOT LF	**.015**	**.049**	**.108**	**.108**	**.320**	**.398**	**.710**	**.911**	**.916**
PVOT ICS-1	**.014**	**.049**	**.107**	**.221**	**.435**	**.486**	**.830**	**.942**	**.932**
Nonidentification: $\beta_n = \beta_0 = 0$									
sup $\mathcal{T}_n$	.066	.164	.249	.358	.584	.696	.902	.970	.983
aver $\mathcal{T}_n$	.062	.148	.226	.233	.438	.548	.716	.872	.911
rand $\mathcal{T}_n$	.044	.107	.186	.184	.380	.505	.634	.793	.864
rand LF	**.013**	**.046**	**.115**	**.069**	**.191**	**.327**	**.498**	**.725**	**.818**
rand ICS-1	**.013**	**.047**	**.116**	**.137**	**.298**	**.481**	**.583**	**.769**	**.847**
sup p_n	.018	.080	.167	.117	.272	.363	.514	.710	.807
sup p_n LF	**.011**	**.043**	**.083**	**.042**	**.122**	**.221**	**.383**	**.612**	**.740**
sup p_n ICS-1	**.011**	**.044**	**.086**	**.093**	**.205**	**.293**	**.464**	**.683**	**.783**
PVOT	.049	.134	.190	.322	.554	.624	.890	.962	.957

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

TABLE 1. Continued

	H_0: LSTAR			H_1-weak			H_1-strong		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
PVOT LF	**.015**	**.061**	**.117**	**.122**	**.322**	**.415**	**.740**	**.911**	**.936**
PVOT ICS-1	**.015**	**.057**	**.116**	**.253**	**.464**	**.570**	**.847**	**.939**	**.954**

Note: Numerical values are rejection frequency at the given level. LSTAR is Logistic STAR. Empirical power is not size-adjusted. *sup* $\mathcal{T}_n$ and *ave* $\mathcal{T}_n$ tests are based on a wild bootstrapped p-value. *rand* $\mathcal{T}_n$: $\mathcal{T}_n(\lambda)$ with uniformly randomly chosen λ on [1,5]. *sup* p_n is the supremum p-value test where p-values are computed from the chi-squared distribution. PVOT uses the chi-squared distribution. LF implies the *least favorable* p-value is used, and ICS-1 implies the type 1 *identification category selection* p-value is used with threshold $\kappa_n = \ln(\ln(n))$.

Theorems 6.2–6.4 provide the theory demonstrating robustness and correct asymptotic size or level for tests (*vi*)–(*xi*). Uniform bootstrap p-value convergence, and uniform size control, require Assumption 6: logistic $F(\cdot)$ is in the $\mathcal{V}(\mathcal{C})$ class. We need to assume $\{F_{n,\lambda}(c) : \lambda \in \Lambda, c \in [0,\infty)\}$ and $\{F^*_{n,\lambda,h}(c) : \lambda \in \Lambda, c \in [0,\infty)\}$ belong to the $\mathcal{V}(\mathcal{C})$ class due to their nonlinear complexity.

The computation of LF and ICS p-values using $\hat{p}^*_{n,\mathcal{M}}(\lambda, h)$ requires a grid of nuisance parameters $h = (\pi, b)$. We use $\pi \in \{-2, -1.5, \ldots, 1.5, 2\}$ and $b \in \{-.5, -.3, -.2, -.1, 0, .1, .2, .3, .5\}$. Finer grids lead to significant increases in computation time. The ICS-1 p-value requires the threshold κ_n: we use $\kappa_n = a\ln(\ln(n))$ with $a = 1$. Values of a close to 1 lead to similar results, while under rejection of the null is more prominent as a increases. Larger rates of increase for κ_n like $c(\ln(n))^\delta$ for some $c, \delta > 0$ generally result in the ICS-1 p-value being nearly equal to the LF p-value, at least within our chosen design. Indeed, under $\kappa_n = (\ln(n))^{1/2}$ there is little difference between LF and ICS-1 values. Finally, the selection matrix $\mathcal{S}_\beta \equiv [I_{k_\beta} : 0_{k_x \times k_x}]$ reduces to $\mathcal{S}_\beta \equiv [1, 0]$ since $k_\beta = k_x = 1$.

7.2. Results

Rejection frequencies are given in Tables 1–3.

7.2.1. *Strong Identification.* Consider the strong identification case $\beta_n = .3$. Under the null, $\mathcal{T}_n(\lambda^*)$ and PVOT-χ^2 exhibit rejection rates close to the nominal levels. The supremum test is over-sized and exhibits the largest size distortion, while the average test is slightly over-sized.

The LF p-value leads to under-rejection for $\mathcal{T}_n(\lambda^*)$-LF and PVOT-LF at each n. The ICS-1 p-value with $\kappa_n = \ln(\ln(n))$ corrects for the size distortion in nearly every case. The exceptions are when $n = 100$ at the 10% level for both $\mathcal{T}_n(\lambda^*)$-ICS and PVOT-ICS, and when $n = 250$ at the 10% level for $\mathcal{T}_n(\lambda^*)$-ICS. In these cases, empirical size is roughly 7%. The PVOT-ICS therefore yields sharp size in nearly every case. The slight advantage of ICS-1 over LF is not surprising since the LF p-value is generally larger.

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

TABLE 2. STAR test rejection frequencies: Sample size $n = 250$

	H_0: LSTAR			H_1-weak			H_1-strong		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
Strong identification: $\beta_n = .3$									
sup $\mathcal{T}_n$	.018	.088	.163	.359	.468	.551	.953	.984	.990
aver $\mathcal{T}_n$	.014	.077	.133	.262	.387	.468	.873	.949	.975
rand $\mathcal{T}_n$	.014	.064	.126	.165	.299	.396	.793	.912	.952
rand LF	**.001**	**.010**	**.025**	**.067**	**.235**	**.368**	**.688**	**.888**	**.936**
rand ICS-1	**.008**	**.031**	**.077**	**.076**	**.244**	**.375**	**.762**	**.902**	**.947**
sup p_n	.003	.039	.066	.103	.264	.358	.743	.876	.917
sup p_n LF	**.000**	**.007**	**.021**	**.032**	**.214**	**.303**	**.605**	**.838**	**.899**
sup p_n ICS-1	**.003**	**.035**	**.063**	**.038**	**.217**	**.316**	**.714**	**.870**	**.912**
PVOT	.016	.067	.125	.328	.437	.517	.952	.983	.991
PVOT LF	**.004**	**.020**	**.041**	**.132**	**.348**	**.417**	**.938**	**.972**	**.976**
PVOT ICS-1	**.011**	**.051**	**.108**	**.147**	**.370**	**.433**	**.947**	**.978**	**.985**
Weak identification: $\beta_n = .3/\sqrt{n}$									
sup $\mathcal{T}_n$	.051	.139	.224	.764	.922	.957	.992	1.00	1.00
aver $\mathcal{T}_n$	.046	.118	.215	.539	.779	.853	.969	.992	.998
rand $\mathcal{T}_n$	.027	.086	.169	.451	.695	.785	.911	.979	.993
rand LF	**.018**	**.060**	**.097**	**.180**	**.481**	**.641**	**.851**	**.961**	**.980**
rand ICS-1	**.018**	**.058**	**.098**	**.298**	**.633**	**.770**	**.926**	**.975**	**.991**
sup p_n	.017	.056	.097	.330	.615	.712	.858	.975	.991
sup p_n LF	**.008**	**.026**	**.067**	**.115**	**.416**	**.587**	**.698**	**.926**	**.978**
sup p_n ICS-1	**.008**	**.030**	**.072**	**.294**	**.580**	**.687**	**.852**	**.975**	**.991**
PVOT	.051	.122	.201	.740	.894	.934	1.00	1.00	1.00
PVOT LF	**.014**	**.061**	**.110**	**.380**	**.708**	**.805**	**.990**	**1.00**	**1.00**
PVOT ICS-1	**.015**	**.060**	**.111**	**.618**	**.848**	**.878**	**.999**	**1.00**	**1.00**
Nonidentification: $\beta_n = \beta_0 = 0$									
sup $\mathcal{T}_n$	.061	.152	.223	.751	.922	.956	1.00	1.00	1.00
aver $\mathcal{T}_n$	.054	.145	.200	.526	.765	.849	.975	.996	.999
rand $\mathcal{T}_n$	.036	.123	.184	.417	.696	.803	.025	.976	.988
rand LF	**.008**	**.047**	**.108**	**.205**	**.504**	**.655**	**.838**	**.955**	**.973**
rand ICS-1	**.008**	**.049**	**.109**	**.411**	**.653**	**.770**	**.923**	**.977**	**.989**
sup p_n	.026	.068	.123	.380	.650	.772	.850	.946	.968
sup p_n LF	**.008**	**.038**	**.079**	**.132**	**.430**	**.592**	**.728**	**.915**	**.946**
sup p_n ICS-1	**.008**	**.004**	**.081**	**.340**	**.629**	**.750**	**.842**	**.945**	**.968**

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

Table 2. Continued

	H_0: LSTAR			H_1-weak			H_1-strong		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
PVOT	.036	.145	.211	.732	.885	.930	1.00	1.00	1.00
PVOT LF	**.010**	**.058**	**.114**	**.373**	**.717**	**.806**	**.990**	**1.00**	**1.00**
PVOT ICS-1	**.010**	**.059**	**.116**	**.682**	**.853**	**.898**	**1.00**	**1.00**	**1.00**

Note: Numerical values are rejection frequency at the given level. LSTAR is Logistic STAR. Empirical power is not size-adjusted. *sup* $\mathcal{T}_n$ and *ave* $\mathcal{T}_n$ tests are based on a wild bootstrapped p-value. *rand* $\mathcal{T}_n$: $\mathcal{T}_n(\lambda)$ with uniformly randomly chosen λ on [1,5]. *sup* p_n is the supremum p-value test where p-values are computed from the chi-squared distribution. PVOT uses the chi-squared distribution. LF implies the *least favorable* p-value is used, and ICS-1 implies the type 1 *identification category selection* p-value is used with threshold $\kappa_n = \ln(\ln(n))$.

The supremum and PVOT-χ^2 tests have the largest size corrected power under the weak alternative (raw power is displayed). The robust PVOT tests perform better than the robust tests based on $\mathcal{T}_n(\lambda^*)$; PVOT-ICS performs better than PVOT-LF at $n \in \{100, 250\}$, but the two are essentially identical at $n = 500$; and PVOT-ICS approaches the size corrected power of supremum and PVOT-χ^2 tests as n increases.

Under the strong alternative, supremum, average and PVOT-ICS tests are comparable, although the average test is weaker at $n = 100$. Both $\mathcal{T}_n(\lambda^*)$-LF and $\mathcal{T}_n(\lambda^*)$ -ICS yield lower power than PVOT-LF and PVOT-ICS. Generally the LF p-value results in lower rejection rates than the ICS-1 p-value since the LF p-value is larger (hence rejection is less likely).

7.2.2. *Weak Identification.* Now consider weak and nonidentification $\beta_n = 3/n^{1/2}$ and $\beta_n = 0$. Each nonrobust test is strongly over-sized, up to an order of 3 to 5, depending on the significance level. As an example, at $n = 100$ under nonidentification $\beta_n = 0$ the rejection rates for PVOT–χ^2 are $\{.050, .140, .205\}$ at nominal sizes $(1\%, 5\%, 10\%)$, compared to $\{.015, .065, .124\}$ under strong identification. At $n = 500$ the rates are $\{.061, .148, .208\}$ and $\{.014, .055, .115\}$ under non- and strong identification. Thus $\mathcal{T}_n(\lambda)$ is strongly positively skewed relative to the $\chi^2(1)$ distribution. The remaining tests are qualitatively similar. For example, the average test based on $\int_{\Lambda_n} \mathcal{T}_n(\lambda)\mu(d\lambda)$ generates rejection frequencies $\{.057, .146, .219\}$ when $n = 100$ under weak identification. These drop to $\{.029, .125, .176\}$ when $n = 500$.

LF and ICS-1 p-values lead to correct size for both $\mathcal{T}_n(\lambda^*)$-ICS and PVOT-ICS tests. Under the alternative, however, the ICS-1 p-value leads to a power gain that reaches close to 25%. Thus, the construction of the LF p-value works well under the null, but not surprisingly weakens empirical power. The maximum gain for ICS-1 occurs under the weak alternative, with weak identification and $n = 100$: see Table 1 (middle panel, fourth column). Bold values reflect LF- and ICS-based rejection frequencies. The typical gain is 5%–10% depending on the alternative and sample size.

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

TABLE 3. STAR test rejection frequencies: Sample size $n = 500$

	H_0: LSTAR			H_1-weak			H_1-strong		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
Strong identification: $\beta_n = .3$									
sup $\mathcal{T}_n$	.029	.069	.153	.441	.590	.676	.997	.999	.999
aver $\mathcal{T}_n$	.022	.055	.120	.382	.546	.624	.988	.996	.997
rand $\mathcal{T}_n$	.008	.049	.098	.328	.488	.598	.976	.999	.996
rand LF	**.001**	**.018**	**.042**	**.227**	**.450**	**.565**	**.967**	**.989**	**.998**
rand ICS-1	**.009**	**.046**	**.096**	**.230**	**.449**	**.565**	**.974**	**.990**	**.998**
sup p_n	.005	.039	.078	.295	.457	.536	.961	.990	.997
sup p_n LF	**.002**	**.010**	**.033**	**.223**	**.427**	**.528**	**.949**	**.985**	**.997**
sup p_n ICS-1	**.005**	**.039**	**.077**	**.228**	**.432**	**.528**	**.962**	**.990**	**.997**
PVOT	.014	.055	.115	.423	.568	.655	.996	.999	.999
PVOT LF	**.002**	**.023**	**.051**	**.311**	**.509**	**.618**	**.995**	**.998**	**1.00**
PVOT ICS-1	**.013**	**.058**	**.106**	**.314**	**.510**	**.618**	**.995**	**.998**	**1.00**
Weak identification: $\beta_n = .3/\sqrt{n}$									
sup $\mathcal{T}_n$	.044	.134	.184	.984	.998	1.00	1.00	1.00	1.00
aver $\mathcal{T}_n$	.029	.125	.176	.883	.968	/989	1.00	1.00	1.00
rand $\mathcal{T}_n$	.032	.096	.162	.817	.929	.970	.995	.998	.998
rand LF	**.009**	**.051**	**.108**	**.519**	**.835**	**.914**	**.984**	**.996**	**.998**
rand ICS-1	**.009**	**.051**	**.120**	**.785**	**.921**	**.954**	**.990**	**.998**	**1.00**
sup p_n	.020	.047	.093	.721	.892	.943	.985	.998	1.00
sup p_n LF	**.015**	**.025**	**.054**	**.451**	**.772**	**.883**	**.961**	**.992**	**1.00**
sup p_n ICS-1	**.014**	**.026**	**.056**	**.710**	**.890**	**.940**	**.986**	**.998**	**1.00**
PVOT	.050	.118	.194	.981	.995	1.00	1.00	1.00	1.00
PVOT LF	**.012**	**.053**	**.109**	**.823**	**.965**	**.975**	**1.00**	**1.00**	**1.00**
PVOT ICS-1	**.012**	**.054**	**.109**	**.958**	**.987**	**.993**	**1.00**	**1.00**	**1.00**
Nonidentification: $\beta_n = \beta_0 = 0$									
sup $\mathcal{T}_n$	.051	.151	.196	.981	.998	.998	1.00	1.00	1.00
aver $\mathcal{T}_n$	.043	.136	.189	.886	.968	.984	1.00	1.00	1.00
rand $\mathcal{T}_n$	.047	.111	.177	.826	.938	.967	.997	1.00	1.00
rand LF	**.006**	**.058**	**.110**	**.549**	**.859**	**.926**	**1.00**	**1.00**	**1.00**
rand ICS-1	**.006**	**.058**	**.109**	**.827**	**.940**	**.973**	**1.00**	**1.00**	**1.00**
sup p_n	.032	.081	.126	.718	.904	.934	.995	.999	.999
sup p_n LF	**.013**	**.051**	**.085**	**.414**	**.778**	**.875**	**.965**	**.999**	**1.00**
sup p_n ICS-1	**.013**	**.051**	**.086**	**.704**	**.903**	**.934**	**.995**	**.999**	**1.00**

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

TABLE 3. Continued

	H_0: LSTAR			H_1-weak			H_1-strong		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
PVOT	.061	.148	.208	.977	.993	.996	1.00	1.00	1.00
PVOT LF	**.014**	**.058**	**.108**	**.853**	**.970**	**.989**	**1.00**	**1.00**	**1.00**
PVOT ICS-1	**.013**	**.057**	**.107**	**.978**	**.996**	**.998**	**1.00**	**1.00**	**1.00**

Note: Numerical values are rejection frequency at the given level. LSTAR is Logistic STAR. Empirical power is not size-adjusted. *sup* $\mathcal{T}_n$ and *ave* $\mathcal{T}_n$ tests are based on a wild bootstrapped p-value. *rand* $\mathcal{T}_n$: $\mathcal{T}_n(\lambda)$ with uniformly randomly chosen λ on [1,5]. *sup* $p_n$is the supremum p-value test where p-values are computed from the chi-squared distribution. PVOT uses the chi-squared distribution. LF implies the *least favorable* p-value is used, and ICS-1 implies the type 1 *identification category selection* p-value is used with threshold $\kappa_n = \ln(\ln(n))$.

The PVOT-ICS test generally has the greatest size corrected power, in particular under (*i*) the strong alternative at $n = 100$, and the 5% and 10% levels; (*ii*) the strong alternative at $n \geq 250$; and (*iii*) the weak alternative at the 5% and 10% levels, when $n \geq 250$. Under those alternatives and sample sizes the supremum and PVOT-ICS tests are comparable at the 1% level. In the remaining cases, the supremum test has the largest size corrected power with a margin of about 5%–10%.

When empirical size and size corrected power are considered jointly, PVOT-ICS is the most promising test in this study across cases, and in terms of robustness against weak and nonidentification. Average and supremum tests with wild bootstrapped p-values exhibit large size distortions under weak and nonidentification, while PVOT-ICS controls for size, and yields competitive or dominant size-corrected power.

8. CONCLUSION

We offer a new bootstrap procedure that is robust to any degree of (non) identification in nonlinear regression models. The procedure targets case specific degrees of identification, avoiding the breakdown of uniform bootstrap asymptotic validity over the parameter space. We focus on a CM test of functional form, but the method extends to a wide variety of tests. An occupation time smoothed bootstrapped p-value leads to a test that achieves uniform size control, and is consistent.

The procedure works well in a simulation experiment, in particular when the proposed bootstrapped p-value is imbedded in the p-value occupation time. Future work may include expanding the proposed bootstrap to other tests where identification is a potential problem, including tests of white noise for model residuals, structural break tests, and so on.

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

SUPPLEMENTARY MATERIAL

To view supplementary material for this article, please visit http://dx.doi.org/10.1017/S0266466620000201

References

Andrews, D.W.K. & X. Cheng (2012a) Estimation and inference with weak, semi-strong and strong identification. *Econometrica* 80, 2153–2211.

Andrews, D.W.K. & X. Cheng (2012b) *Supplement to "Estimation and Inference with Weak, Semi-Strong, and Strong Indentification"*. Department of Economics, Yale University.

Andrews, D.W.K. & X. Cheng (2013) Maximum likelihood estimation and uniform inference with sporadic identification failure. *Journal of Econometrics* 173, 36–56.

Andrews, D.W.K. & X. Cheng (2014) GMM estimation and uniform subvector inference with possible identification failure. *Econometric Theory* 30, 287–333.

Andrews, D.W.K. & P. Guggenberger (2010) Asymptotic size and a problem with subsampling and the m out of n bootstrap. *Econometric Theory* 26, 426–468.

Andrews, D.W.K., M.J. Moreira & J. Stock (2006) Optimal two-sided invariant similar tests for instrumental variables regression. *Econometrica* 74, 715–752.

Andrews, D.W.K. & W. Ploberger (1994) Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica* 62, 1383–1414.

Andrews, I. & A. Mikusheva (2016) A geometric approach to nonlinear econometric models. *Econometrica* 84, 1249–1264.

Arcones, M.A. & B. Yu (1994) Central limit theorems for empirical U-processes of stationary mixing sequences. *Journal of Theoretical Probability* 7, 47–71.

Bickel, P.J. & D.A. Freedman (1981) Some asymptotic theory for the bootstrap. *Annals of Statistics* 9, 1196–1217.

Bierens, H.J. (1982) Consistent model specification tests. *Journal of Econometrics* 20, 105–134.

Bierens, H.J. (1990) A consistent conditional moment test of functional form. *Econometrica* 58, 1443–1458.

Bierens, H.J. & W. Ploberger (1997) Asymptotic theory for integrated conditional moment tests. *Econometrica* 65, 1129–1151.

Caner, M. (2010) Testing, estimation in GMM and CUE with nearly weak identification. *Econometric Reviews* 29, 330–363.

Chan, K.S. & H. Tong (1986) On estimating thresholds in autoregressive models. *Journal of Time Series Analysis* 7, 179–190.

Cheng, X. (2015) Robust inference in nonlinear models with mixed identification strength. *Journal of Econometrics* 189, 207–228.

Chernoff, H. & S. Zacks (1964) Estimating the current mean of a normal distribution which is subject to changes in time. *Annals of Mathematical Statistics* 35, 999–1028.

Choi, I. & P.C.B. Phillips (1992) Asymptotic and finite sample distribution theory for IV estimators and tests in partially identified structural equations. *Journal of Econometrics* 51, 113–150.

Cox, G. (2016) Weak Identification in a Class of Generically Identified Models with an Application to Factor Models. Discussion Paper, Yale University.

Davidson, J. & A.G. Halunga (2014) Consistent testing of functional form in time series models. In N. Haldrup, M. Meitz & P. Saikkonen (eds.), *Essays in Nonlinear Time Series Econometrics*. Oxford University Press.

Davies, R. (1977) Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* 64, 247–254.

Davies, R. (1987) Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* 74, 33–43.

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

de Jong, R.M. (1996) On the Bierens test under data dependence. *Journal of Econometrics* 72, 1–32.

de Jong, R.M. (1997) Central limit theorems for dependent heterogeneous random variables. *Econometric Theory* 13, 353–367.

Delgado, M.A., M.A. Dominguez & P. Lavergne (2006) Consistent tests of conditional moment restrictions. *Annales d'Economie et de Statistique* 81, 33–67.

Dette, H. (1999) A consistent test for the functional form of a regression based on a difference of variance estimators. *Annals of Statistics* 27, 1012–1040.

Dudley, R.M. (1978) Central limit theorems for empirical measures. *Annals of Probability* 6, 899–929.

Dudley, R.M. (1984) A course on empirical processes. *Lecture Notes in Math, vol. 1097, chap. École d'Été de Probabilités de Saint-Flour XII—1982*, pp. 2–142. Springer, Berliln.

Dufour, J.-M. (1997) Some impossibility theorems in econometrics with applications to structural and dynamic models. *Econometrica* 65, 1365–1387.

Eberlein, E. (1984) Weak rates of convergence of partial sums of absolute regular sequences. *Statistics and Probability Letters* 2, 291–293.

Elliott, G., U. Muller & M. Watson (2015) Nearly optimal tests when a nuisance parameter is present under the null hypothesis. *Econometrica* 83, 771–811.

Gine, E. & J. Zinn (1986) Lectures on the central limit theorem for empirical processes. *Probability and Banach Spaces: Lecture Notes in Math*, vol. 1221, pp. 50–113, Berliln. Springer.

Gine, E. & J. Zinn (1990) Bootstrapping general empirical measures. *Annals of Statistics* 18, 851–869.

Gonzalez-Rivera, G. (1998) Smooth-transition GARCH models. *Studies in Nonlinear Dynamics and Econometrics* 3, 61–78.

Granger, C.W.J. & T. Terasvirta (1993) *Modelling Non-Linear Economic Relationships*. Oxford University Press.

Han, S. & A. McCloskey (2016) Estimation and Inference with a (Nearly) Singular Jacobian. Discussion Paper, Brown University.

Hansen, B.E. (1996) Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* 64, 413–430.

Härdle, W. & E. Mammen (1993) Comparing nonparametric vs parametric regression fits. *Annals Statistics* 21, 1926–1947.

Hill, J.B. (2008) Consistent and non-degenerate model specification tests against smooth transition and neural network alternatives. *Annales D' Economie et de Statistique* 90, 145–179.

Hill, J.B. (2018) A Smoothed p-Value Test When There is a Nuisance Parameter under the Alternative. Discussion Paper, Department of Economics, University of North Carolina.

Hoffman-Jørgensen, J. (1984) Convergence of Stochastic Processes on Polish Spaces, mimeo.

Hoffman-Jørgensen, J. (1991) Convergence of stochastic processes on Polish spaces. *Various Publications Series 39*. Aarhus Universitet, mimeo.

Hong, Y. & H. White (1995) Consistent specification testing via nonparametric series regression. *Econometrica* 63, 1133–1159.

Kullback, S. & R.A. Leibler (1951) On Information and sufficiency. *Annals of Mathematical Statistics* 22, 79–86.

Lehmann, E.L. (1994) *Testing Statistical Hypotheses*, 2nd edition. Chapman and Hall New York.

Li, H., Q. Li & R. Liu (2016) Consistent model specification tests based on k-nearest-neighbor estimation method. *Journal of Econometrics* 194, 187–202.

Li, Q. (1999) Consistent model specification tests for time series econometric models. *Journal of Econometrics* 92, 101–147.

Liu, R.Y. (1988) Bootstrap procedures under some NON-I.I.D. models. *Annals of Statistics* 16, 1696–1708.

Lundbergh, S. & T. Terasvirta (2006) A time series model for an exchange rate in a target zone with applications. *Journal of Econometrics* 131, 579–609.

Luukkonen, R., P. Saikkonen & T. Terasvirta (1988) Testing linearity against smooth transition autoregressive models. *Biometrika* 75, 491–499.

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

McCloskey, A. (2017) Bonferroni-based size-correction for nonstandard testing problems. *Journal of Econometrics* 200, 17–35.
Moreira, M.J. (2003) A conditional likelihood ratio test for structural models. *Econometrica* 71, 1027–1048.
Newey, W.K. (1985) Maximum likelihood specification testing and conditional moment test. *Econometrica* 53, 1047–1070.
Phillips, P.C.B. (1990) Partially identified econometric models. *Econometric Theory* 5, 181–240.
Pollard, D. (1984) *Convergence of Stochastic Processes*. Springer, New York.
Pollard, D. (1990) Empirical processes: Theory and applications. *Regional Conference Series in Probability and Statistics*, Volume 2.
Romano, J.P. (1989) Do bootstrap confidence procedures behave well uniformly in *P*? *Canadian Journal of Statistics* 17, 75–80.
Sargan, J.D. (1983) Identification and lack of identification. *Econometrica* 51, 1605–1633.
Sawa, T. (1978) Information criteria for discriminating among alternative regression models. *Econometrica* 46, 1273–1291.
Sheehy, A. & J. Wellner (1992) Uniform Donsker classes of functions. *Annals of Statistics* 20, 1983–2030.
Stinchcombe, M. & H. White (1998) Consistent specification testing with nuisance parameters present only under the alternative. *Econometric Theory* 14, 295–325.
Stock, J.H., & J. Wright (2000) GMM with weak identification. *Econometrica* 68, 1055–1096.
Tauchen, G. (1985) Diagnostic testing and evaluation of maximum likelihood models. *Journal of Econometrics* 30, 415–443.
Terasvirta, T. (1994) Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association* 89, 208 218.
Tong, H. & K.S. Lim (1980) Threshold autoregression, limit cycles and cyclical data (with discussion). *Journal of the Royal Statistical Society, Series B* 42, 245–292.
van der Vaart, A. & J. Wellner (1996) *Weak Convergence and Empirical Processes*. Springer, New York.
Vapnik, V.K. & A.Y. Červonenkis (1971) On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and its Applications* 16, 264–280.
Whang, Y.-J. (2000) Consistent bootstrap tests of parametric regression functions. *Journal of Econometrics* 98, 27–46.
White, H. (1982) Maximum likelihood estimation of misspecified models. *Econometrica* 50, 1–25.
White, H. (1989) An additional hidden unit test for neglected nonlinearity in multilayer feedforward networks. *Proceeding of the International Joint Conference on Neural Networks, II*. pp. 451–455, Washington, D.C. New York, NY. IEEE Press.
Wu, C.F.J. (1986) Jackknife, bootstrap and other subsampling methods in regression analysis. *Annals of Statistics* 14, 1261–1295.
Zheng, J.X. (1996) A consistent test of functional form via nonparametric estimation techniques. *Journal of Econometrics* 75, 263–289.

A APPENDIX

A.1. Assumption 1.d–f

Assumptions 1.d–f contain technical restrictions on long-run variances, and parameter space details that are useful when any degree of identification is allowed.

In order to conserve space below, we use the following notation. Let $int(\mathcal{M})$ denote the interior of set $\mathcal{M}$. Write compactly $\inf_{\alpha,r,\theta} = \inf_{\alpha'\alpha=1,r'r=1,\{\theta_i\}_{i=1}^m \in \Theta^m}$, $\inf_{\alpha,\pi,\lambda} = \inf_{\alpha'\alpha=1,\{\pi_i,\lambda_i\}_{i=1}^m \in (\Pi\times\Lambda)^m}$, $\inf_{r,\omega,\pi} = \inf_{r'r=1,\omega\in\mathbb{R}^{k_\beta}:||\omega||=1,\pi\in\Pi}$, and so on. $m \in \mathbb{N}$ and

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

$a \in \mathbb{R}^m$ are arbitrary; $\Theta^m \equiv \Theta \times \cdots \times \Theta \subset \mathbb{R}^m$. $r = [r_1, r_2']'$, with $r_1 \in \mathbb{R}$, is an arbitrary vector whose dimension is implicitly defined. We need the following definitions:

$$\begin{aligned}\mathbb{E}_{\psi,n}(\pi;a,r) &\equiv \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\epsilon_t\sum_{i=1}^{m}\alpha_i r' d_{\psi,t}(\pi_i) \text{ and } \mathbb{E}_{\theta,n}(\omega,\pi;a,r)\\ &\equiv \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\epsilon_t\sum_{i=1}^{m}\alpha_i r' d_{\theta,t}(\omega_i,\pi_i)\\ \mathfrak{EG}_{\psi,n}(\lambda;a,r) &\equiv r_1\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\sum_{i=1}^{m}\alpha_i\left\{\epsilon_t(\psi_n,\pi_i)\mathcal{K}_{\psi,t}(\pi_i,\lambda_i)\right.\\ &\quad\left.-E\left[\epsilon_t(\psi_n,\pi_i)\mathcal{K}_{\psi,t}(\pi_i,\lambda_i)\right]\right\}+r_2'\sum_{i=1}^{m}\alpha_i\mathcal{G}_{\psi,n}(\psi_n,\pi_i).\end{aligned}$$

Assumption 1 (Data generating process, test weight)

d. *Long-run variances*:

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

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

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

(iv) $\liminf_{n\to\infty}\inf_{a,r,\pi} E[\mathbb{E}_{\psi,n}(\pi;a,r)^2] > 0$ and $\limsup_{n\to\infty}\sup_{a,r,\pi} E[\mathbb{E}_{\psi,n}(\pi;a,r)^2] < \infty$; and $\liminf_{n\to\infty}\inf_{a,r,\omega,\pi} E[\mathbb{E}_{\theta,n}(\omega,\pi;a,r)^2] > 0$ and $\limsup_{n\to\infty}\sup_{a,r,\omega,\pi} E[\mathbb{E}_{\theta,n}(\omega,\pi;a,r)^2] < \infty$.

(v) Under $\mathcal{C}(i,b)$ with $||b|| < \infty$, $\liminf_{n\to\infty} E[\sup_{\alpha,r,\lambda}\mathfrak{EG}_{\psi,n}(\lambda;a,r)^2] < \infty$.

(vi) Under $\mathcal{C}(ii,\omega_0)$, $E[\sup_{\alpha,r,\lambda}(1/\sqrt{n}\sum_{t=1}^{n}\epsilon_t\mathcal{K}_{\theta,t}(\lambda;a,m))^2] < \infty$ 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) $0_{k_\beta} \in \text{int}\,(\mathcal{B}^*)$.

(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 β.

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

Remark 11. (d.i,ii) are standard for nondegenerate finite dimensional asymptotics for the least squares first-order equations, under stationarity. (d.iii,iv) likewise imply the components of the least-squares asymptotic variance are nondegenerate and positive definite. Each is standard for ensuring nondegenerate asymptotics. (d.v) promotes a joint weak

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

limit theory for the test statistic numerator $1/\sqrt{n}\sum_{t=1}^{n}\epsilon_t(\psi_n,\pi)F(\lambda'\mathcal{W}(x_t))$ and $\hat{\pi}_n$ under weak identification, which in turn leads to a limit theory for $1/\sqrt{n}\sum_{t=1}^{n}\epsilon_t(\hat{\theta}_n)F(\lambda'\mathcal{W}(x_t))$. (d.vi) similarly covers $1/\sqrt{n}\sum_{t=1}^{n}\epsilon_t(\hat{\theta}_n)F(\lambda'\mathcal{W}(x_t))$ under strong identification. In the latter cases (d.v) and (d.vi), we cannot assume strict positivity due to possible degeneracy: see the discussion leading to Assumptions 3 and 4.

Remark 12. (e.ii) ensures nonidentification ($\beta = 0$) and near nonidentification (β close to 0) points are in $\mathcal{B}^*$. (e.iii) ensures the true space $\mathcal{Z}^*(\beta)$ is bounded from the empty set for values of β near the nonidentification point $\beta = 0$, and therefore allows for derivatives of certain moments with respect to β near $\beta = 0$ (cf. Andrews and Cheng, 2013, comments following Assumptions B2(iii) and STAR4(iv)).[17] The (f.i) property $\Theta^* \subset \text{int}(\Theta)$ implies the true value does not lie on the boundary of the optimization space. This is nonessential, but allows a focus on weak identification.

A.2. Proofs of Main Results

We assume all random variables exist on a complete measure space such that majorants and integrals over uncountable families of measurable functions are measurable, and probabilities where applicable are outer probability measures.[18]

In order to conserve space, write processes on compact spaces variously as $\{f(a,b,c) : A,B,C\} = \{f(a,b,c) : a \in A, b \in B, c \in C\}$. $A_n(\lambda) = o_{p,\lambda}(1)$ implies $\sup_{\lambda\in\Lambda}||A_n(\lambda)|| \xrightarrow{p} 0$. All Gaussian processes below have a version that has *almost surely* continuous and uniformly bounded sample paths, hence we just say *Gaussian process*.

The following proofs require supporting results presented in Appendix B and proven in the supplemental material Hill (2020, Appendix B). Recall $\omega(\beta) \equiv \{\beta/||\beta|| \text{ if } \beta \neq 0, 1_{k_\beta}/||1_{k_\beta}|| \text{ if } \beta = 0\}$.

Recall the augmented parameter $\theta^+ \equiv [||\beta||, \omega', \zeta', \pi']' \in \Theta^+$ where $\Theta^+ \equiv \{\theta^+ \in \mathbb{R}^{k_\beta+k_x+k_\pi+1} : \theta^+ = [||\beta||, \omega(\beta), \zeta, \pi]' : \beta \in \mathcal{B}, \zeta \in \mathcal{Z}(\beta), \pi \in \Pi\}$.

Proof of Theorem 4.2.

Claim a. Let drift case $\mathcal{C}(i,b)$ hold with $||b|| < \infty$. Recall $\epsilon_t(\theta) = \epsilon_t(\psi,\pi)$, and define $f(x_t,\theta) = \zeta' x_t + \beta' g(x_t,\pi)$.

Step 1. We prove the following expansion:

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\epsilon_t(\hat{\theta}_n)F\left(\lambda'\mathcal{W}(x_t)\right) \tag{A.1}$$

$$= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\{\epsilon_t(\psi_n,\hat{\pi}_n)\mathcal{K}_{\psi,t}(\hat{\pi}_n,\lambda) - E\left[\epsilon_t(\psi_n,\hat{\pi}_n)\mathcal{K}_{\psi,t}(\hat{\pi}_n,\lambda)\right]\right\}$$

$$+ E\left[\mathcal{K}_{\psi,t}(\hat{\pi}_n,\lambda)\left\{g(x_t,\pi_0) - g(x_t,\hat{\pi}_n)\right\}'\right]b$$

$$+ \mathfrak{b}_\psi(\hat{\pi}_n,\lambda)'\mathcal{H}_\psi^{-1}(\hat{\pi}_n)E\left[d_{\psi,t}(\hat{\pi}_n)\left\{g(x_t,\pi_0) - g(x_t,\hat{\pi}_n)\right\}'\right]b$$

[17] In the Logistic STAR model in the preceding footnote it suffices to assume compact $\mathcal{B}^* \subset (-\infty,\infty)\times(-1+\iota, 1-\iota)$ for some infinitesimal $\iota > 0$, and $\mathcal{Z}^*(\beta)$ is a compact subset of $\{\zeta \in \mathbb{R}^2 : \zeta_1 \in (-\infty,\infty), -1 < \zeta_2 + \beta_2 < 1\}$. Now assume $\mathcal{Z}_0^* = \{\zeta \in \mathbb{R}^2 : \zeta_1 \in (-\infty,\infty), -1+\iota < \zeta_2 < 1-\iota\}$ and pick $\delta = \iota$. The same idea extends to $\mathcal{Z}_0$.

[18] See Pollard's (1984, Appendix C) *permissibility* criteria, and see Dudley's (1986, p. 101) *admissible Suslin* property.

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

$$
\begin{aligned}
&+ \mathfrak{b}_\psi(\hat{\pi}_n,\lambda)' \left\{ \mathcal{H}_\psi^{-1}(\hat{\pi}_n)\mathcal{D}_\psi(\hat{\pi}_n)b + \left[b, 0'_{k_\beta}\right]' \right\} + o_{p,\lambda}(1) \\
&= \mathfrak{Z}_n(\hat{\pi}_n,\lambda) + \mathcal{R}(\hat{\pi}_n,\lambda) + o_{p,\lambda}(1),
\end{aligned}
$$

where $\mathcal{R}(\pi,\lambda)$ is implicitly defined, and

$$
\mathfrak{Z}_n(\pi,\lambda) \equiv \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left\{ \epsilon_t(\psi_n,\pi)\mathcal{K}_{\psi,t}(\pi,\lambda) - E\left[\epsilon_t(\psi_n,\pi)\mathcal{K}_{\psi,t}(\pi,\lambda)\right] \right\}.
$$

Recall $\hat{\theta}_n = [\hat{\psi}_n(\hat{\pi}_n)', \hat{\pi}_n']'$ and write $\hat{\psi}_n = \hat{\psi}_n(\hat{\pi}_n)$. By the mean value theorem, there exists $\psi_n^* \in \Psi$, $||\psi_n^* - \psi_n|| \leq ||\hat{\psi}_n - \psi_n||$, such that:

$$
\begin{aligned}
&\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \epsilon_t(\hat{\theta}_n) F\left(\lambda'\mathcal{W}(x_t)\right) \\
&\quad = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \epsilon_t(\psi_n,\hat{\pi}_n) F\left(\lambda'\mathcal{W}(x_t)\right) \\
&- \frac{1}{n} \sum_{t=1}^{n} F\left(\lambda'\mathcal{W}(x_t)\right) \frac{\partial}{\partial \psi'} f(x_{t,}, \left[\psi_n^*, \hat{\pi}_n\right]) \sqrt{n}\left(\hat{\psi}_n - \psi_n\right) \\
&\quad = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \epsilon_t(\psi_n,\hat{\pi}_n) F\left(\lambda'\mathcal{W}(x_t)\right) - \hat{\mathfrak{b}}_{\psi,n}(\hat{\pi}_n,\lambda)' \sqrt{n}\left(\hat{\psi}_n - \psi_n\right).
\end{aligned}
$$

By Lemma B.10, $\sup_{\pi\in\Pi,\lambda\in\Lambda} ||\hat{\mathfrak{b}}_{\psi,n}(\pi,\lambda) - \mathfrak{b}_\psi(\pi,\lambda)|| \xrightarrow{p} 0$. The proof of Theorem 4.1 verifies (see (C.19) in Hill, 2020):

$$
\sup_{\pi\in\Pi} \left\| \sqrt{n}\left(\hat{\psi}_n(\pi) - \psi_n\right) - \left(-\mathcal{H}_\psi^{-1}(\pi) \left\{ \mathcal{G}_{\psi,n}(\psi_{0,n},\pi) + \mathcal{D}_\psi(\pi)b \right\} - \left[b, 0'_{k_\beta}\right]' \right) \right\| \xrightarrow{p} 0. \tag{A.2}
$$

Now apply (A.2) for $\sqrt{n}(\hat{\psi}_n - \psi_n)$ and the Lemma A.2 uniform consistency of $\widehat{\mathcal{H}}_{\psi,n}(\pi)$ to yield:

$$
\begin{aligned}
&\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \epsilon_t(\hat{\theta}_n) F\left(\lambda'\mathcal{W}(x_t)\right) \\
&\quad = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \epsilon_t(\psi_n,\hat{\pi}_n) F\left(\lambda'\mathcal{W}(x_t)\right) \\
&\qquad - \mathfrak{b}_\psi(\hat{\pi}_n,\lambda)' \left\{ -\mathcal{H}_\psi^{-1}(\hat{\pi}_n) \left\{ \mathcal{G}_{\psi,n}(\psi_{0,n},\hat{\pi}_n) + \mathcal{D}_\psi(\hat{\pi}_n)b \right\} - \left[b, 0'_{k_\beta}\right]' \right\} + o_{p,\lambda}(1) \\
&\quad = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \epsilon_t(\psi_n,\hat{\pi}_n) F\left(\lambda'\mathcal{W}(x_t)\right) + \mathfrak{b}_\psi(\hat{\pi}_n,\lambda)' \mathcal{H}_\psi^{-1}(\hat{\pi}_n) \mathcal{G}_{\psi,n}(\psi_{0,n},\hat{\pi}_n) \\
&\qquad + \mathfrak{b}_\psi(\hat{\pi}_n,\lambda)' \left\{ \mathcal{H}_\psi^{-1}(\hat{\pi}_n)\mathcal{D}_\psi(\hat{\pi}_n)b + \left[b, 0'_{k_\beta}\right]' \right\} + o_{p,\lambda}(1).
\end{aligned}
$$

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

Next, by the construction of $\mathcal{G}_{\psi,n}(\theta)$ in (4):

$$
\begin{aligned}
-\mathcal{G}_{\psi,n}(\psi_{0,n},\hat{\pi}_n) &= -\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\{\epsilon_t(\psi_{0,n},\hat{\pi}_n)d_{\psi,t}(\hat{\pi}_n) - E\left[\epsilon_t(\psi_{0,n},\hat{\pi}_n)d_{\psi,t}(\hat{\pi}_n)\right]\right\} \\
&= -\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\{\epsilon_t(\psi_n,\hat{\pi}_n)d_{\psi,t}(\hat{\pi}_n) - E\left[\epsilon_t(\psi_n,\hat{\pi}_n)d_{\psi,t}(\hat{\pi}_n)\right]\right\} \\
&\quad + \sqrt{n}\frac{1}{n}\sum_{t=1}^{n}\left\{d_{\psi,t}(\hat{\pi}_n)g(x_t,\hat{\pi}_n)' - E\left[d_{\psi,t}(\hat{\pi}_n)g(x_t,\hat{\pi}_n)'\right]\right\}\beta_n.
\end{aligned}
$$

Combine $\sqrt{n}||\beta_n|| \to [0,\infty)$, Lemma A.2, and Theorem 4.1 to yield: $\sqrt{n}n^{-1}\sum_{t=1}^{n}\{d_{\psi,t}(\hat{\pi}_n)g(x_t,\hat{\pi}_n)' - E[d_{\psi,t}(\hat{\pi}_n)g(x_t,\hat{\pi}_n)']\}\beta_n \xrightarrow{p} 0$. Therefore,

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\epsilon_t(\hat{\theta}_n)F\left(\lambda'\mathcal{W}(x_t)\right) && \textbf{(A.3)} \\
&\quad = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\epsilon_t(\psi_n,\hat{\pi}_n)F\left(\lambda'\mathcal{W}(x_t)\right) \\
&\qquad - \mathfrak{b}_\psi(\hat{\pi}_n,\lambda)'\mathcal{H}_\psi^{-1}(\hat{\pi}_n)\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\{\epsilon_t(\psi_n,\hat{\pi}_n)d_{\psi,t}(\hat{\pi}_n) - E\left[\epsilon_t(\psi_n,\hat{\pi}_n)d_{\psi,t}(\hat{\pi}_n)\right]\right\} \\
&\qquad + \mathfrak{b}_\psi(\hat{\pi}_n,\lambda)'\left\{\mathcal{H}_\psi^{-1}(\hat{\pi}_n)\mathcal{D}_\psi(\hat{\pi}_n)b + \left[b,0'_{k_\beta}\right]'\right\} + o_{p,\lambda}(1) \\
&\quad = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\epsilon_t(\psi_n,\hat{\pi}_n)\mathcal{K}_{\psi,t}(\hat{\pi}_n,\lambda) && \textbf{(A.4)} \\
&\qquad + \mathfrak{b}_\psi(\hat{\pi}_n,\lambda)'\mathcal{H}_\psi^{-1}(\hat{\pi}_n)\sqrt{n}E\left[\epsilon_t(\psi_n,\hat{\pi}_n)d_{\psi,t}(\hat{\pi}_n)\right] && \textbf{(A.5)} \\
&\qquad + \mathfrak{b}_\psi(\hat{\pi}_n,\lambda)'\left\{\mathcal{H}_\psi^{-1}(\hat{\pi}_n)\mathcal{D}_\psi(\hat{\pi}_n)b + \left[b,0'_{k_\beta}\right]'\right\} + o_{p,\lambda}(1).
\end{aligned}
$$

By adding and subtracting $E[\epsilon_t(\psi_n,\hat{\pi}_n)\mathcal{K}_{\psi,t}(\hat{\pi}_n,\lambda)]$, summand (A.4) satisfies:

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\epsilon_t(\psi_n,\hat{\pi}_n)\mathcal{K}_{\psi,t}(\hat{\pi}_n,\lambda) && \textbf{(A.6)} \\
&\quad = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\{\epsilon_t(\psi_n,\hat{\pi}_n)\mathcal{K}_{\psi,t}(\hat{\pi}_n,\lambda) - E\left[\epsilon_t(\psi_n,\hat{\pi}_n)\mathcal{K}_{\psi,t}(\hat{\pi}_n,\lambda)\right]\right\} \\
&+ \sqrt{n}E\left[\epsilon_t(\psi_n,\hat{\pi}_n)\mathcal{K}_{\psi,t}(\hat{\pi}_n,\lambda)\right] \\
&\quad = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\{\epsilon_t(\psi_n,\hat{\pi}_n)\mathcal{K}_{\psi,t}(\hat{\pi}_n,\lambda) - E\left[\epsilon_t(\psi_n,\hat{\pi}_n)\mathcal{K}_{\psi,t}(\hat{\pi}_n,\lambda)\right]\right\} \\
&\qquad + \sqrt{n}E\left[\epsilon_t\mathcal{K}_{\psi,t}(\hat{\pi}_n,\lambda)\right] + E\left[\mathcal{K}_{\psi,t}(\hat{\pi}_n,\lambda)\left\{g(x_t,\pi_0) - g(x_t,\hat{\pi}_n)\right\}'\right]\sqrt{n}\beta_n.
\end{aligned}
$$

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

Under H_0, trivially $\sup_{\pi\in\Pi}||E[\epsilon_t\mathcal{K}_{\psi,t}(\pi,\lambda)]|| = 0$. Turning to the expectations in (A.5):

$$E\left[\epsilon_t(\psi_n,\hat{\pi}_n)d_{\psi,t}(\hat{\pi}_n)\right] = E\left[\epsilon_t d_{\psi,t}(\hat{\pi}_n)\right] + E\left[d_{\psi,t}(\hat{\pi}_n)\left\{g(x_t,\pi_0)-g(x_t,\hat{\pi}_n)\right\}'\right]\beta_n \tag{A.7}$$

$$= E\left[d_{\psi,t}(\hat{\pi}_n)\left\{g(x_t,\pi_0)-g(x_t,\hat{\pi}_n)\right\}'\right]\beta_n.$$

Combine (A.3), (A.6), and (A.7) with $\sqrt{n}\beta_n \to b$, $||b|| < \infty$, to arrive at (A.1).

Step 2. We will show $\{1/\sqrt{n}\sum_{t=1}^n \epsilon_t(\hat{\theta}_n)F(\lambda'\mathcal{W}(x_t)) : \Lambda\} \Rightarrow^* \{\mathfrak{Z}_\psi(\pi^*(b),\lambda)+\mathcal{R}(\pi^*(b),\lambda) : \Lambda\}$. Lemma A.8.a states $\{\mathfrak{Z}_n(\pi,\lambda) : \Pi,\Lambda\} \Rightarrow^* \{\mathfrak{Z}_\psi(\pi,\lambda) : \Pi,\Lambda\}$, a zero-mean Gaussian process, and by Theorem 4.1.a $\hat{\pi}_n \xrightarrow{d} \pi^*(b)$ where $\pi^*(b)$ is defined by Assumption 2. Step 3 proves joint weak convergence

$$\left\{\mathfrak{Z}_n(\pi,\lambda),\hat{\pi}_n : \pi\in\Pi,\lambda\in\Lambda\right\} \Rightarrow^* \left\{\mathfrak{Z}_\psi(\pi,\lambda),\pi^*(b) : \pi\in\Pi,\lambda\in\Lambda\right\}. \tag{A.8}$$

The mapping theorem and expansion (A.1) deliver the desired result.

Step 3. We need to show (A.8). By the proof of Theorem 4.1.a, $\hat{\pi}_n$ is a continuous function of $\mathcal{G}_{\psi,n}(\psi_{0,n},\pi)$ and $\widehat{\mathcal{H}}_{\psi,n}(\pi)$. Furthermore, $\{\mathcal{G}_{\psi,n}(\psi_{0,n},\pi) : \Pi\} \Rightarrow^* \{\mathcal{G}_\psi(\pi) : \Pi\}$ by Lemma A.1, and $\widehat{\mathcal{H}}_{\psi,n}(\pi)$ has a nonrandom limit uniformly on Π by Lemma A.2. Therefore, (A.8) follows from the mapping theorem and Cramé r's theorem provided jointly:

$$\left\{\begin{bmatrix}\mathfrak{Z}_n(\pi,\lambda)\\ \mathcal{G}_{\psi,n}(\psi_{0,n},\pi)\end{bmatrix} : \Pi,\Lambda\right\}$$

$$= \left\{\begin{bmatrix}\frac{1}{\sqrt{n}}\sum_{t=1}^n\left\{\epsilon_t(\pi,\lambda)\mathcal{K}_{\psi,t}(\pi,\lambda) - E\left[\epsilon_t(\psi_n,\pi)\mathcal{K}_{\psi,t}(\pi,\lambda)\right]\right\}\\ -\frac{1}{\sqrt{n}}\sum_{t=1}^n\left\{\epsilon_t(\psi,\pi)d_{\psi,t}(\pi) - E\left[\epsilon_t(\psi,\pi)d_{\psi,t}(\pi)\right]\right\}\end{bmatrix} : \Pi,\Lambda\right\}$$

$$\Rightarrow^* \left\{\begin{bmatrix}\mathfrak{Z}_\psi(\pi,\lambda)\\ \mathcal{G}_\psi(\pi)\end{bmatrix} : \Pi,\Lambda\right\}.$$

The latter holds by the same arguments used to prove Lemmas A.1 and A.8, hence we only provide a sketch of the proof. First, $[\mathfrak{Z}_n(\pi,\lambda),\mathcal{G}_{\psi,n}(\psi_{0,n},\pi)']'$ converges in finite dimensional distributions over $\Pi\times\Lambda$ to a zero mean, finite variance Gaussian random vector. This follows because linear combinations $\sum_{i=1}^m a_i\{r_1\mathfrak{Z}_n(\pi_i,\lambda_i)+r_2'\mathcal{G}_{\psi,n}(\psi_{0,n},\pi_i)\}$ for any $m\in\mathbb{N}$ and $a\in\mathbb{R}^m$ with $a'a=1$, and any $r=[r_1,r_2']'$, $r'r=1$, satisfy a Gaussian central theorem under the moment and memory properties of Assumption 1.b–d(vi). Second, $[\mathfrak{Z}_n(\pi,\lambda),\mathcal{G}_{\psi,n}(\psi_{0,n},\pi)']'$ is stochastically equicontinuous because, by probability subadditivity, we require $\mathfrak{Z}_n(\pi,\lambda)$ and each element $\mathcal{G}_{\psi,n,i}(\psi_{0,n},\pi)$ of $\mathcal{G}_{\psi,n}(\psi_{0,n},\pi) = [\mathcal{G}_{\psi,n,i}(\psi_{0,n},\pi)]_{i=1}^{k_x+k_\beta}$ to be stochastically equicontinuous, and these properties are established in the proofs of Lemmas A.1 and A.8.

Step 4. We now tackle $\hat{v}_n(\hat{\theta}_n,\lambda)$ and complete the proof. $\hat{v}_n(\hat{\theta}_n,\lambda)$ is a function of $\hat{\mathfrak{b}}_{\theta,n}(\omega(\hat{\beta}_n),\hat{\pi}_n,\lambda)$ and $d_{\theta,t}(\omega(\hat{\beta}_n),\hat{\pi}_n)$. By Lemma A.10.a, $\sup_{\theta^+\in\Theta^+,\lambda\in\Lambda}||\hat{v}_n^2(\theta^+,\lambda) - v^2(\theta^+,\lambda)|| \xrightarrow{p} 0$. Furthermore,

$$\omega(\hat{\beta}_n(\hat{\pi}_n)) = \frac{\sqrt{n}\mathcal{S}_\beta\hat{\psi}_n(\hat{\pi}_n)}{\left\|\sqrt{n}\mathcal{S}_\beta\hat{\psi}_n(\hat{\pi}_n)\right\|} = \frac{\sqrt{n}\mathcal{S}_\beta\left(\hat{\psi}_n(\hat{\pi}_n)-\psi_n\right)+\sqrt{n}\beta_n}{\left\|\sqrt{n}\mathcal{S}_\beta\left(\hat{\psi}_n(\hat{\pi}_n)-\psi_n\right)+\sqrt{n}\beta_n\right\|} \equiv \omega_n(\hat{\pi}_n), \tag{A.9}$$

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

by construction of $\mathcal{S}_\beta$. Notice ω_n is an implicitly defined stochastic function of $\hat{\pi}_n$. Now, by Theorem 4.1.a, the mapping theorem and $\sqrt{n}\beta_n \to b$, $||b|| < \infty$:

$$\omega_n(\hat{\pi}_n) \xrightarrow{d} \frac{\mathcal{S}_\beta \tau(\pi^*(b), b)}{\left\|\mathcal{S}_\beta \tau(\pi^*(b), b)\right\|} = \frac{\tau_\beta(\pi^*(b), b)}{\left\|\tau_\beta(\pi^*(b), b)\right\|}.$$

Joint weak convergence for $\sqrt{n}(\hat{\psi}_n(\pi) - \psi_n), \hat{\pi}_n$ and $\omega_n(\hat{\pi}_n)$ follows from arguments in the proof of Theorem 4.1.a because $\omega_n(\pi)$ is a continuous function of $\sqrt{n}(\hat{\psi}_n(\pi) - \psi_n)$, and $\sqrt{n}(\hat{\psi}_n(\pi) - \psi_n)$ and $\hat{\pi}_n$ are continuous functions of $\mathcal{G}_{\psi,n}(\psi_{0,n}, \pi)$ and $\widehat{\mathcal{H}}_{\psi,n}(\pi)$. Hence,

$$\left[\sqrt{n}\left(\hat{\psi}_n(\hat{\pi}_n) - \psi_n\right)', \hat{\pi}_n', \omega_n(\hat{\pi}_n)'\right]' \xrightarrow{d} \left[\tau(\pi^*(b), b)', \pi^*(b)', \frac{\tau_\beta(\pi^*(b), b)'}{\left\|\tau_\beta(\pi^*(b), b)\right\|}\right]'. \tag{A.10}$$

Using $v^2(\cdot)$ defined in (11), we may therefore write:

$$\hat{v}_n^2(\hat{\theta}_n, \lambda) = v^2(\omega_n(\hat{\pi}_n), \hat{\pi}_n, \lambda) + o_{p,\lambda}(1) \text{ where } v^2(\omega, \pi, \lambda) = v^2(\left[\beta_0, \omega, \zeta_0, \pi\right], \lambda), \tag{A.11}$$

and $\liminf_{n\to\infty} v^2(\omega_n(\hat{\pi}_n), \hat{\pi}_n, \lambda) > 0$ a.s. $\forall \lambda \in \Lambda$ by Assumption 4. The claim now follows from Step 2, $\sup_{\theta^+ \in \Theta^+, \lambda \in \Lambda} ||\hat{v}_n^2(\theta^+, \lambda) - v^2(\theta^+, \lambda)|| \xrightarrow{p} 0$, (A.10), (A.11), and the mapping theorem:

$$\begin{aligned}\{\mathcal{T}_n(\lambda) : \Lambda\} &= \left\{\frac{\left(\mathfrak{Z}_n(\hat{\pi}_n, \lambda) + \mathcal{R}(\hat{\pi}_n, \lambda)\right)^2}{v^2(\omega_n(\hat{\pi}_n), \hat{\pi}_n, \lambda)} + o_{p,\lambda}(1) : \Lambda\right\} \\ &\Rightarrow^* \left\{\frac{\left(\mathfrak{Z}(\pi^*(b), \lambda) + \mathcal{R}(\pi^*(b), \lambda)\right)^2}{v^2(\omega(\pi^*(b)), \pi^*(b), \lambda)} : \Lambda\right\}.\end{aligned} \tag{A.12}$$

Claim b. Let $\mathcal{C}(ii, \omega_0)$ apply. A first-order expansion yields for some midpoint θ_n^*, $||\theta_n^* - \theta_n|| \leq ||\hat{\theta}_n - \theta_n||$:

$$\begin{aligned}&\frac{1}{\sqrt{n}}\sum_{t=1}^n \epsilon_t(\hat{\theta}_n) F\left(\lambda' \mathcal{W}(x_t)\right) \\ &= \frac{1}{\sqrt{n}}\sum_{t=1}^n \epsilon_t F\left(\lambda' \mathcal{W}(x_t)\right) - \frac{1}{n}\sum_{t=1}^n F\left(\lambda' \mathcal{W}(x_t)\right) \frac{\partial}{\partial \theta'} f(x_{t,}, \theta_n^*) \sqrt{n}\left(\hat{\theta}_n - \theta_n\right) \\ &= \frac{1}{\sqrt{n}}\sum_{t=1}^n \epsilon_t F\left(\lambda' \mathcal{W}(x_t)\right) - \hat{\mathfrak{b}}_{\theta,n}(\beta_n^* / \left\|\beta_n^*\right\|, \pi_n^*, \lambda)' \sqrt{n} \mathfrak{B}(\beta_n)\left(\hat{\theta}_n - \theta_n\right).\end{aligned}$$

The proof of Theorem 4.1.b shows (see (C.21) in Hill, 2020):

$$\begin{aligned}\sqrt{n}\mathfrak{B}(\beta_n)\left(\hat{\theta}_n - \theta_n\right) &= \widehat{\mathcal{H}}_n^{-1}(\omega(\beta_n^*), \pi_n^*) \mathfrak{B}(\beta_n)^{-1} \sqrt{n} \frac{\partial}{\partial \theta} Q_n(\theta_n) \\ &= \widehat{\mathcal{H}}_n^{-1}(\omega(\beta_n^*), \pi_n^*) \mathfrak{B}(\beta_n)^{-1} \mathcal{G}_{\theta,n}(\theta_n).\end{aligned}$$

By Lemma A.5, $\sup_{\omega \in \mathbb{R}^{k_\beta} : ||\omega|| = 1, \pi \in \Pi} ||\widehat{\mathcal{H}}_n(\omega, \pi) - \mathcal{H}_\theta(\omega, \pi)|| \xrightarrow{p} 0$, and $\hat{\theta}_n \xrightarrow{p} \theta_0$ by the proof of Theorem 4.1. Thus, by definition of $\mathcal{G}_{\theta,n}(\cdot)$:

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

$$\sqrt{n}\mathfrak{B}(\beta_n)\left(\hat{\theta}_n-\theta_n\right)=\mathcal{H}_\theta^{-1}\mathfrak{B}(\beta_n)^{-1}\mathcal{G}_{\theta,n}(\theta_n)+o_p(1)$$
$$=-\mathcal{H}_\theta^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\epsilon_t d_{\theta,t}(\beta_n/\|\beta_n\|,\pi_0)+o_p(1).$$

Moreover, $\sup_{\lambda\in\Lambda}||\hat{\mathfrak{b}}_{\theta,n}(\beta_n^*/||\beta_n^*||,\hat{\pi}_n,\lambda)-\mathfrak{b}_\theta(\omega_0,\pi_0,\lambda)||\xrightarrow{p}0$ by Lemma A.9 and $\hat{\theta}_n\xrightarrow{p}\theta_0$. Combined, we obtain,

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\epsilon_t(\hat{\theta}_n)F\left(\lambda'\mathcal{W}(x_t)\right) \tag{A.13}$$
$$=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\epsilon_t F\left(\lambda'\mathcal{W}(x_t)\right)+\mathfrak{b}_\theta(\omega_0,\pi_0,\lambda)'\mathcal{H}_\theta^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\epsilon_t d_{\theta,t}(\beta_n/\|\beta_n\|,\pi_0)+o_{p,\lambda}(1)$$
$$=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\epsilon_t\left(F\left(\lambda'\mathcal{W}(x_t)\right)+\mathfrak{b}_\theta(\omega_0,\pi_0,\lambda)'\mathcal{H}_\theta^{-1}d_{\theta,t}(\beta_n/\|\beta_n\|,\pi_0)\right)+o_{p,\lambda}(1).$$

By Lemma A.8.b, therefore, $\{1/\sqrt{n}\sum_{t=1}^{n}\epsilon_t(\hat{\theta}_n)F\left(\lambda'\mathcal{W}(x_t)\right):\Lambda\}\Rightarrow^*\{\mathfrak{Z}_\theta(\lambda):\Lambda\}$, a zero-mean Gaussian process with covariance kernel $E[\mathfrak{Z}_\theta(\lambda)\mathfrak{Z}_\theta(\tilde{\lambda})]=E[\epsilon_t^2\mathcal{K}_{\theta,t}(\lambda)\mathcal{K}_{\theta,t}(\tilde{\lambda})]$ where $\mathcal{K}_{\theta,t}(\lambda)\equiv F\left(\lambda'\mathcal{W}(x_t)\right)-\mathfrak{b}_\theta(\omega_0,\pi_0,\lambda)'\mathcal{H}_\theta^{-1}d_{\theta,t}(\omega_0,\pi_0)$.

Now turn to $\hat{v}_n(\hat{\theta}_n,\lambda)$. By Lemma A.10.b and $\hat{\theta}_n\xrightarrow{p}\theta_0$, $\sup_{\lambda\in\Lambda}||\hat{v}_n^2(\hat{\theta}_n,\lambda)-v^2(\theta_0,\lambda)||\xrightarrow{p}0$ where by construction and Assumption 1.d(vi) $v^2(\theta_0,\lambda)=E[\epsilon_t^2\mathcal{K}_{\theta,t}^2(\lambda)]<\infty$. By Assumption 4, $v^2(\theta_0,\lambda)>0\forall\lambda\in\Lambda$. Since $E[\mathfrak{Z}_\theta^2(\lambda)]=E[\epsilon_t^2\mathcal{K}_{\theta,t}^2(\lambda)]$, the proof is complete by the mapping theorem. $\mathcal{QED}$.

Proof of Theorem 4.3. By the proof of Theorem of 4.2, $\hat{v}_n^2(\hat{\theta}_n,\lambda)\to(0,\infty)$ asymptotically with probability approaching one under any identification case, and $\forall\lambda\in\Lambda$.

It remains to prove $|1/n\sum_{t=1}^{n}\epsilon_t(\hat{\theta}_n)F(\lambda'\mathcal{W}(x_t))|\xrightarrow{p}(0,\infty)\forall\lambda\in\Lambda/S$ where $S\subset\Lambda$ has Lebesgue measure zero. Consider identification case $\mathcal{C}(i,b)$ with $||b||<\infty$. Arguments in the proof of Theorem 4.2.a imply

$$\frac{1}{n}\sum_{t=1}^{n}\epsilon_t(\hat{\theta}_n)F\left(\lambda'\mathcal{W}(x_t)\right)$$
$$=\frac{1}{n}\sum_{t=1}^{n}\epsilon_t(\psi_n,\hat{\pi}_n)F\left(\lambda'\mathcal{W}(x_t)\right)-\hat{\mathfrak{b}}_{\psi,n}(\hat{\pi}_n,\lambda)'\left(\hat{\psi}_n-\psi_n\right)$$
$$=\frac{1}{n}\sum_{t=1}^{n}\epsilon_t(\psi_n,\hat{\pi}_n)F\left(\lambda'\mathcal{W}(x_t)\right)+O_{p,\lambda}(1/\sqrt{n})$$
$$=E\left[\epsilon_t F\left(\lambda'\mathcal{W}(x_t)\right)\right]+\frac{1}{n}\sum_{t=1}^{n}\left\{\epsilon_t F\left(\lambda'\mathcal{W}(x_t)\right)-E\left[\epsilon_t F\left(\lambda'\mathcal{W}(x_t)\right)\right]\right\}$$
$$-\beta_n'\frac{1}{n}\sum_{t=1}^{n}\left\{g(x_t,\hat{\pi}_n)-g(x_t,\pi_0)\right\}F\left(\lambda'\mathcal{W}(x_t)\right)+O_{p,\lambda}(1/\sqrt{n}).$$

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

Under $\sqrt{n}||\beta_n|| \to [0,\infty)$, by Lemma A.12, we have $\sup_{\pi\in\Pi,\lambda\in\Lambda}||1/n\sum_{t=1}^n \epsilon_t(\hat{\theta}_n) F(\lambda'\mathcal{W}(x_t)) - E[\epsilon_t F(\lambda'\mathcal{W}(x_t))]|| \xrightarrow{p} 0$. The claim now follows from $E[\epsilon_t F(\lambda'\mathcal{W}(x_t))] \neq 0\ \forall\lambda\in\Lambda/S$ for some $S\subset\Lambda$ with Lebesgue measure zero by Thm. 2.3 in Stinchcombe and White (1998), cf. Bierens (1990, Lem. 1).

Under $\mathcal{C}(ii,\omega_0)$, and the proof of Theorem 4.2.b,

$$\begin{aligned}
&\sum_{t=1}^n \epsilon_t(\hat{\theta}_n)F(\lambda'\mathcal{W}(x_t)) - E\left[\epsilon_t F(\lambda'\mathcal{W}(x_t))\right] \\
&\quad = \frac{1}{n}\sum_{t=1}^n \left\{\epsilon_t F\left(\lambda'\mathcal{W}(x_t)\right) - E\left[\epsilon_t F\left(\lambda'\mathcal{W}(x_t)\right)\right]\right\} \\
&\qquad + \mathfrak{b}_\theta(\omega_0,\pi_0,\lambda)'\mathcal{H}_\theta^{-1}\frac{1}{n}\sum_{t=1}^n \epsilon_t d_{\theta,t}(\beta_n/\|\beta_n\|,\pi_0) + O_{p,\lambda}(1/\sqrt{n}).
\end{aligned}$$

Mixing under Assumption 1.b implies ergodicity. The ergodic theorem therefore yields $1/n\sum_{t=1}^n \epsilon_t d_{\theta,t}(\beta_n/\|\beta_n\|,\pi_0) \xrightarrow{p} 0$ in view of identification Assumption 1 .a(ii) and stationarity. By Lemma B.13, $1/n\sum_{t=1}^n\{\epsilon_t F(\lambda'\mathcal{W}(x_t)) - E[\epsilon_t F(\lambda'\mathcal{W}(x_t))]\} \xrightarrow{p} 0$, and Thm. 2.3 in Stinchcombe and White (1998) yields $E[\epsilon_t F(\lambda'\mathcal{W}(x_t))] \neq 0 \forall\lambda\in\Lambda/S$. $\mathcal{QED}$.

Proof of Theorem 5.1.

Claim a. Let $\mathcal{C}(i,b)$ hold.

Step 1. We first show $||\hat{\Sigma}_n - \bar{\Sigma}(\pi^*(b),b)|| \xrightarrow{p} 0$, where $\bar{\Sigma}(\pi,b) \equiv \Sigma(\omega^*(\pi,b),\pi) = \Sigma(||\beta_0||,\omega^*(\pi,b),\zeta_0,\pi)$. Recall $\Sigma(||\beta||,\omega,\zeta,\pi) = \Sigma(\theta^+) \equiv \mathcal{H}_\theta(\theta^+)^{-1}\mathcal{V}(\theta^+)\mathcal{H}_\theta(\theta^+)^{-1}$, cf. (16). Recall $\hat{\beta}_n = \hat{\beta}_n(\hat{\pi}_n)$ hence we write $\omega(\hat{\beta}_n) = \omega_n(\hat{\pi}_n)$ as in (A.9). Thus joint weak convergence for $\sqrt{n}(\hat{\psi}_n(\hat{\pi}_n) - \psi_n), \hat{\pi}_n$ and $\omega(\hat{\beta}_n)$ holds by (A.10). Hence,

$$\left[\sqrt{n}\left(\hat{\psi}_n(\hat{\pi}_n) - \psi_n\right)', \hat{\pi}_n', \omega(\hat{\beta}_n)'\right]' \xrightarrow{d} \left[\tau(\pi^*(b),b)', \pi^*(b)', \frac{\tau_\beta(\pi^*(b),b)'}{\left\|\tau_\beta(\pi^*(b),b)\right\|}\right]'.$$

Joint convergence for $\hat{\pi}_n$ and $\omega(\hat{\beta}_n)$, $\sup_{\pi\in\Pi}||\hat{\psi}_n(\pi) - \psi_n|| \xrightarrow{p} 0$ by the proof of Theorem 4.1, uniform convergence Lemma B.6 for $\widehat{\mathcal{H}}_n(\theta^+)$ and $\hat{\mathcal{V}}_n(\theta^+)$, and the mapping theorem, together yield $||\hat{\Sigma}_n - \bar{\Sigma}(\pi^*(b),b)|| \xrightarrow{p} 0$.

Step 2. Now invoke $(\sqrt{n}(\hat{\psi}_n(\hat{\pi}_n) - \psi_n), \hat{\pi}_n) \xrightarrow{d} (\tau(\pi^*(b),b), \pi^*(b))$ by Theorem 4.1.a, and the mapping theorem, to complete the proof.

Claim b. Let $\mathcal{C}(ii,\omega_0)$ hold, and let $\{\kappa_n\}$ be a sequence of positive constants, $\kappa_n \to \infty$ and $\kappa_n = o(\sqrt{n})$. Write

$$\begin{aligned}
\kappa_n^{-2}\mathcal{A}_n^2 = &\frac{1}{k_\beta}\frac{n}{\kappa_n^2}\left(\hat{\beta}_n - \beta_n\right)'\hat{\Sigma}_{\beta,\beta,n}^{-1}\left(\hat{\beta}_n - \beta_n\right) \\
&+ 2\frac{1}{k_\beta}\frac{\sqrt{n}}{\kappa_n}\left(\hat{\beta}_n - \beta_n\right)'\hat{\Sigma}_{\beta,\beta,n}^{-1}\frac{\sqrt{n}}{\kappa_n}\beta_n \\
&+ \frac{1}{k_\beta}\frac{\sqrt{n}}{\kappa_n}\beta_n'\hat{\Sigma}_{\beta,\beta,n}^{-1}\frac{\sqrt{n}}{\kappa_n}\beta_n \equiv \mathcal{B}_{n,1} + \mathcal{B}_{n,2} + \mathcal{B}_{n,3}.
\end{aligned}$$

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

Lemma B.6 yields that $\hat{\Sigma}_{\beta,\beta,n}$ is positive definite asymptotically with probability approaching one. Hence $\mathcal{B}_{n,1} \xrightarrow{p} 0$ by application of Theorem 4.1.b. If $\sqrt{n}||\beta_n||/\kappa_n = O(1)$ then $|\mathcal{B}_{n,2}| \xrightarrow{p} 0$ because $\sqrt{n}\kappa_n^{-1}(\hat{\beta}_n - \beta_n) = O_p(1/\kappa_n) = o_p(1)$, and $\mathcal{B}_{n,3} \xrightarrow{p} [0,\infty)$, hence $\kappa_n^{-1}\mathcal{A}_n \xrightarrow{p} [0,\infty)$.

If $\sqrt{n}||\beta_n||/\kappa_n \to \infty$ then $|\mathcal{B}_{n,2}| = O_p(\mathcal{B}_{n,3})$ because $\sqrt{n}\kappa_n^{-1}(\hat{\beta}_n - \beta_n) = o_p(1)$, and $\mathcal{B}_{n,3} \xrightarrow{p} \infty$, hence $\kappa_n^{-1}\mathcal{A}_n \xrightarrow{p} \infty$. An example of this final case is $\beta_0 \neq 0$: $||\beta_n|| \to ||\beta_0|| > 0$ while $\kappa_n = o(\sqrt{n})$ by supposition hence $\sqrt{n}||\beta_n||/\kappa_n \to \infty$. $\mathcal{QED}$.

Proof of Theorem 6.1.

Claim (a). Recall $h \equiv (\gamma_0, b) \in \mathfrak{H} \equiv \{h : \gamma_0 \in \Gamma^*, \text{ and } ||b|| < \infty, \text{ with } \beta_0 = 0\}$, where Γ^* is the true parameter space in (12).

Let F_γ be the distribution function of $W_t = [y_t, x_t']'$ under some $\gamma \in \Gamma^*$. Let P_γ denote probability under F_γ. Recall $\mathcal{F}_\infty(c) \equiv P(\mathcal{T}(\lambda) \leq c)$, where $\{\mathcal{T}(\lambda) : \lambda \in \Lambda\}$ is the asymptotic null chi-squared process under strong identification, and $\mathcal{F}_{\lambda,h}(c) \equiv P(\mathcal{T}_\psi(\lambda,h) \leq c)$ where $\{\mathcal{T}_\psi(\lambda,h)\colon \lambda \in \Lambda\}$ is the asymptotic null process under weak identification. Write the finite sample p-values $p_n^\infty(\lambda) \equiv 1 - \mathcal{F}_\infty(\mathcal{T}_n(\lambda)) = \bar{\mathcal{F}}_\infty(\mathcal{T}_n(\lambda))$ and $p_n(\lambda,h) \equiv 1 - \mathcal{F}_{\lambda,h}(\mathcal{T}_n(\lambda)) = \bar{\mathcal{F}}_{\lambda,h}(\mathcal{T}_n(\lambda))$. Recall that $\mathcal{F}_{\lambda,\tilde{h}}(\cdot)$ is continuous by Assumption 5.

Step 1 (LF). The asymptotic size $AsySz \equiv \sup_{\lambda\in\Lambda} AsySz(\lambda)$ can be written as:

$$
\begin{aligned}
AsySz &= \sup_{\lambda\in\Lambda} \limsup_{n\to\infty} \sup_{\gamma\in\Gamma^*} P_\gamma \left(\max \left\{ \sup_{h\in\mathfrak{H}} \{p_n(\lambda,h)\}, p_n^\infty(\lambda) \right\} < \alpha | H_0 \right) \\
&= \sup_{\lambda\in\Lambda} \limsup_{n\to\infty} \sup_{\gamma\in\Gamma^*} P_\gamma \left(\max \left\{ \sup_{h\in\mathfrak{H}} \left\{ \bar{\mathcal{F}}_{\lambda,h}(\mathcal{T}_n(\lambda)) \right\}, \bar{\mathcal{F}}_\infty(\mathcal{T}_n(\lambda)) \right\} < \alpha | H_0 \right) \equiv \mathfrak{A},
\end{aligned}
$$

say. By Theorem 4.2.a, $\{\mathcal{T}_n(\lambda) : \Lambda\} \Rightarrow^* \{\mathcal{T}_\psi(\lambda,h) : \Lambda\}$ under $\mathcal{C}(i,b)$ with $||b|| < \infty$. Weak convergence implies convergence in finite dimensional distributions. By the definition of distribution convergence, and the mapping theorem, weak convergence therefore yields:

$$
\begin{aligned}
\mathfrak{A} &= \sup_{\lambda\in\Lambda} \sup_{\tilde{h}\in\mathfrak{H}} P \left(\max \left\{ \sup_{h\in\mathfrak{H}} \left\{ \bar{\mathcal{F}}_{\lambda,h} \left(\mathcal{T}_\psi(\lambda,\tilde{h}) \right) \right\}, \bar{\mathcal{F}}_\infty \left(\mathcal{T}_\psi(\lambda,\tilde{h}) \right) \right\} < \alpha \right) \qquad \textbf{(A.14)} \\
&\leq \sup_{\lambda\in\Lambda} \sup_{\tilde{h}\in\mathfrak{H}} P \left(\sup_{h\in\mathfrak{H}} \left\{ \bar{\mathcal{F}}_{\lambda,h} \left(\mathcal{T}_\psi(\lambda,\tilde{h}) \right) \right\} < \alpha \right) \\
&\leq \sup_{\lambda\in\Lambda} \sup_{\tilde{h}\in\mathfrak{H}} P \left(\bar{\mathcal{F}}_{\lambda,\tilde{h}} \left(\mathcal{T}_\psi(\lambda,\tilde{h}) \right) < \alpha \right) = \alpha.
\end{aligned}
$$

In the first inequality notice $\sup_{h\in\mathfrak{H}}$ operates only on the distribution function $\bar{\mathcal{F}}_{\lambda,h}$, whereas $\sup_{\tilde{h}\in\mathfrak{H}}$ operates on the limit process $\mathcal{T}_\psi(\lambda,\tilde{h})$. The inequalities hold since $\sup_{h\in\mathfrak{H}}\{\bar{\mathcal{F}}_{\lambda,h}(\mathcal{T}_\psi(\lambda,\tilde{h}))\} \leq \max\{\sup_{h\in\mathfrak{H}}\{\bar{\mathcal{F}}_{\lambda,h}(\mathcal{T}_\psi(\lambda,\tilde{h}))\}, \bar{\mathcal{F}}_\infty(\mathcal{T}_\psi(\lambda,\tilde{h}))\}$ and $\bar{\mathcal{F}}_{\lambda,\tilde{h}}(\mathcal{T}_\psi(\lambda,\tilde{h})) \leq \sup_{h\in\mathfrak{H}}\{\bar{\mathcal{F}}_{\lambda,h}(\mathcal{T}_\psi(\lambda,\tilde{h}))\}$. The last equality in (A.14) applies since $\mathcal{T}_\psi(\lambda,\tilde{h})$ is distributed $\mathcal{F}_{\lambda,\tilde{h}}$ which is continuous: hence $P(\bar{\mathcal{F}}_{\lambda,\tilde{h}}(\mathcal{T}_\psi(\lambda,\tilde{h})) < \alpha) = \alpha$ for any $\tilde{h}$ and λ.

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

Under $\mathcal{C}(ii,\omega_0)$, $\{\mathcal{T}_n(\lambda):\Lambda\} \Rightarrow^* \{\mathcal{T}(\lambda):\Lambda\}$ by Theorem 4.2.b. Since $\mathcal{T}(\lambda)$ is distributed $\mathcal{F}_\infty$, which is continuous:

$$\mathfrak{A} = \sup_{\lambda\in\Lambda}\sup_{\gamma\in\Gamma^*} P\left(\max\left\{\sup_{h\in\mathfrak{H}}\left\{\bar{\mathcal{F}}_{\lambda,h}(\mathcal{T}(\lambda))\right\}, \bar{\mathcal{F}}_\infty(\mathcal{T}(\lambda))\right\} < \alpha\right)$$
$$\leq \sup_{\lambda\in\Lambda}\sup_{\gamma\in\Gamma^*} P\left(\bar{\mathcal{F}}_\infty(\mathcal{T}(\lambda)) < \alpha\right) = \alpha.$$

Step 2 (ICS-1). By Theorem 5.1.a, $\mathcal{A}_n = O_p(1)$ under $\mathcal{C}(i,b)$ with $||b|| < \infty$. Further $\kappa_n \to \infty$ and $\kappa_n = o(n^{1/2})$. Hence

$$p_n^{(ICS-1)}(\lambda) \equiv \left\{p_n^{(LF)}(\lambda) \text{ if } \mathcal{A}_n \leq \kappa_n,\ p_n^\infty(\lambda) \text{ if } \mathcal{A}_n > \kappa_n\right\} = p_n^{(LF)}(\lambda)$$

asymptotically with probability approaching one.

Under $\mathcal{C}(ii,\omega_0)$, $\mathcal{A}_n \xrightarrow{p} \infty$ by Theorem 5.1.a. If $\kappa_n^{-1}\mathcal{A}_n \xrightarrow{p} [0,\infty)$ then again $p_n^{(ICS-1)}(\lambda) = p_n^{(LF)}(\lambda)$ asymptotically with probability approaching one. If $\kappa_n^{-1}\mathcal{A}_n \xrightarrow{p} \infty$, for example when $\beta_0 = 0$ (see Theorem 5.1.b), then $p_n^{(ICS-1)}(\lambda) = p_n^\infty(\lambda)$ asymptotically with probability approaching one.

In each case, a p-value is chosen that leads to correct asymptotic level $AsySz \leq \alpha$ in view of Step 1.

Claim (b). Let H_0 be false. By Theorem 4.3 $\mathcal{T}_n(\lambda) \xrightarrow{p} \infty\ \forall\lambda \in \Lambda/S$ where $S \subset \Lambda$ has Lebesgue measure zero. Theorem 4.2.a states $\sup_{\lambda\in\Lambda}\{\mathcal{T}_\psi(\lambda,b)\} < \infty$ a.s., hence the distribution of $\mathcal{T}_\psi(\lambda,b)$ has support $[0,\infty)$. Therefore $\sup_{h\in\mathfrak{H}}\{p_n(\lambda,h)\} \xrightarrow{p} 0$ and $p_n^\infty(\lambda) \xrightarrow{p} 0$, hence by construction $p_n^{(LF)}(\lambda) \xrightarrow{p} 0$ and by arguments under Step 2 above $p_n^{(ICS-1)}(\lambda) \xrightarrow{p} 0$. $\mathcal{QED}$.

Proof of Theorem 6.2.

Claim (a).

Step 1. Operate conditionally on the sample $\mathfrak{W}_n \equiv \{(y_t,x_t)\}_{t=1}^n$. In this step, we prove the bootstrapped test statistic converges weakly in probability to the Theorem 4.2.a null limit process under weak identification:

$$\left\{\hat{\mathcal{T}}^*_{\psi,n}(\lambda,h):\lambda\in\Lambda\right\} \Rightarrow^p \left\{\left(\frac{\mathfrak{T}_\psi(\pi^*(b),\lambda,b)}{\bar{\nu}(\pi^*(b),\lambda,b)}\right)^2 : \lambda\in\Lambda\right\} = \left\{\mathcal{T}_\psi(\lambda,h):\lambda\in\Lambda\right\}. \quad \textbf{(A.15)}$$

We then prove the claim in Step 2.

Step 1.1 Recall $\sigma_0\widehat{\mathcal{G}}^*_{\psi,n}(\pi) \equiv 1/\sqrt{n}\sum_{t=1}^n \sigma_0 z_t \widehat{\mathcal{H}}^{-1/2}_{\psi,n}(\pi) d_{\psi,t}(\pi)$ where z_t is i.i.d. $N(0,1)$. We will prove $\{\sigma_0\widehat{\mathcal{G}}^*_{\psi,n}(\pi):\pi\in\Pi\} \Rightarrow^p \{\mathcal{H}_\psi^{-1/2}(\pi)\mathcal{G}_\psi(\pi):\pi\in\Pi\}$, where $\mathcal{G}_\psi(\pi) = \mathcal{G}_\psi(\psi_0,\pi)$, and $\mathcal{G}_\psi(\theta) \equiv \mathcal{G}_\psi(\psi,\pi)$ is the Lemma B.1 case $\mathcal{C}(i,b)$ limit process. We need to prove convergence in finite dimensional distributions, and demonstrate stochastic equicontinuity.[19]

In order to establish convergence in finite dimensional distributions, we use Hansen's (1996, proof of Thm. 2) argument. Denote $E_{\mathfrak{W}_n}[\cdot] = E[\cdot|\mathfrak{W}_n]$. By Gaussianicity of z_t,

[19] See Thm. 3.1 in Gine and Zinn (1990), and see Dudley (1978) and Gine and Zinn (1986, Thm. 1.1.3).

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

$\sigma_0\widehat{\mathcal{G}}^*_{\psi,n}(\pi)$ is (conditionally) normally distributed with mean zero and covariance kernel:

$$E_{\mathfrak{W}_n}\left[\sigma_0^2\widehat{\mathcal{G}}^*_{\psi,n}(\pi)\widehat{\mathcal{G}}^*_{\psi,n}(\tilde{\pi})'\right] = \sigma_0^2\widehat{\mathcal{H}}^{-1/2}_{\psi,n}(\pi)\frac{1}{n}\sum_{t=1}^{n} d_{\psi,t}(\pi)d_{\psi,t}(\tilde{\pi})'\widehat{\mathcal{H}}^{-1/2}_{\psi,n}(\tilde{\pi})$$
$$= \sigma_0^2\widehat{\mathcal{H}}^{-1/2}_{\psi,n}(\pi)\widehat{\mathcal{H}}_{\psi,n}(\pi,\tilde{\pi})\widehat{\mathcal{H}}^{-1/2}_{\psi,n}(\tilde{\pi}),$$

where $\widehat{\mathcal{H}}_{\psi,n}(\pi,\tilde{\pi})$ is implicitly defined. Let $\mathcal{W}$ be the set of asymptotic sample $\{(y_t,x_t)\}_{t=1}^{\infty}$ such that

$$\sup_{\pi,\tilde{\pi}\in\Pi\times\Pi}\left\|E_{\mathfrak{W}_n}\left[\widehat{\mathcal{G}}^*_{\psi,n}(\pi)\widehat{\mathcal{G}}^*_{\psi,n}(\tilde{\pi})'\right] - \mathcal{H}_\psi^{-1/2}(\pi)\mathcal{H}_\psi(\pi,\tilde{\pi})\mathcal{H}_\psi^{-1/2}(\pi)\right\| \xrightarrow{p} 0. \qquad \textbf{(A.16)}$$

By Lemma B.2 $\sup_{\pi\in\Pi}||\widehat{\mathcal{H}}_{\psi,n}(\pi) - \mathcal{H}_\psi(\pi)|| \xrightarrow{p} 0$, and by Assumption 1.d(iii) $\mathcal{H}_\psi(\pi)$ is positive definite uniformly on Π. By the same argument used to prove Lemma B.2, $\sup_{\pi,\tilde{\pi}\in\Pi}||\widehat{\mathcal{H}}_{\psi,n}(\pi,\tilde{\pi}) - \mathcal{H}_\psi(\pi,\tilde{\pi})|| \xrightarrow{p} 0$ where $\mathcal{H}_\psi(\pi,\tilde{\pi}) \equiv E[d_{\psi,t}(\pi)d_{\psi,t}(\tilde{\pi})']$. This proves $P(\mathfrak{W}_n \in \mathcal{W}) = 1$. Therefore, $\sigma_0\widehat{\mathcal{G}}^*_{\psi,n}(\pi)$ converges in finite dimensional distributions to a zero-mean Gaussian law with covariance kernel $\sigma_0^2\mathcal{H}_\psi^{-1/2}(\pi)\mathcal{H}_\psi(\pi,\tilde{\pi})\mathcal{H}_\psi^{-1/2}(\pi)$. $\mathcal{H}_\psi^{-1/2}(\pi)\mathcal{G}_{\psi,n}(\pi)$ has the same limit under the null and case $\mathcal{C}(i,b)$ with $||b|| < \infty$. This follows by Lemma B.1, the $\mathcal{G}_{\psi,n}(\psi_{0,n},\pi)$ identify (17), and the following moment under the null and Assumption 1.a(i):

$$E\left[\mathcal{H}_\psi^{-1/2}(\pi)\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\epsilon_t d_{\psi,t}(\pi)\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\epsilon_t d_{\psi,t}(\tilde{\pi})'\mathcal{H}_\psi^{-1/2}(\tilde{\pi})\right]$$
$$= \sigma_0^2\mathcal{H}_\psi^{-1/2}(\pi)\mathcal{H}_\psi(\pi,\tilde{\pi})\mathcal{H}_\psi^{-1/2}(\tilde{\pi}).$$

Now apply Lemma B.1 to yield that $\mathcal{H}_\psi^{-1/2}(\pi)\mathcal{G}_{\psi,n}(\pi)$ converges in finite dimensional distributions to $\mathcal{H}_\psi^{-1/2}(\pi)\mathcal{G}_\psi(\pi)$, a zero-mean Gaussian law with kernel $\sigma_0^2\mathcal{H}_\psi^{-1/2}(\pi)$ $\mathcal{H}_\psi(\pi,\tilde{\pi})\mathcal{H}_\psi^{-1/2}(\pi)$. Since Gaussian processes are fully characterized by their mean and covariance functions, $\sigma_0\widehat{\mathcal{G}}^*_{\psi,n}(\pi)$ therefore converges in finite dimensional distributions to $\mathcal{H}_\psi^{-1/2}(\pi)\mathcal{G}_\psi(\pi)$.

Next we establish stochastic equicontinuity. Let $r \in \mathbb{R}^{k_x+k_\beta}$, $r'r = 1$, be arbitrary. By the mean value theorem, for some $\mathring{\pi} \in \Pi$, $||\mathring{\pi} - \pi|| \le ||\tilde{\pi} - \pi||$:

$$r'\widehat{\mathcal{H}}^{-1/2}_{\psi,n}(\pi)d_{\psi,t}(\pi) - r'\widehat{\mathcal{H}}^{-1/2}_{\psi,n}(\tilde{\pi})d_{\psi,t}(\tilde{\pi})$$
$$= \left(\left[r'\frac{\partial}{\partial\pi_i}\widehat{\mathcal{H}}^{-1/2}_{\psi,n}(\mathring{\pi})d_{\psi,t}(\mathring{\pi}) + r'\widehat{\mathcal{H}}^{-1/2}_{\psi,n}(\mathring{\pi})\frac{\partial}{\partial\pi_i}d_{\psi,t}(\mathring{\pi})\right]_{i=1}^{k_\pi}\right)'(\pi - \tilde{\pi}).$$

The derivatives are

$$\frac{\partial}{\partial\pi}d_{\psi,t}(\pi) = \begin{bmatrix}\frac{\partial}{\partial\pi}g(x_t,\pi)\\ 0_{k_x\times k_\pi}\end{bmatrix} \text{ and } \frac{\partial}{\partial\pi_i}\widehat{\mathcal{H}}^{-1/2}_{\psi,n}(\pi) = -\widehat{\mathcal{H}}^{-1/2}_{\psi,n}(\pi)\frac{\partial}{\partial\pi_i}\widehat{\mathcal{H}}_{\psi,n}(\pi)\widehat{\mathcal{H}}^{-1/2}_{\psi,n}(\pi)$$
$$\frac{\partial}{\partial\pi_i}\widehat{\mathcal{H}}_{\psi,n}(\pi) = \frac{1}{n}\sum_{t=1}^{n}\begin{bmatrix}\frac{\partial}{\partial\pi_i}g(x_t,\pi)g(x_t,\pi)' + g(x_t,\pi)\frac{\partial}{\partial\pi_i}g(x_t,\pi)' & \frac{\partial}{\partial\pi_i}g(x_t,\pi)x_t'\\ x_t\frac{\partial}{\partial\pi_i}g(x_t,\pi)' & 0_{k_x\times k_x}\end{bmatrix}.$$

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

Invoke Chebyshev's inequality, and the fact that z_t is i.i.d., independent of $\mathfrak{W}_n$, and not a function of π, to yield

$$\mathcal{P}_n(\eta) \equiv$$

$$P\left(\sup_{\pi,\tilde{\pi}\in\Pi:||\pi-\tilde{\pi}||\leq\delta}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}z_t\left\{r'\widehat{\mathcal{H}}_{\psi,n}^{-1/2}(\pi)d_{\psi,t}(\pi)-r'\widehat{\mathcal{H}}_{\psi,n}^{-1/2}(\tilde{\pi})d_{\psi,t}(\tilde{\pi})\right\}\right|>\eta|\mathfrak{W}_n\right)$$

$$\leq\frac{1}{\eta^2}E\left[\sup_{\pi,\tilde{\pi}\in\Pi:||\pi-\tilde{\pi}||\leq\delta}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}z_t\left\{r'\widehat{\mathcal{H}}_{\psi,n}^{-1/2}(\pi)d_{\psi,t}(\pi)\right.\right.\right.$$

$$\left.\left.\left.-r'\widehat{\mathcal{H}}_{\psi,n}^{-1/2}(\tilde{\pi})d_{\psi,t}(\tilde{\pi})\right\}\right)^2|\mathfrak{W}_n\right]$$

$$=\frac{1}{\eta^2}\frac{1}{n}\sum_{t=1}^{n}\sup_{\pi,\tilde{\pi}\in\Pi:||\pi-\tilde{\pi}||\leq\delta}\left\{r'\widehat{\mathcal{H}}_{\psi,n}^{-1/2}(\pi)d_{\psi,t}(\pi)-r'\widehat{\mathcal{H}}_{\psi,n}^{-1/2}(\tilde{\pi})d_{\psi,t}(\tilde{\pi})\right\}^2$$

$$\leq\frac{\delta^2}{\eta^2}\frac{1}{n}\sum_{t=1}^{n}\sup_{\pi\in\Pi}\left|\left[r'\frac{\partial}{\partial\pi_i}\widehat{\mathcal{H}}_{\psi,n}^{-1/2}(\pi)d_{\psi,t}(\pi)+r'\widehat{\mathcal{H}}_{\psi,n}^{-1/2}(\pi)\frac{\partial}{\partial\pi_i}d_{\psi,t}(\pi)\right]_{i=1}^{k_\pi}\right|^2=\frac{\delta^2}{\eta^2}\mathcal{C}_n,$$

say. We prove below that $\mathcal{C}_n\xrightarrow{p}\mathcal{C}$ a finite non-negative constant. We can therefore choose any $\delta>0$ if $\mathcal{C}=0$, and otherwise $0<\delta\leq[\epsilon\eta^2/\mathcal{C}]^{1/2}$, such that for each $(\epsilon,\eta)>0$ there exists $\delta>0$ yielding $\lim_{n\to\infty}\mathcal{P}_n(\eta)<\epsilon$ asymptotically with probability approaching one with respect to the sample draw $\mathfrak{W}_n$. This establishes stochastic equicontinuity.

We now prove $\mathcal{C}_n\xrightarrow{p}\mathcal{C}\in[0,\infty)$. Since $\widehat{\mathcal{H}}_{\psi,n}(\pi)\equiv 1/n\sum_{t=1}^{n}d_{\psi,t}(\pi)d_{\psi,t}(\pi)'$, note that:

$$\frac{1}{n}\sum_{t=1}^{n}\sup_{\pi\in\Pi}\left|\left[r'\frac{\partial}{\partial\pi_i}\widehat{\mathcal{H}}_{\psi,n}^{-1/2}(\pi)d_{\psi,t}(\pi)+r'\widehat{\mathcal{H}}_{\psi,n}^{-1/2}(\pi)\frac{\partial}{\partial\pi_i}d_{\psi,t}(\pi)\right]_{i=1}^{k_\pi}\right|^2$$

$$=\sum_{i,j=1}^{k_\pi}\sup_{\pi\in\Pi}r'\frac{\partial}{\partial\pi_i}\widehat{\mathcal{H}}_{\psi,n}^{-1/2}(\pi)\widehat{\mathcal{H}}_{\psi,n}(\pi)\frac{\partial}{\partial\pi_j}\widehat{\mathcal{H}}_{\psi,n}^{-1/2}(\pi)r$$

$$+\sum_{i,j=1}^{k_\pi}\sup_{\pi\in\Pi}r'\widehat{\mathcal{H}}_{\psi,n}^{-1/2}(\pi)\frac{1}{n}\sum_{t=1}^{n}\frac{\partial}{\partial\pi_i}d_{\psi,t}(\pi)\frac{\partial}{\partial\pi_j}d_{\psi,t}(\pi)'\widehat{\mathcal{H}}_{\psi,n}^{-1/2}(\pi)r$$

$$+2\sum_{i,j=1}^{k_\pi}\sup_{\pi\in\Pi}r'\frac{\partial}{\partial\pi_i}\widehat{\mathcal{H}}_{\psi,n}^{-1/2}(\pi)\frac{1}{n}\sum_{t=1}^{n}d_{\psi,t}(\pi)\frac{\partial}{\partial\pi_i}d_{\psi,t}(\pi)'\widehat{\mathcal{H}}_{\psi,n}^{-1/2}(\pi)r.$$

The argument used to prove Lemmas B.2 and B.5 extend to each component of $(\partial/\partial\pi_i)\widehat{\mathcal{H}}_{\psi,n}^{-1/2}(\pi)$, $1/n\sum_{t=1}^{n}d_{\psi,t}(\pi)(\partial/\partial\pi_i)d_{\psi,t}(\pi)'$ and $1/n\sum_{t=1}^{n}(\partial/\partial\pi_i)d_{\psi,t}(\pi)(\partial/\partial\pi_i)d_{\psi,t}(\pi)'$, in view of the Assumption 1.b,c mixing and moment bounds. Thus, each summand has a uniformly bounded uniform probability limit under any case $\mathcal{C}(i,b)$ or $\mathcal{C}(ii,\omega_0)$. Hence, by Slutsky's theorem,

$$\mathcal{C}_n\xrightarrow{p}\sum_{i,j=1}^{k_\pi}\sup_{\pi\in\Pi}\left\{r'\frac{\partial}{\partial\pi_i}\mathcal{H}_\psi^{-1/2}(\pi)\mathcal{H}_\psi(\pi)\frac{\partial}{\partial\pi_j}\mathcal{H}_\psi^{-1/2}(\pi)r\right\}$$

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

$$+\sum_{i,j=1}^{k_\pi}\sup_{\pi\in\Pi}\left\{r'\mathcal{H}_\psi^{-1/2}(\pi)E\left[\frac{\partial}{\partial\pi_i}d_{\psi,t}(\pi)\frac{\partial}{\partial\pi_j}d_{\psi,t}(\pi)'\right]\widehat{\mathcal{H}}_{\psi,n}^{-1/2}(\pi)r\right\}$$

$$+2\sum_{i,j=1}^{k_\pi}\sup_{\pi\in\Pi}\left\{r'\frac{\partial}{\partial\pi_i}\mathcal{H}_\psi^{-1/2}(\pi)E\left[d_{\psi,t}(\pi)\frac{\partial}{\partial\pi_i}d_{\psi,t}(\pi)'\right]\mathcal{H}_\psi^{-1/2}(\pi)r\right\}\equiv\mathcal{C}<\infty.$$

Non-negativity $\mathcal{C}\geq 0$ is trivial in view of the quadratic form of $\mathcal{C}_n$.

Step 1.2 Next, we prove the bootstrapped $\hat{\pi}_n^*(\pi_0,b)\equiv\arg\min_{\pi\in\Pi}\{\hat{\xi}_{\psi,n}^*(\pi,\pi_0,b)\}$ satisfies:

$$\hat{\pi}_n^*(\pi_0,b)\Rightarrow^p\arg\min_{\pi\in\Pi}\left\{-\frac{1}{2}\left(\mathcal{G}_\psi(\pi)+\mathcal{D}_\psi(\pi,\pi_0)b\right)'\mathcal{H}_\psi^{-1}(\pi)\left(\mathcal{G}_\psi(\pi)+\mathcal{D}_\psi(\pi,\pi_0)b\right)\right\}$$
$$\equiv\pi^*(b). \qquad \textbf{(A.17)}$$

$\widehat{\mathcal{H}}_{\psi,n}(\pi)$ and $\hat{\mathcal{D}}_{\psi,n}(\pi,\pi_0)$ have uniform probability limits $\mathcal{H}_\psi(\pi)$ and $\mathcal{D}_\psi(\pi,\pi_0)$ by Lemma B.6. The Step 1.1 result of weak convergence in probability, the mapping theorem, and $\hat{\sigma}_n\xrightarrow{p}\sigma_0$ (see Remark 7), together yield:

$$\left\{\hat{\xi}_{\psi,n}^*(\pi,\pi_0,b):\pi\in\Pi\right\}$$
$$=\left\{-\frac{1}{2}\left(\hat{\sigma}_n\widehat{\mathcal{G}}_{\psi,n}^*(\pi)+\widehat{\mathcal{H}}_{\psi,n}^{-1/2}(\pi)\hat{\mathcal{D}}_{\psi,n}(\pi,\pi_0)\times b\right)'\right.$$
$$\left.\times\left(\hat{\sigma}_n\widehat{\mathcal{G}}_{\psi,n}^*(\pi)+\widehat{\mathcal{H}}_{\psi,n}^{-1/2}(\pi)\hat{\mathcal{D}}_{\psi,n}(\pi,\pi_0)\times b\right):\pi\in\Pi\right\}$$
$$\Rightarrow^p-\frac{1}{2}\left\{\mathcal{G}_\psi(\pi)+\mathcal{D}_\psi(\pi,\pi_0)b\right\}'\mathcal{H}_\psi^{-1}(\pi)\left\{\mathcal{G}_\psi(\pi)+\mathcal{D}_\psi(\pi,\pi_0)b:\pi\in\Pi\right\}.$$

Apply the mapping theorem again, and Assumption 2.a, to yield (A.17).

Step 1.3 Define $\mathcal{K}_{n,t}(\pi,\lambda)\equiv F(\lambda'\mathcal{W}(x_t))-\hat{\mathfrak{b}}_{\psi,n}(\pi,\lambda)'\widehat{\mathcal{H}}_{\psi,n}^{-1}(\pi)d_{\psi,t}(\pi)$ and recall $\mathcal{K}_{\psi,t}(\pi,\lambda)\equiv F(\lambda'\mathcal{W}(x_t))-\mathfrak{b}_\psi(\pi,\lambda)'\mathcal{H}_\psi^{-1}(\pi)d_{\psi,t}(\pi)$. We will show $\hat{\mathfrak{T}}_{\psi,n}^*(\pi,\lambda,\pi_0,b)$ defined in (20) satisfies:

$$\left\{\hat{\mathfrak{T}}_{\psi,n}^*(\pi,\lambda,\pi_0,b):\Pi,\Lambda\right\}\Rightarrow^p\left\{\mathfrak{T}_\psi(\pi,\lambda,b):\Pi,\Lambda\right\},$$

where, as in (10),

$$\mathfrak{T}_\psi(\pi,\lambda,b)\equiv\mathfrak{Z}_\psi(\pi,\lambda)+\mathfrak{b}_\psi(\pi,\lambda)'\left(\mathcal{H}_\psi^{-1}(\pi)\mathcal{D}_\psi(\pi)b+\left[b,0'_{k_\beta}\right]'\right)$$
$$+\mathfrak{b}_\psi(\pi,\lambda)'\mathcal{H}_\psi^{-1}(\pi)E\left[d_{\psi,t}(\pi)\{g(x_t,\pi_0)-g(x_t,\pi)\}'\right]b$$
$$+E\left[\mathcal{K}_{\psi,t}(\pi,\lambda)\{g(x_t,\pi_0)-g(x_t,\pi)\}'\right]b,$$

and $\mathfrak{Z}_\psi(\pi,\lambda)$ is the Lemma B.9 zero mean Gaussian limit process of $1/\sqrt{n}\sum_{t=1}^n\epsilon_t\mathcal{K}_{\psi,t}(\pi,\lambda)$.

Observe that $\widehat{\mathcal{H}}_{\psi,n}(\pi)$, $\hat{\mathcal{D}}_{\psi,n}(\pi,\pi_0)$, $\hat{\mathfrak{b}}_{\psi,n}(\hat{\psi}_n,\pi,\lambda)$, $1/n\sum_{t=1}^n d_{\psi,t}(\pi)g(x_t,\pi)'$ and $1/n\sum_{t=1}^n\mathcal{K}_{n,t}(\pi,\lambda)g(x_t,\pi)$ have uniform probability limits $\mathfrak{b}_\psi(\pi,\lambda)$, $\mathcal{H}_\psi(\pi)$, $\mathcal{D}_\psi(\pi)$, $E[d_{\psi,t}(\pi)g(x_t,\pi)']$, and $E[\mathcal{K}_{\psi,t}(\pi,\lambda)g(x_t,\pi)]$ by applications of Lemmas B.2, B.10, and B.13.

It therefore suffices to prove $\{\sigma_0\hat{\mathfrak{Z}}_{\psi,n}^*(\pi,\lambda):\Pi,\Lambda\}\Rightarrow^p\{\mathfrak{Z}_\psi(\pi,\lambda):\Pi,\Lambda\}$ where $\hat{\mathfrak{Z}}_{\psi,n}^*(\pi,\lambda)\equiv 1/\sqrt{n}\sum_{t=1}^n\sigma_0 z_t\mathcal{K}_{n,t}(\pi,\lambda)$. Note that $\hat{\mathfrak{Z}}_{\psi,n}^*(\pi,\lambda)$ is (conditionally) normally

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

distributed with zero-mean and covariance kernel $\sigma_0^2 1/n\sum_{t=1}^n \mathcal{K}_{n,t}(\pi,\lambda)\mathcal{K}_{n,t}(\tilde{\pi},\tilde{\lambda})$. Let $\tilde{\mathcal{W}}$ be the set of asymptotic sample $\{(y_t,x_t)\}_{t=1}^{\infty}$ such that

$$\sup_{\pi,\tilde{\pi}\in\Pi\times\Pi,\lambda,\tilde{\lambda}\in\Lambda}\left\|\frac{1}{n}\sum_{t=1}^n \mathcal{K}_{n,t}(\pi,\lambda)\mathcal{K}_{n,t}(\tilde{\pi},\tilde{\lambda})-E\left[\mathcal{K}_{\psi,t}(\pi,\lambda)\mathcal{K}_{\psi,t}(\tilde{\pi},\tilde{\lambda})\right]\right\|\xrightarrow{p}0.$$

By Lemmas B.2 and B.10: $\sup_{\pi,\tilde{\pi}\in\Pi\times\Pi,\lambda,\tilde{\lambda}\in\Lambda}|1/n\sum_{t=1}^n\{\mathcal{K}_{n,t}(\pi,\lambda)\mathcal{K}_{n,t}(\tilde{\pi},\tilde{\lambda})-\mathcal{K}_{\psi,t}(\pi,\lambda)\mathcal{K}_{\psi,t}(\tilde{\pi},\tilde{\lambda})\}|\xrightarrow{p}0$ and by the same arguments used to prove Lemma B.10: $\sup_{\pi,\tilde{\pi}\in\Pi\times\Pi,\lambda,\tilde{\lambda}\in\Lambda}|1/n\sum_{t=1}^n\{\mathcal{K}_{\psi,t}(\pi,\lambda)\mathcal{K}_{\psi,t}(\tilde{\pi},\tilde{\lambda})-E[\mathcal{K}_{\psi,t}(\pi,\lambda)\mathcal{K}_{\psi,t}(\tilde{\pi},\tilde{\lambda})]\}|\xrightarrow{p}0$. This proves $P(\mathfrak{W}_n\in\tilde{\mathcal{W}})=1$. Hence, $\hat{\mathfrak{Z}}^*_{\psi,n}(\pi,\lambda)$ converges in finite dimensional distributions to a zero-mean Gaussian law with kernel $\sigma_0^2E[\mathcal{K}_{\psi,t}(\pi,\lambda)\mathcal{K}_{\psi,t}(\tilde{\pi},\tilde{\lambda})]$.

By Lemma B.9.a, under the null $\{\mathfrak{Z}_{\psi,n}(\pi,\lambda):\Pi,\Lambda\}\Rightarrow^*\{\mathfrak{Z}_{\psi}(\pi,\lambda):\Pi,\Lambda\}$, a zero-mean Gaussian process with covariance kernel $\sigma_0^2E[\mathcal{K}_{\psi,t}(\pi,\lambda)\mathcal{K}_{\psi,t}(\tilde{\pi},\tilde{\lambda})]$. Therefore, the finite dimensional distributions of $\{\hat{\mathfrak{Z}}^*_{\psi,n}(\pi,\lambda)):\Pi,\Lambda\}$ converge to those of $\{\mathfrak{Z}_{\psi}(\pi,\lambda):\Pi,\Lambda\}$ under the null.

It remains to prove stochastic equicontinuity for $\hat{\mathfrak{Z}}^*_{\psi,n}(\pi,\lambda)$. By construction, we need only show the sequence of distributions of $1/\sqrt{n}\sum_{t=1}^n z_tF(\lambda'\mathcal{W}(x_t))$ and $1/\sqrt{n}\sum_{t=1}^n z_t\hat{\mathfrak{b}}_{\psi,n}(\pi,\lambda)'\widehat{\mathcal{H}}^{-1}_{\psi,n}(\pi)d_{\psi,t}(\pi)$ are stochastically equicontinuous, and invoke probability subadditivity. For $1/\sqrt{n}\sum_{t=1}^n z_tF(\lambda'\mathcal{W}(x_t))$, by the mean value theorem and Chebyshev's inequality and the fact $\{z_t\}_{t=1}^n$ is independent of the sample $\mathfrak{W}_n$:

$$\begin{aligned}
&P\left(\sup_{\lambda,\tilde{\lambda}\in\Lambda:||\lambda-\tilde{\lambda}||\le\delta}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^n z_t\left\{F(\lambda'\mathcal{W}(x_t))-F(\tilde{\lambda}'\mathcal{W}(x_t))\right\}\right|>\eta|\mathfrak{W}_n\right)\\
&\le\frac{1}{\eta^2}\frac{1}{n}\sum_{t=1}^n\sup_{\lambda,\tilde{\lambda}\in\Lambda:||\lambda-\tilde{\lambda}||\le\delta}\left\{F(\lambda'\mathcal{W}(x_t))-F(\tilde{\lambda}'\mathcal{W}(x_t))\right\}^2\\
&\le\frac{1}{\eta^2}\frac{1}{n}\sum_{t=1}^n\sup_{\lambda\in\Lambda}\left\|\frac{\partial}{\partial\lambda}F(\lambda'\mathcal{W}(x_t))\right\|^2\times\delta^2.
\end{aligned}$$

The Assumption 1.c envelope bounds and ergodicity imply $1/n\sum_{t=1}^n\sup_{\lambda\in\Lambda}||(\partial/\partial\lambda)F(\lambda'\mathcal{W}(x_t))||^2\xrightarrow{p}E[\sup_{\lambda\in\Lambda}||(\partial/\partial\lambda)F(\lambda'\mathcal{W}(x_t))||^2]\le K<\infty$. Now pick $0<\delta\le[\epsilon\eta^2/K]^{1/2}$ to complete the proof of stochastic equicontinuity asymptotically with probability approaching one with respect to the draw $\mathfrak{W}_n$.

Next, for $1/\sqrt{n}\sum_{t=1}^n z_t\hat{\mathfrak{b}}_{\psi,n}(\pi,\lambda)'\widehat{\mathcal{H}}^{-1}_{\psi,n}(\pi)d_{\psi,t}(\pi)$, write

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{t=1}^n\sigma_0z_t\left\{\hat{\mathfrak{b}}_{\psi,n}(\pi,\lambda)'\widehat{\mathcal{H}}^{-1}_{\psi,n}(\pi)d_{\psi,t}(\pi)-\hat{\mathfrak{b}}_{\psi,n}(\tilde{\pi},\tilde{\lambda})'\widehat{\mathcal{H}}^{-1}_{\psi,n}(\tilde{\pi})d_{\psi,t}(\tilde{\pi})\right\}\\
&\quad=\hat{\mathfrak{b}}_{\psi,n}(\tilde{\pi},\tilde{\lambda})'\widehat{\mathcal{H}}^{-1}_{\psi,n}(\tilde{\pi})\frac{1}{\sqrt{n}}\sum_{t=1}^n\sigma_0z_t\left\{d_{\psi,t}(\pi)-d_{\psi,t}(\tilde{\pi})\right\} \qquad \textbf{(A.18)}\\
&\qquad+\hat{\mathfrak{b}}_{\psi,n}(\tilde{\pi},\tilde{\lambda})'\widehat{\mathcal{H}}^{-1}_{\psi,n}(\pi)\left\{\widehat{\mathcal{H}}_{\psi,n}(\tilde{\pi})-\widehat{\mathcal{H}}_{\psi,n}(\pi)\right\}\widehat{\mathcal{H}}^{-1}_{\psi,n}(\tilde{\pi})\frac{1}{\sqrt{n}}\sum_{t=1}^n\sigma_0z_td_{\psi,t}(\pi)
\end{aligned}$$

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

$$+\left\{\hat{\mathfrak{b}}_{\psi,n}(\pi,\lambda)-\hat{\mathfrak{b}}_{\psi,n}(\tilde{\pi},\tilde{\lambda})\right\}'\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\sigma_0 z_t\widehat{\mathcal{H}}_{\psi,n}^{-1}(\pi)d_{\psi,t}(\pi).$$

By Lemmas B.2 and B.10, $\sup_{\pi\in\Pi}||\widehat{\mathcal{H}}_{\psi,n}(\pi)-\mathcal{H}_{\psi}(\pi)||\xrightarrow{p}0$ and $\sup_{\pi\in\Pi,\lambda\in\Lambda}||\hat{\mathfrak{b}}_{\psi,n}(\pi,\lambda)-\mathfrak{b}_{\psi}(\pi,\lambda)||\xrightarrow{p}0$. Step 1.1 gives first-order expansions for both $1/\sqrt{n}\sum_{t=1}^{n}\sigma_0 z_t\{d_{\psi,t}(\pi)-d_{\psi,t}(\tilde{\pi})\}$ and $\widehat{\mathcal{H}}_{\psi,n}(\tilde{\pi})-\widehat{\mathcal{H}}_{\psi,n}(\pi)$ around π. Arguments there suffice to prove the first two summands in (A.18) are stochastically equicontinuous asymptotically with probability approaching one with respect to the sample draw.

Consider the third summand in (A.18). By the Step 1.1 argument and Lemma B.2, $\sup_{\pi\in\Pi}||1/\sqrt{n}\sum_{t=1}^{n}\sigma_0 z_t\widehat{\mathcal{H}}_{\psi,n}^{-1}(\pi)d_{\psi,t}(\pi)||=O_p(1)$.

Next, write $\hat{\mathfrak{b}}_{\psi,n}(\chi)=\hat{\mathfrak{b}}_{\psi,n}(\pi,\lambda)$ where $\chi=[\pi',\lambda']'\in\mathcal{X}\equiv\Pi\times\Lambda$. Two applications of Minkowski's inequality yields:

$$\begin{aligned}
&\left|\sup_{\chi,\tilde{\chi}\in\mathcal{X}}\left\|\hat{\mathfrak{b}}_{\psi,n}(\chi)-\hat{\mathfrak{b}}_{\psi,n}(\tilde{\chi})\right\|-\sup_{\chi,\tilde{\chi}\in\mathcal{X}}\left\|\mathfrak{b}_{\psi}(\chi)-\mathfrak{b}_{\psi}(\tilde{\chi})\right\|\right|\\
&\leq\left|\sup_{\chi,\tilde{\chi}\in\mathcal{X}}\left\|\hat{\mathfrak{b}}_{\psi,n}(\chi)-\mathfrak{b}_{\psi}(\chi)-\hat{\mathfrak{b}}_{\psi,n}(\tilde{\chi})+\mathfrak{b}_{\psi}(\tilde{\chi})\right\|\right|\\
&\leq\left|\sup_{\chi\in\mathcal{X}}\left\|\hat{\mathfrak{b}}_{\psi,n}(\chi)-\mathfrak{b}_{\psi}(\chi)\right\|\right|+\left|\sup_{\tilde{\chi}\in\mathcal{X}}\left\|\hat{\mathfrak{b}}_{\psi,n}(\tilde{\chi})-\mathfrak{b}_{\psi}(\tilde{\chi})\right\|\right|.
\end{aligned}$$

The right hand side is $o_p(1)$ by Lemma B.10. Now apply the mean value theorem, the Cauchy–Schwartz inequality, and the Assumption 1.c envelope bounds to yield:

$$\begin{aligned}
&\sup_{\chi,\tilde{\chi}\in\mathcal{X}:||\chi-\tilde{\chi}||\leq\delta}\left\|\mathfrak{b}_{\psi}(\chi)-\mathfrak{b}_{n}(\tilde{\chi})\right\|\leq\sup_{\chi,\tilde{\chi}\in\mathcal{X}:||\chi-\tilde{\chi}||\leq\delta}\left|\mathfrak{b}_{\psi}(\chi)-\mathfrak{b}_{n}(\tilde{\chi})\right|\\
&\quad\leq\sup_{\chi,\tilde{\chi}\in\mathcal{X}:||\chi-\tilde{\chi}||\leq\delta}\left\{\left|E\left[\frac{\partial}{\partial\chi}F(\lambda'\mathcal{W}(x_t))d_{\psi,t}(\pi)\right]\right|\times|\chi-\tilde{\chi}|\right\}\\
&\quad\leq\left\{\left(\sup_{\lambda\in\Lambda}E\left[\left|\frac{\partial}{\partial\lambda}F(\lambda'\mathcal{W}(x_t))\right|^2\right]\times\sup_{\pi\in\Pi}E\left[\left|d_{\psi,t}(\pi)\right|^2\right]\right)^{1/2}\right.\\
&\qquad\left.+\left(\sup_{\lambda\in\Lambda}E\left[\left\{F(\lambda'\mathcal{W}(x_t)^2\right\}\right]\times\sup_{\pi\in\Pi}E\left[\left|\frac{\partial}{\partial\pi}d_{\psi,t}(\pi)\right|^2\right]\right)^{1/2}\right\}\times\delta=K\times\delta<\infty.
\end{aligned}$$

Stochastic equicontinuity asymptotically with probability approaching one with respect to the sample draw therefore follows for the third summand in (A.18) by the Step 1.1 argument.

Step 1.4 We prove joint weak convergence in probability:

$$\left\{\hat{\mathfrak{T}}_{\psi,n}^{*}(\pi,\lambda,\pi_0,b),\hat{\pi}_{n}^{*}(\pi_0,b):\Pi,\Lambda\right\}\Rightarrow^{p}\left\{\mathfrak{T}_{\psi}(\pi,\lambda,b),\pi^{*}(b):\Pi,\Lambda\right\}. \tag{A.19}$$

$\hat{\pi}_{n}^{*}(\pi_0,b)$ is a continuous function of $\sigma_0\widehat{\mathcal{G}}_{\psi,n}^{*}(\pi)$, $\widehat{\mathcal{H}}_{\psi,n}(\pi)$ and $\hat{\mathcal{D}}_{\psi,n}(\pi,\pi_0)$, where $\widehat{\mathcal{H}}_{\psi,n}(\pi)$ and $\hat{\mathcal{D}}_{\psi,n}(\pi,\pi_0)$ have uniform probability limits. Hence, by construction of $\hat{\mathfrak{T}}_{\psi,n}^{*}(\pi,\lambda,\pi_0,b)$, and uniform convergence in probability of key summands in

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

Steps 1.1–1.3, cf. Lemmas B.2, B.10, and B.13, it suffices to show:

$$\left\{\sigma_0\hat{\mathfrak{Z}}^*_{\psi,n}(\pi,\lambda),\sigma_0\widehat{\mathcal{G}}^*_{\psi,n}(\pi):\pi\in\Pi,\lambda\in\Lambda\right\}$$
$$\Rightarrow^p\left\{\mathfrak{Z}_\psi(\pi,\lambda),\mathcal{H}_\psi^{-1/2}(\pi)\mathcal{G}_\psi(\pi):\pi\in\Pi,\lambda\in\Lambda\right\}.$$

The required result then follows from the mapping theorem.

Let $r=[r_1,r_2']'$, $r_1\in\mathbb{R}$, $r_2\in\mathbb{R}^{k_x+k_\beta}$, $r'r=1$, and define $\hat{\mathcal{L}}_{n,t}(\pi,\lambda;r)\equiv r_1\mathcal{K}_{n,t}(\pi,\lambda)+r_2'\widehat{\mathcal{H}}_{\psi,n}^{-1/2}(\pi)\times d_{\psi,t}(\pi)$ and $\mathcal{L}_{n,t}(\pi,\lambda;r)\equiv r_1\mathcal{K}_{n,t}(\pi,\lambda)+r_2'\mathcal{H}_\psi^{-1/2}(\pi)\times d_{\psi,t}(\pi)$. Any linear combination $\sigma_0 r_1\hat{\mathfrak{Z}}^*_{\psi,n}(\pi,\lambda)+\sigma_0 r_2'\widehat{\mathcal{G}}^*_{\psi,n}(\pi)=1/\sqrt{n}\sum_{t=1}^n\sigma_0 z_t\mathcal{L}_{n,t}(\pi,\lambda;r)$ is normally distributed with zero mean and covariance kernel $\sigma_0^2 n^{-1}\sum_{t=1}^n\mathcal{L}_{n,t}(\pi,\lambda;r)\mathcal{L}_{n,t}(\tilde{\pi},\tilde{\lambda};r)$. By arguments used to prove Lemmas B.10 and B.13:

$$\sup_{\pi,\tilde{\pi}\in\Pi\times\Pi,\lambda,\tilde{\lambda}\in\Lambda}\left\|\frac{1}{n}\sum_{t=1}^n\left\{\hat{\mathcal{L}}_{n,t}(\pi,\lambda;r)\hat{\mathcal{L}}_{n,t}(\tilde{\pi},\tilde{\lambda};r)-\mathcal{L}_{n,t}(\pi,\lambda;r)\mathcal{L}_{n,t}(\tilde{\pi},\tilde{\lambda};r)\right\}\right\|\xrightarrow{p}0$$
$$\sup_{\pi,\tilde{\pi}\in\Pi\times\Pi,\lambda,\tilde{\lambda}\in\Lambda}\left\|\frac{1}{n}\sum_{t=1}^n\mathcal{L}_{n,t}(\pi,\lambda;r)\mathcal{L}_{n,t}(\tilde{\pi},\tilde{\lambda};r)-E\left[\mathcal{L}_{n,t}(\pi,\lambda;r)\mathcal{L}_{n,t}(\tilde{\pi},\tilde{\lambda};r)\right]\right\|\xrightarrow{p}0.$$

$(\sigma_0\hat{\mathfrak{Z}}^*_{\psi,n}(\pi,\lambda),\sigma_0\widehat{\mathcal{G}}^*_{\psi,n}(\pi))$ therefore convergences in finite dimensional distributions to $(\mathfrak{Z}_\psi(\pi,\lambda),\mathcal{H}_\psi^{-1/2}(\pi)\mathcal{G}_\psi(\pi))$. Stochastic equicontinuity for $1/\sqrt{n}\sum_{t=1}^n\sigma_0 z_t\mathcal{L}_{n,t}(\pi,\lambda;r)$ follows from Step 1.1 for $\widehat{\mathcal{G}}^*_{\psi,n}(\pi)$, Step 1.3 for $\hat{\mathfrak{Z}}^*_{\psi,n}(\pi,\lambda)$, and probability subadditivity.

Step 1.5 Finally, consider the test statistic denominator $\hat{v}_n^2(\omega,\pi,\lambda)$. By Lemma B.11:

$$\sup_{\substack{\{\omega\in\mathbb{R}^{k_\beta}:\omega'\omega=1\}\\ \times\Pi\times\Lambda}}\left|\hat{v}_n^2(\omega,\pi,\lambda)-E\left[\epsilon_t^2(\psi_0,\pi)\left\{F\left(\lambda'\mathcal{W}(x_t)\right)-\mathfrak{b}_\theta(\omega,\pi,\lambda)'\mathcal{H}_\theta^{-1}(\omega,\pi)d_{\theta,t}(\omega,\pi)\right\}^2\right]\right|\xrightarrow{p}0. \tag{A.20}$$

In conclusion, by weak convergence in probability results in Steps 1.1 and 1.2, plus the mapping theorem, $\hat{\tau}^*_{\beta,n}(\pi_0,b)\Rightarrow^p-\mathcal{S}_\beta\mathcal{H}_\psi^{-1}(\pi^*(b))\{\sigma_0\mathcal{G}_\psi(\pi^*(b))+\mathcal{D}_\psi(\pi^*(b),\pi_0)b\}=\tau_\beta(\pi^*(b),b)$. Hence $\hat{\omega}^*_n(\pi_0,b)\Rightarrow^p\tau_\beta(\pi^*(b),b)/||\tau_\beta(\pi^*(b),b)||$. The preceding limit, (A.17)–(A.20), and the mapping theorem yield (A.15).

Step 2. Assume $\mathcal{C}(i,b)$ with $||b||<\infty$. By independence and the Glivenko–Cantelli theorem $\hat{p}^*_{n,\mathcal{M}}(\lambda,h)\xrightarrow{p}P(\hat{\mathcal{T}}^*_{\psi,n,1}(\lambda,h)\geq\mathcal{T}_n(\lambda)|\mathfrak{W}_n)$ as $\mathcal{M}\to\infty$. Now define $F_{n,\lambda}(c)\equiv P(\mathcal{T}_n(\lambda)\leq c)$, hence $p_n(\lambda,h)\equiv 1-\mathcal{F}_{\lambda,h}(\mathcal{T}_n(\lambda))$. $\{\mathcal{T}_n(\lambda):\lambda\in\Lambda\}\Rightarrow^*\{\mathcal{T}_\psi(\lambda,b):\lambda\in\Lambda\}$ by Theorem 4.2. By (A.15), conditionally on the sample draw $\mathfrak{W}_n\equiv\{(y_t,x_t)\}_{t=1}^n$, $\{\hat{\mathcal{T}}^*_{\psi,n,j}(\lambda,h)\}_{j=1}^{\mathcal{M}}$ is a sequence of i.i.d. draws from $\{\mathcal{T}_\psi(\lambda,h):\Lambda\}$, asymptotically with probability approaching one with respect to $\mathfrak{W}_n$. Since $\mathcal{T}_n(\lambda)$, and $\hat{\mathcal{T}}^*_{\psi,n,j}(\lambda,h)$ conditionally on $\mathfrak{W}_n$, have the same weak limits in probability under H_0, uniformly on Λ, it follows that $\sup_{c\geq0}|P(\hat{\mathcal{T}}^*_{\psi,n,j}(\lambda,h)\leq c|\mathfrak{W}_n)-F_{n,\lambda}(c)|\xrightarrow{p}0\forall\lambda\in\Lambda$ (Gine and Zinn, 1990, Sect. 3, eqn. (3.4) and (3.5)). Therefore, as claimed $\hat{p}^*_{n,\mathcal{M}_n}(\lambda,h)=1-F_{n,\lambda}(\mathcal{T}_n(\lambda))+o_p(1)=p_n(\lambda,h)+o_p(1)$ given $\mathcal{M}_n\to\infty$.

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

Claim (b). Recall $F_{n,\lambda}(c) \equiv P(\mathcal{T}_n(\lambda) \leq c)$ and $F^*_{n,\lambda,h}(c) \equiv P(\hat{\mathcal{T}}^*_{\psi,n,1}(\lambda,h) \leq c|\mathfrak{W}_n)$.

Step 1. In order to prove $\sup_{\lambda\in\Lambda} |\hat{p}^*_{n,\mathcal{M}_n}(\lambda,h) - p_n(\lambda,h)| \xrightarrow{p} 0$, it suffices to show:

$$\sup_{\lambda\in\Lambda} \left| \hat{p}^*_{n,\mathcal{M}}(\lambda,h) - P\left(\hat{\mathcal{T}}^*_{\psi,n,1}(\lambda,h) \geq \mathcal{T}_n(\lambda)|\mathfrak{W}_n\right)\right| \tag{A.21}$$

$$\sup_{\lambda\in\Lambda} \left| P\left(\hat{\mathcal{T}}^*_{\psi,n,1}(\lambda,h) \geq \mathcal{T}_n(\lambda)|\mathfrak{W}_n\right) - \left\{1 - F_{n,\lambda}(\mathcal{T}_n(\lambda))\right\}\right| \xrightarrow{p} 0. \tag{A.22}$$

Consider (A.21). Under (a), we have pointwise as $\mathcal{M} \to \infty$:

$$\hat{p}^*_{n,\mathcal{M}}(\lambda,h) - P\left(\hat{\mathcal{T}}^*_{\psi,n,1}(\lambda,h) \geq \mathcal{T}_n(\lambda)|\mathfrak{W}_n\right)$$
$$= \frac{1}{\mathcal{M}} \sum_{j=1}^{\mathcal{M}} \left\{ I\left(\hat{\mathcal{T}}^*_{\psi,n,j}(\lambda,h) > \mathcal{T}_n(\lambda)\right) - E\left[I\left(\hat{\mathcal{T}}^*_{\psi,n,j}(\lambda,h) > \mathcal{T}_n(\lambda)\right)|\mathfrak{W}_n\right]\right\} \xrightarrow{p} 0.$$

It remains to establish equicontinuity on Λ for a uniform Glivenko–Cantelli theorem (A.21) (van der Vaart and Wellner, 1996, Thm. 2.8.1). The $\mathcal{V}(\mathcal{C})$ class satisfies the required condition (van der Vaart and Wellner, 1996, p. 168). Therefore, we need only demonstrate $\{I(\hat{\mathcal{T}}^*_{\psi,n,j}(\lambda,h) > \mathcal{T}_n(\lambda)) - E[I(\hat{\mathcal{T}}^*_{\psi,n,j}(\lambda,h) > \mathcal{T}_n(\lambda))|\mathfrak{W}_n] : \lambda \in \Lambda\}$ lies in $\mathcal{V}(\mathcal{C})$. In the following, we use properties of $\mathcal{V}(\mathcal{C})$ functions without citation. See, for example, van der Vaart and Wellner (1996, Chap. 2.6), and see the discussion above Assumption 6.

The test weight $F(\cdot)$ is in $\mathcal{V}(\mathcal{C})$ under Assumption 6, and trivially $\{\lambda' W(x_t) : \lambda \in \Lambda\}$ lies in $\mathcal{V}(\mathcal{C})$, hence $\{F(\lambda' W(x_t)) : \lambda \in \Lambda\}$ lies in $\mathcal{V}(\mathcal{C})$. Therefore, $\{\hat{\mathcal{T}}^*_{\psi,n,j}(\lambda,h), \mathcal{T}_n(\lambda) : \lambda \in \Lambda\}$ are in $\mathcal{V}(\mathcal{C})$ because they involve ratios and products of linearly combined $\{F(\lambda' W(x_t))\}_{t=1}^n$. Therefore, $\{I(\hat{\mathcal{T}}^*_{\psi,n,j}(\lambda,h) > \mathcal{T}_n(\lambda)) : \lambda \in \Lambda\}$ is in $\mathcal{V}(\mathcal{C})$.

Next, by Assumption 6 $\{F^*_{n,\lambda,h}(c) : \lambda \in \Lambda, c \in [0,\infty)\}$ belongs to the $\mathcal{V}(\mathcal{C})$ class. Hence $\{E[I(\hat{\mathcal{T}}^*_{\psi,n,j}(\lambda,h) > \mathcal{T}_n(\lambda))|\mathfrak{W}_n] : \lambda \in \Lambda\}$ is in $\mathcal{V}(\mathcal{C})$, and therefore $\{I(\hat{\mathcal{T}}^*_{\psi,n,j}(\lambda,h) > \mathcal{T}_n(\lambda)) - E[I(\hat{\mathcal{T}}^*_{\psi,n,j}(\lambda,h) > \mathcal{T}_n(\lambda))|\mathfrak{W}_n] : \lambda \in \Lambda\}$ lies in $\mathcal{V}(\mathcal{C})$ as required, proving (A.21).

Now consider (A.22). Under (a) pointwise $P(\hat{\mathcal{T}}^*_{\psi,n,1}(\lambda,h) \geq \mathcal{T}_n(\lambda)|\mathfrak{W}_n) - \{1 - F_{n,\lambda}(\mathcal{T}_n(\lambda))\} \xrightarrow{p} 0$. We have from above that $\hat{\mathcal{T}}^*_{\psi,n,1}(\lambda,h)$ and $\mathcal{T}_n(\lambda)$ are in $\mathcal{V}(\mathcal{C})$. Under Assumption 6$\{F_{n,\lambda}(c), F^*_{n,\lambda,h}(c) : \lambda \in \Lambda, c \in [0,\infty)\}$ lie in $\mathcal{V}(\mathcal{C})$. Hence $\{P(\hat{\mathcal{T}}^*_{\psi,n,1}(\lambda,h) \geq \mathcal{T}_n(\lambda)|\mathfrak{W}_n) - \{1 - F_{n,\lambda}(\mathcal{T}_n(\lambda))\} : \lambda \in \Lambda\}$ lies in $\mathcal{V}(\mathcal{C})$, promoting the required uniform convergence (A.22) (cf. van der Vaart and Wellner, 1996, Theorem 2.8.1).

Step 2. Finally, we prove $AsySz^* \leq \alpha$. Consider the LF p-value $\hat{p}^{(LF)}_{n,\mathcal{M}_n}(\lambda)$. $AsySz^*$ can be written as:

$$AsySz^* = \sup_{\lambda\in\Lambda} \limsup_{n\to\infty} \sup_{\gamma\in\Gamma^*} P_\gamma \left(\max\left\{ \sup_{h\in\mathfrak{H}} \left\{\hat{p}^*_{n,\mathcal{M}_n}(\lambda,h)\right\}, p^\infty_n(\lambda)\right\} < \alpha | H_0\right)$$
$$= \sup_{\lambda\in\Lambda} \limsup_{n\to\infty} \sup_{\gamma\in\Gamma^*} P_\gamma \left(\sup_{h\in\mathfrak{H}} \left\{\hat{p}^*_{n,\mathcal{M}_n}(\lambda,h), \bar{\mathcal{F}}_\infty(\mathcal{T}_n(\lambda))\right\} < \alpha | H_0\right).$$

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

By Step 1 $\sup_{\lambda\in\Lambda}|\hat{p}^{*}_{n,\mathcal{M}_n}(\lambda,h)-p_n(\lambda,h)|\xrightarrow{p}0$, hence

$$AsySz^{*}=\sup_{\lambda\in\Lambda}\limsup_{n\to\infty}\sup_{\gamma\in\Gamma^{*}}P_{\gamma}\left(\sup_{h\in\mathfrak{H}}\left\{\bar{\mathcal{F}}_{\lambda,h}(\mathcal{T}_n(\lambda))+o_{p,\lambda}(1),\bar{\mathcal{F}}_{\infty}(\mathcal{T}_n(\lambda))\right\}<\alpha|H_0\right)\equiv\mathfrak{A}^{*},$$

say. By Theorem 4.2.a, $\{\mathcal{T}_n(\lambda):\Lambda\}\Rightarrow^{*}\{\mathcal{T}_{\psi}(\lambda,h):\Lambda\}$ under $\mathcal{C}(i,b)$ with $||b||<\infty$. Weak convergence implies convergence in finite dimensional distributions. By the definition of distribution convergence, and the mapping theorem, weak convergence therefore yields:

$$\mathfrak{A}^{*}=\sup_{\lambda\in\Lambda}\sup_{\tilde{h}\in\mathfrak{H}}P\left(\max\left\{\sup_{h\in\mathfrak{H}}\left\{\bar{\mathcal{F}}_{\lambda,h}\left(\mathcal{T}_{\psi}(\lambda,\tilde{h})\right)\right\},\bar{\mathcal{F}}_{\infty}\left(\mathcal{T}_{\psi}(\lambda,\tilde{h})\right)\right\}<\alpha\right).$$

The remainder of the proof under $\mathcal{C}(i,b)$ and $\mathcal{C}(ii,\omega_0)$, and for $\hat{p}^{(ICS-1)}_{n,\mathcal{M}_n}(\lambda)$, now follows directly from the proof of Theorem 6.1. $\mathcal{QED}$.

Proof of Theorem 6.3. Recall $\hat{p}^{(\cdot)}_{n,\mathcal{M}}(\lambda)$ is the LF or ICS-1 p-value computed with the weak identification p-value approximation $\hat{p}^{*}_{n,\mathcal{M}}(\lambda,h)$. Write $\hat{\mathcal{P}}_n(\alpha)=\hat{\mathcal{P}}_{n,\mathcal{M}_n}(\alpha)\equiv\int_{\Lambda}I(\hat{p}^{(\cdot)}_{n,\mathcal{M}_n}(\lambda)<\alpha)d\lambda$. Define the infeasible PVOT $\mathcal{P}^{(\cdot)}_n(\alpha)\equiv\int_{\Lambda}I(p^{(\cdot)}_n(\lambda)<\alpha)d\lambda$, where $p^{(\cdot)}_n(\lambda)$ is the infeasible LF or ICS-1 p-value (see Section 5).

By Theorem 6.2.a with $\mathcal{M}=\mathcal{M}_n\to\infty$ as $n\to\infty$, and Lebesgue's dominated convergence theorem: $\hat{\mathcal{P}}_n(\alpha)-\mathcal{P}_n(\alpha)=\int_{\Lambda}\{I(\hat{p}^{(\cdot)}_{n,\mathcal{M}_n}(\lambda)<\alpha)-I(p^{(\cdot)}_n(\lambda)<\alpha)\}d\lambda\xrightarrow{p}0$. It therefore suffices to prove the claim for $\mathcal{P}_n(\alpha)$.

Step 1 (H_0). Consider identification case $\mathcal{C}(i,b)$. The LF p-value satisfies $p^{(LF)}_n(\lambda)\geq p_n(\lambda,h)=\bar{\mathcal{F}}_{\lambda,h}(\mathcal{T}_n(\lambda))$ for any fixed h, hence in the LF case:

$$P\left(\mathcal{P}_n(\alpha)>\alpha\right)\leq P\left(\int_{\Lambda}I\left(\bar{\mathcal{F}}_{\lambda,h}(\mathcal{T}_n(\lambda))<\alpha\right)d\lambda>\alpha\right). \tag{A.23}$$

By Theorem 4.2.a and Assumption 5, and using the notation of Section 5, $\{\mathcal{T}_n(\lambda):\lambda\in\Lambda\}\Rightarrow^{*}\{\mathcal{T}_{\psi}(\lambda,h):\lambda\in\Lambda\}$. The limit process $\mathcal{T}_{\psi}(\lambda,h)$ satisfies Assumption 1.a in Hill (2018) , and $\bar{\mathcal{F}}_{\lambda,h}(\mathcal{T}_n(\lambda))$ trivially satisfies Assumption 1.b in Hill (2018). Hence, $\lim_{n\to\infty}P(\int_{\Lambda}I(\bar{\mathcal{F}}_{\lambda,h}(\mathcal{T}_n(\lambda))<\alpha)d\lambda>\alpha)\leq\alpha$ by Thm. 3.1 in Hill (2018). In view of (A.23), this proves $\lim_{n\to\infty}P(\mathcal{P}_n(\alpha)>\alpha)\leq\alpha$.

The ICS-1 p-value satisfies $p^{(ICS-1)}_n(\lambda)\geq p^{(LF)}_n(\lambda)I(\mathcal{A}_n\leq\kappa_n)\geq\bar{\mathcal{F}}_{\lambda,h}(\mathcal{T}_n(\lambda))I(\mathcal{A}_n\leq\kappa_n)$, hence in the ICS-1 case:

$$P\left(\mathcal{P}_n(\alpha)>\alpha\right)\leq P\left(\int_{\Lambda}I\left(\bar{\mathcal{F}}_{\lambda,h}(\mathcal{T}_n(\lambda))I\left(\mathcal{A}_n\leq\kappa_n\right)<\alpha\right)d\lambda>\alpha\right).$$

Assumption 1.a in Hill (2018) holds by the above argument. By the Cauchy–Schwartz inequality, Theorem 5.1.a and $\kappa_n\to\infty$:

$$\begin{aligned}E\left|\bar{\mathcal{F}}_{\lambda,h}(\mathcal{T}_n(\lambda))I\left(\mathcal{A}_n\leq\kappa_n\right)-\bar{\mathcal{F}}_{\lambda,h}(\mathcal{T}_n(\lambda))\right|&=E\left|\bar{\mathcal{F}}_{\lambda,h}(\mathcal{T}_n(\lambda))I\left(\mathcal{A}_n>\kappa_n\right)\right|\\&\leq\left(E\left[\bar{\mathcal{F}}^{2}_{\lambda,h}(\mathcal{T}_n(\lambda))\right]\right)^{1/2}P\left(\mathcal{A}_n>\kappa_n\right)^{1/2}\to0.\end{aligned}$$

Hence $\bar{\mathcal{F}}_{\lambda,h}(\mathcal{T}_n(\lambda))I(\mathcal{A}_n\leq\kappa_n)-\bar{\mathcal{F}}_{\lambda,h}(\mathcal{T}_n(\lambda))\xrightarrow{p}0$ by Markov's inequality, which verifies Assumption 1.b in Hill (2018). Thm. 3.1 in Hill (2018) now yields

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

$\lim_{n\to\infty} P(\int_\Lambda I(\bar{\mathcal{F}}_{\lambda,h}(\mathcal{T}_n(\lambda))I(\mathcal{A}_n \le \kappa_n) < \alpha)d\lambda > \alpha) \le \alpha$, proving $\lim_{n\to\infty} P(\hat{\mathcal{P}}_n(\alpha) > \alpha) \le \alpha$.

Under identification case $\mathcal{C}(ii,\omega_0)$, the claim follows from Theorem 4.2.b, and Thm. 3.1 in Hill (2018).

Step 2 (H_1). By Theorem 6.1.b, $p_n^{(\cdot)}(\lambda) \xrightarrow{p} 0 \forall \lambda \in \Lambda/S$ where S has Lebesgue measure zero. The claim now follows from Thm. 2.2.b in Hill (2018). $\mathcal{QED}$.

Proof of Theorem 6.4. Consider the LF case. By Theorem 6.2 $\sup_{\lambda\in\Lambda} |\hat{p}^*_{n,\mathcal{M}}(\lambda,h) - p_n(\lambda,h)| \xrightarrow{p} 0$, and by Theorem 4.2.a, $\{\mathcal{T}_n(\lambda) : \Lambda\} \Rightarrow^* \{\mathcal{T}_\psi(\lambda,h) : \Lambda\}$ under $\mathcal{C}(i,b)$ with $||b|| < \infty$. An application of the mapping theorem, compactness of Λ and dominated convergence yield:

$$
\begin{aligned}
AsySz(pvot) &= \limsup_{n\to\infty} \sup_{\gamma\in\Gamma^*} P_\gamma\left(\int_\Lambda I\left(\max\left\{\sup_{h\in\mathfrak{H}}\left\{\hat{p}^*_{n,\mathcal{M}_n}(\lambda,h)\right\}, p_n^\infty(\lambda)\right\} < \alpha\right)d\lambda > \alpha | H_0\right) \\
&= \limsup_{n\to\infty} \sup_{\gamma\in\Gamma^*} P_\gamma\left(\int_\Lambda I\left(\max\left\{\sup_{h\in\mathfrak{H}}\left\{\bar{\mathcal{F}}_{\lambda,h}(\mathcal{T}_n(\lambda))\right\}, p_n^\infty(\lambda)\right\} < \alpha\right)d\lambda + o_p(1) > \alpha | H_0\right) \\
&= \sup_{\tilde{h}\in\mathfrak{H}} P\left(\int_\Lambda I\left(\max\left\{\sup_{h\in\mathfrak{H}}\left\{\bar{\mathcal{F}}_{\lambda,h}(\mathcal{T}_\psi(\lambda,\tilde{h}))\right\}, p_n^\infty(\lambda)\right\} < \alpha\right)d\lambda > \alpha | H_0\right) \\
&\le \sup_{h\in\mathfrak{H}} P\left(\int_\Lambda I\left(\bar{\mathcal{F}}_{\lambda,h}(\mathcal{T}_\psi(\lambda,h)) < \alpha\right)d\lambda > \alpha | H_0\right).
\end{aligned}
$$

Since $\mathcal{T}_\psi(\lambda,h)$ is distributed $\bar{\mathcal{F}}_{\lambda,h}$ we have $P(\int_\Lambda I(\bar{\mathcal{F}}_{\lambda,h}(\mathcal{T}_\psi(\lambda,h)) < \alpha)d\lambda > \alpha | H_0) \le \alpha$ for all $h \in \mathfrak{H}$ (see the proof of Thm. 3.1 in Hill, 2018), hence $AsySz(pvot) \le \alpha$. The ICS-1 case, and under $\mathcal{C}(ii,\omega_0)$, follow similarly. $\mathcal{QED}$.

B APPENDIX : SUPPORTING RESULTS

The following supporting results are proved in the supplemental material Hill (2020, Appendix B). Let $\iota(A)$ and $\bar{\iota}(A)$ denote the minimum and maximum eigenvalues of matrix A.

All subsequent Gaussian processes have *almost surely* uniformly continuous and bounded sample paths, hence in many cases we just say *Gaussian process*. The first two lemmas cover uniform asymptotics for $\mathcal{G}_{\psi,n}(\theta)$ and $\widehat{\mathcal{H}}_{\psi,n}(\pi)$ under non- and weak identification.

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

LEMMA B.2. *Under $\mathcal{C}(i,b)$ and Assumption 1, $\sup_{\pi\in\Pi} ||\widehat{\mathcal{H}}_{\psi,n}(\pi) - \mathcal{H}_\psi(\pi)|| \xrightarrow{p} 0$, where $\iota(\mathcal{H}_\psi(\pi)) > 0$ and $\bar{\iota}(\mathcal{H}_\psi(\pi)) < \infty$ for each $\pi \in \Pi$.*

The next two lemmas are uniform asymptotics under semi-strong and strong identification.

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

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

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* $\mathcal{C}(ii,\omega_0)$ *and Assumption 1,* $\sqrt{n}\mathfrak{B}(\beta_n)^{-1}(\partial/\partial\theta)Q_n(\theta_n)\xrightarrow{d}\mathcal{G}_\theta$*, a zero-mean Gaussian law with a finite, positive definite covariance* $E[\mathcal{G}_\theta\mathcal{G}_\theta']$*, and has a version that has* almost surely *uniformly continuous and bounded sample paths. Moreover,* $E[\mathcal{G}_\theta\mathcal{G}_\theta']=\sigma_0^2E[d_{\theta,t}d_{\theta,t}']$ *under* H_0.

LEMMA B.5. *Under* $\mathcal{C}(ii,\omega_0)$ *and Assumption 1,* $\widehat{\mathcal{H}}_n\xrightarrow{p}\mathcal{H}_\theta$*, and* $\iota(\mathcal{H}_\theta)>0$ *and* $\bar{\iota}(\mathcal{H}_\theta)<\infty$.

Next, we tackle general versions of $\widehat{\mathcal{H}}_n$ and $\hat{\mathcal{V}}_n$ in (2) and (15) that are required for uniform asymptotics. Recall we use $\hat{\mathcal{V}}_n$ for the ICS statistic $\mathcal{A}_n\equiv(k_\beta^{-1}n\hat{\beta}_n'\hat{\Sigma}_{\beta,\beta,n}^{-1}\hat{\beta}_n)^{1/2}$ in (14), where $\hat{\Sigma}_{\beta,\beta,n}$ is the upper $k_\beta\times k_\beta$ block of $\hat{\Sigma}_n\equiv\widehat{\mathcal{H}}_n^{-1}\hat{\mathcal{V}}_n\widehat{\mathcal{H}}_n^{-1}$.

Define the augmented parameter and space $\theta^+\equiv[||\beta||,\omega',\zeta',\pi']'\in\Theta^+\equiv\{\theta^+\in\mathbb{R}^{k_x+k_\beta+k_\pi+1}:\theta^+=[||\beta||,\omega(\beta),\zeta,\pi]'\colon\beta\in\mathcal{B},\zeta\in\mathcal{Z}(\beta),\pi\in\Pi\}$. Define $\epsilon_t(\theta^+)\equiv y_t-\zeta'x_t-||\beta||\omega'g(x_t,\pi)$, and:

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

Hence $\widehat{\mathcal{H}}_n(\hat{\theta}_n^+)=\widehat{\mathcal{H}}_n$ and $\hat{\mathcal{V}}_n(\hat{\theta}_n^+)=\hat{\mathcal{V}}_n$. Define $\mathcal{H}_\theta(\theta^+)\equiv E[d_{\theta,t}(\omega,\pi)d_{\theta,t}(\omega,\pi)']$ and $\mathcal{V}(\theta^+)\equiv E[\epsilon_t^2(\theta^+)d_{\theta,t}(\omega,\pi)d_{\theta,t}(\omega,\pi)']$. In the interest of decreasing notation, we use the same argument θ^+ for both $\widehat{\mathcal{H}}_n(\theta^+)$ and $\hat{\mathcal{V}}_n(\theta^+)$, although $\widehat{\mathcal{H}}_n(\theta^+)$ only depends on $(\omega(\beta),\pi)$.

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

In order to characterize the weak limit of $\hat{\pi}_n$, we need the following results. Andrews and Cheng (2012a) exploit the following normalizing constants for the criterion derivative under $\mathcal{C}(i,b)$:

$$a_n=\begin{cases}\sqrt{n} & \text{if } \mathcal{C}(i,b) \text{ and } \|b\|<\infty\\ \|\beta_n\|^{-1} & \text{if } \mathcal{C}(i,b) \text{ and } \|b\|=\infty\end{cases}$$

By construction $||\beta_n||^{-1}\leq\sqrt{n}$ for large n when $\sqrt{n}||\beta_n||\to\infty$ hence $a_n\leq\sqrt{n}$ for large n when $||b||=\infty$. Recall $\psi_{0,n}\equiv[0_{k_\beta}',\zeta_0']'$ hence $Q_{0,n}\equiv Q_n(\psi_{0,n},\pi)$ does not depend on π.

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

Now define:

$$\mathcal{Z}_n(\pi) = -a_n \widehat{\mathcal{H}}_{\psi,n}^{-1}(\pi) \frac{\partial}{\partial \psi} Q_n(\psi_{0,n}, \pi).$$

Under $\mathcal{C}(i,b)$, $\widehat{\mathcal{H}}_{\psi,n}(\pi)$ is positive definite uniformly on Π, asymptotically with probability approaching one. See Lemma B.2. Write $Q_n^c(\pi) \equiv Q_n(\hat{\psi}_n(\pi), \pi)$.

LEMMA B.7. *Let drift case $\mathcal{C}(i,b)$ and Assumption 1 hold.*

a. *In general* $a_n(\hat{\psi}_n(\pi) - \psi_{0,n}) = \mathcal{Z}_n(\pi)$.

b. $a_n^2\{Q_n^c(\pi) - Q_{0,n}\} = -2^{-1}\mathcal{Z}_n(\pi)'\widehat{\mathcal{H}}_{\psi,n}(\pi)\mathcal{Z}_n(\pi)$ *where* $Q_{0,n} \equiv Q_n(\psi_{0,n}, \pi)$.

Define $\vartheta_\psi(\pi, \omega_0) \equiv -2^{-2}\omega_0' \mathcal{D}_\psi(\pi)' \mathcal{H}_\psi^{-1}(\pi) \mathcal{D}_\psi(\pi)\omega_0$. The following is a key result for characterizing the asymptotic properties of $\hat{\pi}_n$ under weak identification.

LEMMA B.8. *Let drift case $\mathcal{C}(i,b)$ 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* $\omega_0 \in \mathbb{R}^{k_\beta}$, $||\omega_0|| = 1$, *then* $\sup_{\pi\in\Pi} |(1/||\beta_n||^2)$ $(Q_n^c(\pi) - Q_{0,n}) - \vartheta_\psi(\pi, \omega_0)| \xrightarrow{p} 0$.

Write $\epsilon_t(\psi, \pi) = y_t - \zeta' x_t - \beta' g(x_t, \pi)$. Recall ψ_n is the (possibly drifting) true value under H_0.

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

a. *Under $\mathcal{C}(i,b)$ with $||b|| < \infty$:*

$$\left\{ \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left\{ \epsilon_t(\psi_n, \pi)\mathcal{K}_{\psi,t}(\pi, \lambda) - E\left[\epsilon_t(\psi_n, \pi)\mathcal{K}_{\psi,t}(\pi, \lambda)\right]\right\} : \Pi, \Lambda \right\}$$
$$\Rightarrow^* \left\{ \mathfrak{Z}_\psi(\pi, \lambda) : \Pi, \Lambda \right\},$$

a zero-mean Gaussian process with covariance kernel $E[\mathfrak{Z}_\psi(\pi, \lambda)\mathfrak{Z}_\psi(\tilde{\pi}, \tilde{\lambda})]$. *Under H_0,*

$$\sup_{\pi\in\Pi, \lambda\in\Lambda} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left\{ \epsilon_t(\psi_n, \pi)\mathcal{K}_{\psi,t}(\pi, \lambda) - E\left[\epsilon_t(\psi_n, \pi)\mathcal{K}_{\psi,t}(\pi, \lambda)\right]\right\} \right.$$
$$\left. - \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \epsilon_t \mathcal{K}_{\psi,t}(\pi, \lambda) \right| \xrightarrow{p} 0,$$

and $E[\mathfrak{Z}_\psi(\pi, \lambda)\mathfrak{Z}_\psi(\tilde{\pi}, \tilde{\lambda})] = \sigma_0^2 E[\mathcal{K}_{\psi,t}(\pi, \lambda)\mathcal{K}_{\psi,t}(\tilde{\pi}, \tilde{\lambda})]$.

b. *Under* $\mathcal{C}(i, \omega_0)$, $\{1/\sqrt{n}\sum_{t=1}^{n} \epsilon_t \mathcal{K}_{\theta,t}(\lambda) : \lambda \in \Lambda\} \Rightarrow^* \{\mathfrak{Z}_\theta : \lambda \in \Lambda\}$, *a zero-mean Gaussian process with covariance* $E[\mathfrak{Z}_\theta(\lambda)\mathfrak{Z}_\theta\tilde{(\lambda)}] = E[\epsilon_t^2 \mathcal{K}_{\theta,t}(\lambda)\mathcal{K}_{\theta,t}(\tilde{\lambda})]$.

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press

LEMMA B.10. *Under Assumption 1,* $\sup_{\omega\in\mathbb{R}^{k_\beta}:||\omega||=1,\pi\in\Pi,\lambda\in\Lambda}||\hat{\mathfrak{b}}_{\theta,n}(\omega,\pi,\lambda)-\mathfrak{b}_\theta(\omega,\pi,\lambda)||\xrightarrow{p}0$ *and* $\sup_{\pi\in\Pi,\lambda\in\Lambda}||\hat{\mathfrak{b}}_{\psi,n}(\pi,\lambda)-\mathfrak{b}_\psi(\pi,\lambda)||\xrightarrow{p}0$.

LEMMA B.11. *Under Assumption 1* $\sup_{\theta^+\in\Theta^+,\lambda\in\Lambda}||\hat{v}_n^2(\theta^+,\lambda)-v^2(\theta^+,\lambda)||\xrightarrow{p}0$ *and* $\sup_{\theta\in\Theta,\lambda\in\Lambda}||\hat{v}_n^2(\theta,\lambda)-v^2(\theta,\lambda)||\xrightarrow{p}0$.

Define $v^2(\lambda)=v^2(\omega_0,\pi_0,\lambda)$ where:

$$v^2(\omega,\pi,\lambda)\equiv E\left[\epsilon_t^2(\psi_0,\pi)\left\{F\left(\lambda'\mathcal{W}(x_t)\right)-\mathfrak{b}_\theta(\omega,\pi,\lambda)'\mathcal{H}_\theta^{-1}(\omega,\pi)d_{\theta,t}(\omega,\pi)\right\}^2\right].$$

LEMMA B.12. *Let Assumptions 1.a(i) and 3 hold. Under* $\mathcal{C}(i,b)$ *with* $||b||<\infty$, *the set* $\{\lambda\in\Lambda:\inf_{\omega'\omega=1,\pi\in\Pi}v^2(\omega,\pi,\lambda)=0\}$ *has Lebesgue measure zero. Under* $\mathcal{C}(ii,\omega_0)$, *the set* $\{\lambda\in\Lambda:v^2(\lambda)=0\}$ *has Lebesgue measure zero.*

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

LEMMA B.13. *Under Assumption 1,* $\sup_{\pi\in\Pi,\lambda\in\Lambda}|1/n\sum_{t=1}^n\epsilon_tF(\lambda'\mathcal{W}(x_t))-E[\epsilon_tF(\lambda'\mathcal{W}(x_t))]|\xrightarrow{p}0$, $\sup_{\pi\in\Pi,\lambda\in\Lambda}||1/n\sum_{t=1}^n\mathcal{M}_t(\pi,\lambda)-E[\mathcal{M}_t(\pi,\lambda)]||\xrightarrow{p}0$ *and* $\sup_{\pi\in\Pi,\lambda\in\Lambda}||E[\mathcal{M}_t(\pi,\lambda)]||<\infty$, *and* $\sup_{\pi\in\Pi}||1/n\sum_{t=1}^n\tilde{\mathcal{M}}_t(\pi)-E[\tilde{\mathcal{M}}_t(\pi)]||\xrightarrow{p}0$ *and* $\sup_{\pi\in\Pi}||E[\tilde{\mathcal{M}}_t(\pi)]||<\infty$.

https://doi.org/10.1017/S0266466620000201 Published online by Cambridge University Press