A Smoothed P-Value Test When There is a Nuisance Parameter under the Alternative

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

April 11, 2022

Abstract

We present a new test when there is a nuisance parameter λ under the alternative hypothesis. The test exploits the p-value occupation time [PVOT], the measure of the subset of λ on which a p-value test based on a test statistic $\mathcal{T}_n(\lambda)$ rejects the null hypothesis. Key contributions are: (i) An asymptotic critical value upper bound for our test is the significance level α, making inference easy. (ii) We only require $\mathcal{T}_n(\lambda)$ to have a known or bootstrappable limit distribution, hence we do not require $\sqrt{n}$-Gaussian asymptotics, and weak or non-identification is allowed. (iii) A test based on the transformed p-value $\sup_{\lambda \in \Lambda} p_n(\lambda)$ may be conservative and in some cases have nearly trivial power, while the PVOT naturally controls for this by smoothing over the nuisance parameter space. Finally, (iv) the PVOT uniquely allows for bootstrap inference in the presence of nuisance parameters when some estimated parameters may not be identified.

Key words and phrases: p-value test, empirical process test, nuisance parameter, weighted average power, GARCH test, omitted nonlinearity test.
AMS classifications : 62G10, 62M99, 62F35.

1 Introduction

This paper develops a test for cases when a nuisance parameter $\lambda \in \mathbb{R}^k$ is present under the alternative hypothesis H_1, where $k \geq 1$ is finite. Let $\mathcal{Y}_n \equiv \{y_t\}_{t=1}^n$ be the observed sample of data with sample size $n \geq 1$, and let $\mathcal{T}_n(\lambda) \equiv \mathcal{T}(\mathcal{Y}_n, \lambda)$ be a test statistic function of λ for testing a hypothesis H_0 about the data $\mathcal{Y}_n$ against H_1. We assume $\mathcal{T}_n(\lambda) \geq 0$, and that large values are indicative of H_1. We present a simple smoothed p-value test based on the Lebesgue measure of the subset of $\lambda' s$ on which we reject H_0 based on $\mathcal{T}_n(\lambda)$, defined as the *P-Value Occupation Time* [PVOT]. In order to focus ideas, we ignore cases where λ may be a set, interval, or function, or infinite dimensional as in nonparametric estimation problems.

The PVOT was originally explored in Hill and Aguilar (2013) and Hill (2012) as a way to gain inference in the presence of a trimming tuning parameter. We extend the idea to test problems where λ is a nuisance parameter under H_1. We provide a complete asymptotic theory here for the first time.

*Dept. of Economics, University of North Carolina, Chapel Hill; jbhill@email.unc.edu; https://jbhill.web.unc.edu.

Nuisance parameters under H_1 arise in two over-lapping cases. First, λ is part of the data generating process under H_1, e.g. ARMA models with common roots (Andrews and Cheng, 2012); tests of no GARCH effects (Engle, Lilien, and Robins, 1987; Andrews, 2001); tests for common factors (Andrews and Ploberger, 1994); tests for a Box-Cox transformation (Aguirre-Torres and Gallant, 1983); and structural change tests (Andrews, 1993). A standard example is the regression $y_t = \beta_0' x_t + \gamma_0 h(\lambda, x_t) + \epsilon_t$ where x_t are covariates, and $E[\epsilon_t | x_t] = 0$ *a.s.* for unique (β_0, γ_0). If $H_0 : \gamma_0 = 0$ is true then λ is not identified. This class includes the Box-Cox transform, neural networks, flexible functional forms, and regime switching models. See, e.g., Gallant (1981, 1984), Gallant and Golub (1984), White (1989), Andrews and Ploberger (1994), Terasvirta (1994), Hansen (1996) and Andrews and Cheng (2012).

Second, λ is used to compute an estimator, or perform a general model specification test, hence we can only say $\mathcal{Y}_n$ has the joint distribution $f(y, \theta_0)$ under H_0. This includes tests of omitted nonlinearity against general alternatives (White, 1989; Bierens, 1990; Bierens and Ploberger, 1997; Stinchcombe and White, 1998; Hill, 2012); and tests of marginal effects in models with mixed frequency data where λ is used to reduce regressor dimensionality (Ghysels, Santa-Clara, and Valkanov, 2004; Ghysels, Hill, and Motegi, 2016). An example is the regression $y_t = \beta_0' x_t + \epsilon_t$ where we want to test $H_0 : E[\epsilon_t | x_t] = 0$ *a.s.* This is fundamentally different from the preceding example where $E[\epsilon_t | x_t] = 0$ *a.s.* is assumed. A test statistic can be based on the fact that $E[\epsilon_t F(\lambda' x_t)] \neq 0$ *if and only if* $E[\epsilon_t | x_t] = 0$ *a.s.* is false, for all $\lambda \in \Lambda$ outside of a measure zero subset, provided $F : \mathbb{R} \rightarrow \mathbb{R}$ is exponential (Bierens, 1990), logistic (White, 1989), or any real analytic non-polynomial (Stinchcombe and White, 1998), or multinomials of x_t (Bierens, 1982). Notice that λ need not be part of the data generating process since $E[y_t | x_t] = \beta_0' x_t + \gamma_0 F(\lambda' x_t)$ *a.s.* may not be true under H_1.

A classic approach for handling nuisance parameters in the broad sense is to compute a p-value $p_n(\lambda)$ and use $\sup_{\lambda \in \Lambda} p_n(\lambda)$ (see Lehmann, 1994, Chap. 3.1). This may lead to a conservative test, although it promotes a test with the correct asymptotic level.[1] Further, $\sup_{\lambda \in \Lambda} p_n(\lambda)$ may not promote a consistent test even when $\mathcal{T}_n(\lambda)$ and its transforms like $\sup_{\lambda \in \Lambda} \mathcal{T}_n(\lambda)$ do. An example is a Bierens (1990)-type test which is known to be consistent $\forall \lambda \in \Lambda / S$ where S has measure zero. This means $\sup_{\lambda \in \Lambda} p_n(\lambda) \stackrel{p}{\rightarrow} (0, 1)$ under H_1 is possible despite $p_n(\lambda) \stackrel{p}{\rightarrow} 0 \ \forall \lambda \in \Lambda / S$. We find the test where H_0 is rejected at nominal level α when $\sup_{\lambda \in \Lambda} p_n(\lambda) < \alpha$ leads to profound size distortions and trivial power for a test of GARCH effects, and is relatively conservative as a test of omitted nonlinearity. In the case where λ is identified under either hypothesis, Silvapulle (1996) proposes an improvement with better size and power properties.

[1] Let $\tau_n \in [0, 1]$ be a test statistic, and suppose we reject a null hypothesis at nominal significance level α when $\tau_n > \alpha$. Recall that the asymptotic *level* of the test is α if $\lim_{n \rightarrow \infty} P(\tau_n > \alpha | H_0) \leq \alpha$, and if $\lim_{n \rightarrow \infty} P(\tau_n > \alpha | H_0) = \alpha$ then α is the asymptotic size (cf. Lehmann, 1994).

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). Recent contributions include Nyblom (1989), Andrews (1993), Dufour (1997), Andrews and Ploberger (1994, 1995), Hansen (1996), and Andrews and Cheng (2012, 2013, 2014). Nuisance parameters that are not identified under H_1 are either chosen at random, thereby sacrificing power (e.g. White, 1989); or $\mathcal{T}_n(\lambda)$ is smoothed over Λ, resulting in a non-standard limit distribution and in general the necessity of a bootstrap step (e.g. Chernoff and Zacks, 1964; Davies, 1977; Andrews and Ploberger, 1994). Examples 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 an absolutely continuous probability measure (Chernoff and Zacks, 1964; Davies, 1977; Andrews and Ploberger, 1994). See below for a discussion of power optimality of these transforms. The non-standard limit distribution, moreover, cannot be bootstrapped using conventional methods when some parameters may be weakly or non-identified. See Hill (2020b), and see below for discussion.

Let $p_n(\lambda)$ be a p-value or asymptotic p-value based on $\mathcal{T}_n(\lambda)$: $p_n(\lambda)$ may be based on a known limit distribution, or if the limit distribution is non-standard then a bootstrap or simulation method is assumed available for computing an asymptotically valid approximation to $p_n(\lambda)$. Assume that $p_n(\lambda)$ leads to an asymptotically correctly sized test, uniformly on Λ:

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

If $p_n(\lambda)$ is uniformly distributed then α is the size of the test, else by (1) α is the asymptotic size. The terms "asymptotic p-value" and "asymptotic size" are correct when convergence in (1) is uniform over H_0. The latter is not possible here because for generality we do not specify a model or H_0. If $p_n(\lambda)$ is asymptotically free of any other nuisance parameters then uniform convergence over H_0 is immediate given that (1) is uniform over Λ (e.g. Hansen, 1996, p. 417). Since this problem is common, we will not focus on it, and will simply call $p_n(\lambda)$ a "p-value" for brevity.

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

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

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

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

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

$\mathcal{P}_n^*(\alpha)$ is just the Lebesgue measure of the subset of $\lambda's$ on which we reject H_0. Thus, a large occupation time in the rejection region asymptotically indicates H_0 is false.

As long as $\{\mathcal{T}_n(\lambda) : \lambda \in \Lambda\}$ converges weakly under H_0 to a stochastic process $\{\mathcal{T}(\lambda) : \lambda \in \Lambda\}$, and $\mathcal{T}(\lambda)$ has a continuous distribution for all λ outside a set of measure zero, then asymptotically $\mathcal{P}_n^*(\alpha)$ has a mean α and the probability that $\mathcal{P}_n^*(\alpha) > \alpha$ is not greater than α. Evidence against H_0 is therefore simply $\mathcal{P}_n^*(\alpha) > \alpha$. Further, if asymptotically with probability approaching one the PV test (2) rejects H_0 for each λ in a subset of Λ that has Lebesgue measure greater than α, then $\mathcal{P}_n^*(\alpha) > \alpha$ asymptotically with probability one. The PVOT test at the chosen level α is then:

$$\textbf{PVOT Test}\text{: reject } H_0 \text{ if } \mathcal{P}_n^*(\alpha) > \alpha\text{, otherwise fail to reject } H_0. \tag{4}$$

These results are formally derived in Section 2. Thus, an asymptotic level α critical value is simply α, a useful simplification over transforms with non-standard asymptotic distributions, like $\int_\Lambda \mathcal{T}_n(\lambda)\mu(d\lambda)$ and $\sup_{\lambda\in\Lambda} \mathcal{T}_n(\lambda)$. A simulation study in Section 5 suggests the critical value α leads to an asymptotically correctly sized test for tests of omitted nonlinearity and GARCH effects, and strong power in each case. We may therefore expect that similar tests have this property.

The PVOT yields several useful innovations. First, when $\mathcal{T}_n(\lambda)$ is derived from a regression model in which some parameters may be weakly or non-identified, there is no known valid bootstrap or simulation approach for approximating the limit distribution of smoothed test statistics in the class considered in Andrews and Ploberger (1994), including $\int_\Lambda \mathcal{T}_n(\lambda)\mu(d\lambda)$ and $\sup_{\lambda\in\Lambda} \mathcal{T}_n(\lambda)$. This is because a valid bootstrap, for example, must approximate the covariance structure of the limit process $\{\mathcal{T}(\lambda) : \lambda \in \Lambda\}$ which generally requires consistent estimates of model parameters. If some parameters are weakly or non-identified, then they cannot be consistently estimated (see, e.g., Andrews and Cheng, 2012). See also Gallant (1977) for an early contribution to this literature. Hill (2020b) presents an asymptotically valid bootstrap method for the non-smoothed $\mathcal{T}_n(\lambda)$ for any degree of (non)identification. The resulting bootstrapped p-value leads to a valid smoothed p-value test, even though smoothed test statistics *cannot* be consistently bootstrapped. See Section 4.1.

Second, since the PVOT critical value upper bound is simply α under any asymptotic theory for $\mathcal{T}_n(\lambda)$, we only require $\mathcal{T}_n(\lambda)$ to have a known or bootstrappable limit distribution. Thus, $\sqrt{n}$-Gaussian asymptotics is not required as is nearly always assumed (e.g. Andrews and Ploberger, 1994; Hansen, 1996; Andrews and Cheng, 2012). Non-standard asymptotics are therefore allowed, including test statistics when a parameter is weakly identified, GARCH tests (e.g. Andrews, 2001); inference under heavy tails; and non-$\sqrt{n}$ asymptotics are covered, as in heavy tail robust tests (e.g. Hill, 2012; Hill and Aguilar, 2013; Aguilar and Hill, 2015), or when infill asymptotics or nonparametric estimators are involved (e.g. Bandi and Phillips, 2007).

Third, the local power properties of specific PVOT tests appear to be on par with the power optimal exponential class developed in Andrews and Ploberger (1994). We derive general results, and apply them to a test of omitted nonlinearity. We show in a numerical experiment that the PVOT test achieves local power on par with the highest achievable (weighted average) power. In view of the general result, the local power merits of the PVOT test appear to extend to any consistent test on Λ, but any such claim requires a specific test statistic and numerical exercise to verify.

Fourth, although we focus on the PVOT test, in Appendix B of the supplemental material Hill (2020a) we show the PVOT naturally arises as a measure of test optimality when λ is part of the true data generating process under H_1. This requires Andrews and Ploberger's 1994 notion of weighted average local power of a test based on $\mathcal{T}_n(\lambda)$, where the average is computed over λ and a drift parameter (cf. Wald, 1943). In this environment, the PVOT is just a point estimate of the weighted average probability of PV test rejection evaluated under H_0. Since that probability is asymptotically no larger than α when the null is true, the PVOT test rejects H_0 when the PVOT is larger than α. Thus, the PVOT is a natural way to transform a test statistic.

The relevant literature also includes King and Shively (1993) whose re-parameterization leads to a conventional, but not general, test. Hansen (1996) presents a wild bootstrap for computing the p-value for a smoothed LM statistic when λ is part of the data generating process, extending ideas in Wu (1986) and Liu (1988). The method implicitly requires strong identification of regression model parameters. See Shao (2010) for a dependent wild bootstrap method. Our simulation study for tests of functional form and GARCH effects show the PVOT test performs on par with, or is better than, the average and supremum test. Moreover, when model parameters are weakly or non-identified, a PVOT test of functional form substantially dominates $p_n(\lambda^*)$ with randomized λ^*, $\sup_{\lambda\in\Lambda} p_n(\lambda)$, and bootstrapped $\sup_{\lambda\in\Lambda} \mathcal{T}_n(\lambda)$ and $\int_\Lambda \mathcal{T}_n(\lambda)\mu(d\lambda)$. Indeed, the latter two fail to be valid for the reasons explained above.

Bierens (1990) compares supremum and pointwise statistics to achieve standard asymptotics for a functional form test. Bierens and Ploberger (1997) compute a critical value upper bound. We show that the upper bound leads to an under-sized test and potentially low power in a local power numerical exercise and a simulation study.

The remainder of the paper is as follows. We present the formal list of assumptions and the main results for the PVOT test in Section 2, and local power is analyzed in Section 3. Sections 3.2 and 4 contain examples in which we apply the PVOT test to tests of omitted nonlinearity (with possibly weakly identified parameters), and GARCH effects. We perform a simulation study in Section 5 involving these tests. Concluding remarks are left for Section 6.

2 Asymptotic Theory

The following notation is used. $[z]$ rounds z to the nearest integer. $a_n/b_n \sim c$ implies $a_n/b_n \to c$ as $n \to \infty$. $|\cdot|$ is the l_1-matrix norm, and $||\cdot||$ is the Euclidean norm, unless otherwise noted.

We require a notion of weak convergence that can handle a range of applications. A fundamental concern is that the mapping $\mathcal{T}_n : \Lambda \to [0,\infty)$ is not here defined, making measurability of $\{\mathcal{T}_n(\lambda) : \lambda \in \Lambda\}$ and its transforms like $\sup_{\lambda\in\Lambda} \mathcal{T}_n(\lambda)$ a challenge. We therefore assume all random variables in this paper exist on a complete measure space, and probabilities where applicable are outer probability measures. See Pollard's (1984: Appendix C) permissibility criteria, and see Dudley (1978) for related ideas.

We use weak convergence on $l_\infty(\Lambda)$, the space of bounded functions with sup-norm topology, in the sense of Hoffman-Jϕrgensen (1991):

$$\{\mathcal{T}_n(\lambda)\} \Rightarrow^* \{\mathcal{T}(\lambda)\} \text{ in } l_\infty(\Lambda), \text{ where } \{\mathcal{T}_n(\lambda)\} = \{\mathcal{T}_n(\lambda) : \lambda \in \Lambda\}, \text{ etc.}$$

If, for instance, the sample is $\mathcal{Y}_n \equiv \{y_t\}_{t=1}^n \in \mathbb{R}^{nk}$, and $\mathcal{T}_n(\lambda)$ is a measurable mapping $h(\mathcal{Z}(\mathcal{Y}_n, \lambda))$ of a function $\mathcal{Z} : \mathbb{R}^{nk} \times \Lambda \to \mathbb{R}$, then $h(\mathcal{Z}(y,\lambda)) \in l_\infty(\Lambda)$ requires the uniform bound $\sup_{\lambda\in\Lambda} |h(\mathcal{Z}(y,\lambda))| < \infty$ for each $y \in \mathbb{R}^{nk}$.[2] Sufficient conditions for weak convergence to a Gaussian process, for example, are convergence in finite dimensional distributions, and stochastic equicontinuity: $\forall\epsilon > 0$ and $\eta > 0$ there exists $\delta > 0$ such that $\lim_{n\to\infty} P(\sup_{||\lambda-\tilde{\lambda}||\leq\delta} |\mathcal{T}_n(\lambda) - \mathcal{T}_n(\tilde{\lambda})| > \eta) < \epsilon$. Consult, e.g., Dudley (1978), Gine and Zinn (1984), and Pollard (1984).

A large variety of test statistics are known to converge weakly under regularity conditions. In many cases $\mathcal{T}_n(\lambda)$ is a continuous function $h(\mathcal{Z}_n(\lambda))$ of a sequence of sample mappings $\{\mathcal{Z}_n(\lambda)\}_{n\geq1}$ such that $\sup_{x\in A} |h(x)| < \infty$ on every compact subset $A \subset \mathbb{R}$, and $\{\mathcal{Z}_n(\lambda)\} \Rightarrow^* \{\mathcal{Z}(\lambda)\}$ a Gaussian process. Two examples of h are $h(x) = x^2$ for asymptotic chi-squared tests of functional form or structural change; or $h(x) = \max\{0, x\}$ for a GARCH test (Andrews, 2001).

A *version* is a process with the same finite dimensional distributions. If $\{\mathcal{Z}(\lambda)\}$ is Gaussian, then any other Gaussian process with the same mean $E[\mathcal{Z}(\lambda)]$ and covariance kernel $E[\mathcal{Z}(\lambda_1)\mathcal{Z}(\lambda_2)]$ is a version of $\{\mathcal{Z}(\lambda)\}$.[3]

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

a. $\{\mathcal{T}_n(\lambda)\} \Rightarrow^* \{\mathcal{T}(\lambda)\}$, *a process with a version that has* almost surely *uniformly continuous sample*

[2] If more details are available, then boundedness can be refined. For example, if $\mathcal{T}_n(\lambda) = (1/\sqrt{n}\sum_{t=1}^n z(y_t,\lambda))^2$ where $z : \mathbb{R} \times \Lambda \to \mathbb{R}$, then we need $\sup_{\lambda\in\Lambda} |z(y,\lambda)| < \infty$ for each y.

[3] Even in the Gaussian case it is not true that all versions have continuous sample paths, but if a version of $\{\mathcal{Z}(\lambda)\}$ has continuous paths then this is enough to apply the continuous mapping theorem to transforms of $\mathcal{Z}_n(\lambda)$ over Λ. See Dudley (1967, 1978).

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

b. $\sup_{\lambda\in\Lambda}|p_n(\lambda) - \bar{F}_0(\mathcal{T}_n(\lambda))| \stackrel{p}{\to} 0$*, where* $\bar{F}_0(c) \equiv P(\mathcal{T}(\lambda) > c)$*.*

Remark 1. (a) is broadly applicable (see Section 4), while continuity of the distribution of $\mathcal{T}(\lambda)$ and (b) ensure $p_n(\lambda)$ has asymptotically a uniform limit distribution under H_0. This is mild since often $\mathcal{T}_n(\lambda)$ is a continuous transformation of a standardized sample analogue to a population moment. In a great variety of settings for stationary processes, for example, a standardized sample moment has a Gaussian or stable distribution limit, or converges to a function of a Gaussian or stable law. See Gine and Zinn (1984) and Pollard (1984) for weak convergence to stochastic processes, exemplified with Gaussian functional asymptotics, and see Bartkiewicz, Jakubowski, Mikosch, and Wintenberger (2010) for weak convergence to a Stable process for a (possibly dependent) heavy tailed process.

Remark 2. (b) is required when $p_n(\lambda)$ is not computed as the asymptotic p-value $\bar{F}_0(\mathcal{T}_n(\lambda))$, for example when a simulation or bootstrap method is used because $\bar{F}_0$ is unknown or a better small sample approximation is desired. Thus, in order to obtain lower level conditions we need to know how $p_n(\lambda)$ was computed. In Section 4.1, for example, we use Hill's (2020b) weak identification robust bootstrap method for p-value computation; and in Section 4.2 we use Andrews' (2001) simulation method for p-value computation for a GARCH test.

All proofs are presented in Appendix A.

Theorem 2.1. *Let Assumption 1 hold.*

a. In general $\lim_{n\to\infty} P(\mathcal{P}_n^*(\alpha) > \alpha) \leq \alpha$*.*

b. The asymptotic size is exactly $\lim_{n\to\infty} P(\mathcal{P}_n^*(\alpha) > \alpha) = \alpha$ *when*$\mathcal{T}(\lambda) = \mathcal{T}(\lambda^*) =$ *a.s.* $\forall\lambda \in \Lambda$ *and some* $\lambda^* \in \Lambda$ *.Moreover,* $\mathcal{P}_n^*(\alpha) = \alpha$ *a.s. if any h-tuple* $\{\mathcal{T}(\lambda_1), ..., \mathcal{T}(\lambda_h)\}$ *is jointly independent,* $\lambda_i \neq \lambda_j$ *for each* $i \neq j$*, and any* $h \in \mathbb{N}$*.*

c. $\lim_{n\to\infty} P(\mathcal{P}_n^*(\alpha) > \alpha) > 0$ *under the following condition:* $\{\bar{F}_0(\mathcal{T}(\lambda))\}$ *is weakly dependent in the sense that* $P(\bar{F}_0(\mathcal{T}(\lambda)) < \alpha, \bar{F}_0(\mathcal{T}(\tilde{\lambda})) < \alpha) > \alpha^2$ *for each couplet* $\{\lambda, \tilde{\lambda}\}$ *on a subset of* $\Lambda \times \Lambda$ *with positive measure.*

Remark 3. Under H_0 there is asymptotically a probability α we reject at any λ. The above theorem proves this implies asymptotically no more than an α portion of all $\lambda's$ lead to a rejection.

Remark 4. In general the asymptotic *level* of the test is α when the critical value is itself α (Lehmann, 1994, eq. (3.1)). The proof reveals polemic cases: (i) if every h-tuple $\{\mathcal{T}(\lambda_1), ..., \mathcal{T}(\lambda_h)\}$ of the limit process is jointly *independent*, $\lambda_i \neq \lambda_j$ $\forall i \neq j$, then the PVOT $\mathcal{P}_n^*(\alpha) \stackrel{d}{\to} \alpha$ hence $\lim_{n\to\infty} P(\mathcal{P}_n^*(\alpha) > \alpha) = 0$ so that the PVOT has a degenerate limit distribution; or (ii) if $\mathcal{T}(\lambda) = \mathcal{T}(\lambda^*)$ $a.s.$ for some λ^* and all λ such that they are perfectly *dependent*, then $\lim_{n\to\infty} P(\mathcal{P}_n^*(\alpha) > \alpha) = \alpha$ and the asymptotic size is α. Neither case seems plausible in practice, although (ii) occurs when λ is a tuning parameter since these do not appear in the limit process (see Hill and Aguilar, 2013). Case (i) is logical since in this case $\mathcal{P}_n^*(\alpha) \stackrel{d}{\to} \int_\Lambda I(\bar{F}_0(\mathcal{T}(\lambda)) < \alpha)d\lambda$, while $\int_\Lambda I(\bar{F}_0(\mathcal{T}(\lambda)) < \alpha)d\lambda$ has mean α and is just a limiting Riemann sum of bounded independent random variables, hence it has a zero variance by dominated convergence. As long as $\mathcal{T}(\lambda)$ is weakly dependent on a subset of Λ with positive measure then $\lim_{n\to\infty} P(\mathcal{P}_n^*(\alpha) > \alpha) > 0$, ruling out (i). An example is $\mathcal{T}(\lambda) = \mathcal{Z}(\lambda)^2$ where $\{\mathcal{Z}(\lambda)\}$ is a Gaussian process with unit variance and covariance kernel $E[\mathcal{Z}(\lambda)\mathcal{Z}(\tilde{\lambda})]$ that is continuous in $(\lambda, \tilde{\lambda})$ and not everywhere equal to zero. All tests discussed in this paper have weakly dependent $\mathcal{T}(\lambda)$ under standard regularity conditions.

Next, asymptotic global power of PV test (2) translates to global power for PVOT test (4).

Theorem 2.2.
a. $\lim_{n\to\infty} P(\mathcal{P}_n^*(\alpha) > \alpha) > 0$ if and only if $p_n(\lambda) < \alpha$ *on a subset of* Λ *with Lebesgue measure* greater *than* α *asymptotically with positive probability.*
b. The PVOT test is consistent $P(\mathcal{P}_n^*(\alpha) > \alpha) \to 1$ *if the PV test is consistent* $P(p_n(\lambda) < \alpha) \to 1$ *on a subset of* Λ *with measure greater than* α*.*

Remark 5. As long as the PV test is consistent on a subset of Λ with measure greater than α, then the PVOT test is consistent. This trivially holds when the PV test is consistent for any λ outside a set with measure zero, including Andrews' (2001) GARCH test which is consistent on a known compact Λ; White (1989), Bierens (1990) and Bierens and Ploberger (1997) tests (and others) of omitted nonlinearity; Andrews' (1993) structural break test; and a test of an omitted Box-Cox transformation. See Section 4 for examples. At the risk of abusing terminology, we will say a test based on $\mathcal{T}_n(\lambda)$ is *randomized* when λ is uniformly randomized on Λ independent of the data. The randomized test is consistent only if the PV test is consistent for every λ outside a set with measure zero.[4] The transforms $\int_\Lambda \mathcal{T}_n(\lambda)\mu(d\lambda)$ and $\sup_{\lambda\in\Lambda} \mathcal{T}_n(\lambda)$, however, are consistent if the PV test is consistent on a subset of Λ with positive measure. Thus, the PVOT test ranks above the randomized test, but below average and supremum tests in terms of required PV test asymptotic power over Λ. As we discussed in Section 1, it is difficult to find a relevant example in which this matters, outside a toy example. We give such an example below.

[4]Here and elsewhere we refer to a test based on $\mathcal{T}_n(\lambda_*)$ as a *randomized test*, which is generally different from the classical definition of a randomized test (cf. Lehmann, 1994).

The following shows how PV test power transfers to the PVOT test.

Example 2.3. Let λ_* be a random draw from a uniform distribution on Λ. The parameter space is $\Lambda = [0,1]$, $\mathcal{T}_n(\lambda) \stackrel{p}{\to} \infty$ for $\lambda \in [.5,.56]$ such that the PV test is consistent on a subset with measure $\beta = .06$, and $\{\mathcal{T}_n(\lambda) : \lambda \in \Lambda/[.5,.56]\} \Rightarrow^* \{\mathcal{T}(\lambda) : \lambda \in \Lambda/[.5,.56]\}$ such that there is only trivial power. Thus, $\int_\Lambda \mathcal{T}_n(\lambda)\mu(d\lambda)$ and $\sup_{\lambda\in\Lambda} \mathcal{T}_n(\lambda)$ have asymptotic power of one. A uniformly randomized PV test is not consistent at any level, and at level $\alpha < .06$ has trivial power.

In the PVOT case, however, by applying arguments in the proof of Theorem 2.1, we can show $\lim_{n\to\infty} P(\mathcal{P}_n^*(\alpha) > \alpha)$ is identically

$$P\left(\int_{\lambda\in[.5,.56]} d\lambda + \int_{\lambda\notin[.5,.56]} I\left(\mathcal{U}(\lambda) < \alpha\right) d\lambda > \alpha\right) = P\left(\int_{\lambda\notin[.5,.56]} I\left(\mathcal{U}(\lambda) < \alpha\right) d\lambda > \alpha - .06\right)$$

for some process $\{\mathcal{U}(\lambda) : \lambda \in \Lambda/[.5,.56]\}$ where $\mathcal{U}(\lambda)$ is uniform on $[0,1]$. This implies the PVOT test is consistent at level $\alpha \leq .06$ since $\int_{\lambda\notin[.5,.56]} I(\mathcal{U}(\lambda) < \alpha)d\lambda > 0$ $a.s.$

3 Local Power

A characterization of local power requires an explicit hypothesis and some information on the construction of $\mathcal{T}_n(\lambda)$. Assume an observed sequence $\{y_t\}_{t=1}^n$ has a parametric joint distribution $f(y;\theta_0)$, where $\theta_0 = [\beta_0', \delta_0',]$ and $\beta_0 \in \mathbb{R}^r$, $r \geq 1$. Consider testing whether the subvector $\beta_0 = 0$, while δ_0 may contain other distribution parameters. If some additional parameter λ is part of δ_0 only when $\beta_0 \neq 0$, and therefore not identified under H_0, then we have Andrews and Ploberger's (1994) setting, but in general λ need not be part of the true data generating process.

We first treat a general environment. We then study a test of omitted nonlinearity, and perform a numerical experiment in order to compare local power.

3.1 Local Power : General Case

The sequence of local alternatives we consider is:

$$H_1^L : \beta_0 = \mathcal{N}_n^{-1} b \text{ for some } b \in \mathbb{R}^r, \tag{5}$$

where $(\mathcal{N}_n\}$ is a sequence of diagonal matrices $[\mathcal{N}_{n,i,j}]_{i,j=1}^r$, $\mathcal{N}_{n,i,i} \to \infty$. The test statistic is $\mathcal{T}_n(\lambda) \equiv h(\mathcal{Z}_n(\lambda))$ for a sequence of random functions $\{\mathcal{Z}_n(\lambda)\}$ on $\mathbb{R}^q$, $q \geq 1$, and a measurable function $h : \mathbb{R}^q \to [0,\infty)$ where $h(x)$ is monotonically increasing in $||x||$, and $h(x) \to \infty$ as $||x|| \to \infty$.

We assume regularity conditions apply such that under H_1^L

$$\{\mathcal{Z}_n(\lambda) : \lambda \in \Lambda\} \Rightarrow^* \{\mathcal{Z}(\lambda) + c(\lambda)b : \lambda \in \Lambda\}, \tag{6}$$

for some matrix $c(\lambda) \in \mathbb{R}^{q \times r}$, and $\{\mathcal{Z}(\lambda)\}$ is a zero mean process on $\mathbb{R}^q$ with a version that has *almost surely* uniformly continuous sample paths (with respect to some norm $||\cdot||$). In many cases in the literature $\{\mathcal{Z}(\lambda)\}$ is a Gaussian process with $E[\mathcal{Z}(\lambda)\mathcal{Z}(\lambda)'] = I_q$.

Combine (6) and the continuous mapping theorem to deduce under H_0 the limiting distribution function $F_0(x) \equiv P(h(\mathcal{Z}(\lambda)) \leq x)$ for $\mathcal{T}_n(\lambda)$, cf. Billingsley (1999, Theorem 2.7). An asymptotic p-value is $p_n(\lambda) = \bar{F}_0(\mathcal{T}_n(\lambda)) \equiv 1 - F_0(\mathcal{T}_n(\lambda))$, hence $\int_\Lambda I(p_n(\lambda) < \alpha)d\lambda \stackrel{d}{\rightarrow} \int_\Lambda I(\bar{F}_0(h(\mathcal{Z}(\lambda)) + c(\lambda)b)) < \alpha)$ under H_1^L. Similarly, any continuous mapping g over Λ satisfies $g(\mathcal{T}_n(\lambda)) \stackrel{d}{\rightarrow} g(h(\mathcal{Z}(\lambda) + c(\lambda)b))$, including $\int_\Lambda \mathcal{T}_n(\lambda)\mu(d\lambda)$ and $\sup_{\lambda \in \Lambda} \mathcal{T}_n(\lambda)$. Obviously if $c(\lambda)b = 0$ when $b \neq 0$ then local power is trivial at λ. Whether any of the above tests has non-trivial asymptotic local power depends on the measure of the subset of Λ on which $\inf_{\xi'\xi=1} ||\xi' c(\lambda)|| > 0$.

In order to make a fair comparison across tests, we assume each is asymptotically correctly sized for a nominal level α test. The next result follows from the preceding properties, hence a proof is omitted.

Theorem 3.1. *Let (5), (6) and $b \neq 0$ hold, and write $\inf_{\xi'\xi=1} ||\xi' c(\lambda)||$. Assume the randomized statistic $\mathcal{T}_n(\lambda^*)$ uses a draw λ^* from a uniform distribution on Λ. Asymptotic local power is non-trivial for (i) the PVOT test when $\inf_{\xi'\xi=1} ||\xi' c(\lambda)|| > 0$ on a subset of Λ with measure greater than α; and (ii) the uniformly randomized, average and supremum tests when $\inf_{\xi'\xi=1} ||\xi' c(\lambda)|| > 0$ on a subset of Λ with positive measure.*
b. Under cases (i) and (ii), asymptotic local power is monotonically increasing in $|b|$ and converges to one as $|b| \rightarrow \infty$.

Remark 6. The PVOT test ranks lower than randomized, average and supremum tests because it rejects only when the PV tests rejects on a subset of Λ with measure greater than α. Indeed, the PVOT test cannot asymptotically distinguish between PV tests that are consistent on a subset with measure less than α and have trivial power otherwise, or have trivial power everywhere. This cost is slight since a meaningful example in which it matters is difficult to find. The previously cited tests of omitted nonlinearity, one time structural break, and GARCH effects all have randomized, PVOT, average and supremum versions with non-trivial local power, although we only give complete details for a test of omitted nonlinearity below.

3.2 Example : Test of Omitted Nonlinearity

The proposed model to be tested is

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

where ζ_0 lies in the interior of $\mathfrak{Z}$, a compact subset of $\mathbb{R}^q$, $x_t \in \mathbb{R}^k$ contains a constant term and may contain lags of y_t, and $f : \mathbb{R}^k \times \mathfrak{Z} \to \mathbb{R}$ is a known response function. Assume $\{e_t, x_t, y_t\}$ are stationary for simplicity. Let Ψ be a 1-1 bounded mapping from $\mathbb{R}^k$ to $\mathbb{R}^k$, let $\mathcal{F} : \mathbb{R} \to \mathbb{R}$ be analytic and non-polynomial (e.g. exponential or logistic), and assume $\lambda \in \Lambda$, a compact subset of $\mathbb{R}^k$. Misspecification $\sup_{\zeta \in \mathbb{R}^q} P(E[y_t|x_t] = f(x_t, \zeta)) < 1$ implies $E[e_t \mathcal{F}(\lambda'\Psi(x_t))] \neq 0 \; \forall \lambda \in \Lambda/\mathcal{S}$, where $\mathcal{S}$ has Lebesgue measure zero. See White (1989), Bierens (1990) and Stinchcombe and White (1998) for seminal results for iid data, and see de Jong (1996) and Hill (2008) for dependent data. The test statistic for a test of the hypothesis $H_0 : E[y_t|x_t] = f(x_t, \zeta_0)$ $a.s.$ is

$$\mathcal{T}_n(\lambda) = \left(\frac{1}{\hat{v}_n(\lambda)} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} e_t(\hat{\zeta}_n) \mathcal{F}\left(\lambda'\Psi(x_t)\right) \right)^2 \text{ where } e_t(\zeta) \equiv y_t - f(x_t, \zeta). \tag{7}$$

The estimator $\hat{\zeta}_n$ is assumed $\sqrt{n}$-consistent for a strongly identified ζ_0, and $\hat{v}_n^2(\lambda)$ is a consistent estimator of $E[\{1/\sqrt{n} \sum_{t=1}^{n} e_t(\hat{\zeta}_n) \mathcal{F}(\lambda'\Psi(x_t))\}^2]$. By application of Theorem 3.3, below, the asymptotic p-value is $p_n(\lambda) \equiv 1 - F_0(\mathcal{T}_n(\lambda)) \equiv \bar{F}_0(\mathcal{T}_n(\lambda))$ where F_0 is the $\chi^2(1)$ distribution function.

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

$$H_1^L : \beta_0 = b/n^{1/2} \text{ for } b \in \mathbb{R}. \tag{8}$$

We assume for now that regularity conditions apply such that, for some sequence of positive finite non-random numbers $\{c(\lambda)\}$:

$$\text{under } H_1^L : \; \{\mathcal{T}_n(\lambda) : \lambda \in \Lambda\} \Rightarrow^* \{(\mathcal{Z}(\lambda) + bc(\lambda))^2 : \lambda \in \Lambda\}, \tag{9}$$

where $\{\mathcal{Z}(\lambda) + c(\lambda)b\}$ is a Gaussian process with mean $\{c(\lambda)b\}$, and *almost surely* uniformly continuous sample paths. See below for low level assumptions that imply (9). The latter implies by Theorem 2.1 that the PVOT asymptotic probability of rejection $\lim_{n\to\infty} P(\mathcal{P}_n^*(\alpha) > \alpha)$, under H_0, is between $(0, \alpha]$.

Let $F_{J,\nu}(c)$ denote a noncentral $\chi^2(J)$ law with noncentrality ν, hence $(\mathcal{Z}(\lambda) + c(\lambda)b)^2$ is distributed $F_{1,b^2c(\lambda)^2}$. Under the null $b = 0$ by construction $p_n(\lambda) \xrightarrow{d} \bar{F}_{1,0}((\mathcal{Z}(\lambda) + c(\lambda)b)^2) = \bar{F}_{1,0}(\mathcal{Z}(\lambda)^2)$ is uniformly distributed on $[0, 1]$. Under the global alternative $\sup_{\zeta \in \mathbb{R}^q} P(E[y_t|x_t] = f(x_t, \zeta)) < 1$ notice $\mathcal{T}_n(\lambda) \xrightarrow{p} \infty \; \forall \lambda \in \Lambda/S$ implies $p_n(\lambda) \xrightarrow{p} 0 \; \forall \lambda \in \Lambda/S$, hence $\mathcal{P}_n^*(\alpha) \xrightarrow{p} 1$ by Theorem 2.2. The latter implies the PVOT test of $E[y_t|x_t] = f(x_t, \zeta_0)$ *a.s.* is consistent. The following contains the result under H_1^L.

Theorem 3.2. *Under (9), asymptotic local power of the PVOT test is* $P(\int_\Lambda I(\bar{F}_{1,0}(\{\mathcal{Z}(\lambda) + c(\lambda)b\}^2) < \alpha)d\lambda > \alpha)$. *Hence, under H_1^L the probability the PVOT test rejects H_0 increases to unity monotonically as the drift parameter $|b| \to \infty$, for any nominal level $\alpha \in [0, 1)$.*

The following assumptions detail sufficient conditions leading to (9). These are not the most general possible, but are fairly compact for the sake of brevity.

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

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

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

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

d. Variance Estimator*: $\hat{v}_n^2(\lambda) \equiv 1/n \sum_{s,t=1}^n \mathcal{K}((s - t)/\gamma_n) e_s(\hat{\zeta}_n) e_t(\hat{\zeta}_n) \hat{w}_{n,s}(\lambda, \hat{\zeta}_n) \hat{w}_{n,t}(\lambda, \hat{\zeta}_n)$ with kernel $\mathcal{K}$ and bandwidth $\gamma_n \to \infty$ and $\gamma_n = o(\sqrt{n})$. $\mathcal{K}$ is continuous at 0 and all but a finite number of points, $\mathcal{K} : \mathbb{R} \to [-1, 1]$, $\mathcal{K}(0) = 1$, $\mathcal{K}(x) = \mathcal{K}(-x)$ $\forall x \in \mathbb{R}$, $\int_{-\infty}^{\infty} |\mathcal{K}(x)|dx < \infty$; and there exists $\{\delta_n\}$, $\delta_n > 0$, $\delta_n/\sqrt{n} \to \infty$, such that $\int_{\delta_n}^{\infty} \{|\mathcal{K}(x)| + |\mathcal{K}(-x)|\}dx = o(1/\sqrt{n})$.*

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

Remark 7. The kernel variance $\hat{v}_n^2(\lambda)$ form follows from a standard expansion of $1/\sqrt{n} \sum_{t=1}^n e_t(\hat{\zeta}_n)\mathcal{F}(\lambda'\Psi(x_t))$ around ζ_0 under H_0. We exploit a kernel estimator in order to prove uniform convergence of $\hat{v}_n^2(\lambda)$ without the assumption that H_0 is true, a generality that may be of separate interest. See Lemma C.1 in Hill (2020a, Appendix C).

Remark 8. Property (d), other than the requirement that $\mathcal{I}_n \equiv \int_{\delta_n}^{\infty} \{|\mathcal{K}(x)| + |\mathcal{K}(-x)|\}dx = o(1/\sqrt{n})$ for $\delta_n/\sqrt{n} \to \infty$, is similar to properties in Andrews (1991) and elsewhere, covering Bartlett, Parzen, Tukey-Hanning and Quadratic-Spectral kernels. We use $\mathcal{I}_n = o(1/\sqrt{n})$ with $\delta_n/\sqrt{n} \to \infty$ to prove uniform convergence $\sup_{\lambda\in\Lambda} |\hat{v}_n^2(\lambda) - v^2(\lambda)| \xrightarrow{p} 0$. The bound $\mathcal{I}_n = o(1/\sqrt{n})$ is trivially satisfied for any $\delta_n \geq K$ and some finite $K > 0$ for Bartlett, Parzen, and Tukey-Hanning kernels, while the

Quadratic-Spectral kernel obtains $\mathcal{I}_n \leq K \int_{\delta_n}^{\infty} x^{-2}dx = K\delta_n^{-3}$ hence $\mathcal{I}_n = o(1/\sqrt{n})$ for any $\delta_n/n^{1/6} \rightarrow \infty$.

The next claim is proven in Hill (2020a, Appendix C) since it follows from standard arguments.

Theorem 3.3.

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

$$E\left[\tilde{\mathcal{Z}}_n(\lambda)\tilde{\mathcal{Z}}_n(\tilde{\lambda})\right] = \frac{E\left[\epsilon_t^2 w_t(\lambda) w_t(\tilde{\lambda})\right]}{\left(E[\epsilon_t^2 w_t^2(\lambda)]E[\epsilon_t^2 w_t^2(\tilde{\lambda})]\right)^{1/2}}. \tag{10}$$

b. Under H_1^L weak convergence (9) is valid with $c(\lambda) = E[w_t^2(\lambda)]/(E[\epsilon_t^2 w_t^2(\lambda)])^{1/2} > 0$ where $w_t(\lambda) \equiv F_t(\lambda) - E[F_t(\lambda)g_t(\zeta_0)'] \times (E[g_t(\zeta_0)g_t(\zeta_0)'])^{-1}g_t(\zeta_0)$.

Theorem 3.3.a implies under H_0 the test statistic converges weakly $\{\mathcal{T}_n(\lambda) : \lambda \in \Lambda\} \Rightarrow^* \{\mathcal{Z}(\lambda)^2 : \lambda \in \Lambda\}$, where $\{\mathcal{Z}(\lambda)\}$ is weakly dependent in the sense of Theorem 2.1: $P(\bar{F}_0(\mathcal{T}(\lambda)) < \alpha, \bar{F}_0(\mathcal{T}(\tilde{\lambda})) < \alpha) > \alpha^2$ on a subset of $\Lambda \times \Lambda$ with positive measure. This follows instantly from Gaussianicity of $\{\mathcal{Z}(\lambda)\}$ and its continuous covariance kernel (10). This in turn implies by Theorem 2.1 that the PVOT $\mathcal{P}_n^*(\alpha) \equiv \int_\Lambda I(p_n(\lambda) < \alpha)d\lambda$ does not have a degenerate limit distribution, which yields the following result by invoking Theorems 2.1 and 3.3.a.

Theorem 3.4. *Let Assumption 2 and H_0 hold. Then* $\lim_{n\rightarrow\infty} P(\mathcal{P}_n^*(\alpha) > \alpha) \in (0, \alpha]$.

3.3 Numerical Experiment : Test of Omitted Nonlinearity

Our final goal in this section is to compare asymptotic local power for tests based on the PVOT, average $\int_\Lambda \mathcal{T}_n(\lambda)\mu(d\lambda)$ with uniform measure $\mu(\lambda)$, supremum $\sup_{\lambda\in\Lambda} \mathcal{T}_n(\lambda)$, and Bierens and Ploberger's (1997) Integrated Conditional Moment [ICM] statistics. We work with a simple model $y_t = \zeta_0 x_t + \beta_0 \exp\{\lambda x_t\} + \epsilon_t$, where $\zeta_0 = 1$, $\beta_0 = b/\sqrt{n}$, and $\{\epsilon_t, x_t\}$ are iid $N(0, 1)$ distributed. We omit a constant term entirely for simplicity. In order to abstract from the impact of sampling error on asymptotics, we assume $\zeta_0 = 1$ is known, hence the test statistic is

$$\mathcal{T}_n(\lambda) \equiv \frac{\hat{z}_n^2(\lambda)}{\hat{v}_n^2(\lambda)} \text{ where } \hat{z}_n(\lambda) \equiv \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(y_t - \zeta_0 x_t)\exp\{\lambda x_t\},\ \hat{v}_n^2(\lambda) \equiv \frac{1}{n}\sum_{t=1}^{n}(y_t - \zeta_0 x_t)^2 \exp\{2\lambda x_t\}.$$

The nuisance parameter space is $\Lambda = [0, 1]$. A Gaussian setting implies the main results of Andrews and Ploberger (1994) apply: the average $\int_\Lambda \mathcal{T}_n(\lambda)\mu(d\lambda)$ has the highest weighted average local power for alternatives close to the null.

In view of Gaussianicity, and Theorem 3.3, it can be shown $\{\mathcal{T}_n(\lambda)\} \Rightarrow^* \{(\mathcal{Z}(\lambda) + c(\lambda)b)^2\}$, where $c(\lambda) = E[\exp\{2\lambda x_t\}]/(E[\epsilon_t^2 \exp\{2\lambda x_t\}])^{1/2} = (E[\exp\{2\lambda x_t\}])^{1/2} = \exp\{\lambda^2\}$, and $\{\mathcal{Z}(\lambda)\}$ is a zero mean Gaussian process with *almost surely* uniformly continuous sample paths, and covariance function $E[\mathcal{Z}(\lambda)\mathcal{Z}(\tilde{\lambda})] = \exp\{-.5(\lambda - \tilde{\lambda})^2\}$. Local asymptotic power is therefore:

$$\begin{aligned}
&\text{PVOT: } P\left(\int_0^1 I\left(\bar{F}_{1,0}\left(\left\{\mathcal{Z}(\lambda) + b\exp\{\lambda^2\}\right\}^2\right) < \alpha\right) d\lambda > c_\alpha^{(pvot)}\right) \\
&\text{randomized: } P\left(\left\{\mathcal{Z}(\lambda_*) + b\exp\{\lambda_*^2\}\right\}^2 > c_\alpha^{(rand)}\right) \\
&\text{average: } P\left(\int_0^1 \left\{\mathcal{Z}(\lambda) + b\exp\{\lambda^2\}\right\}^2 d\lambda > c_\alpha^{(ave)}\right) \\
&\text{supremum: } P\left(\sup_{\lambda\in[0,1]} \left\{\mathcal{Z}(\lambda) + b\exp\{\lambda^2\}\right\}^2 > c_\alpha^{(\sup)}\right),
\end{aligned}$$

where $\bar{F}_{1,0}$ is the upper tail probability of a $\chi^2(1)$ distribution; λ_* is a uniform random variable on Λ, independent of $\{\epsilon_t, x_t\}$; and $c_\alpha^{(\cdot)}$ are level α asymptotic critical values under the null: $c_\alpha^{(pvot)} \equiv \alpha$, and $c_\alpha^{rand)}$ is the $1-\alpha$ quantile from a $\chi^2(1)$ distribution. See below for approximating $\{c_\alpha^{(ave)}, c_\alpha^{(\sup)}\}$.

Local power for Bierens and Ploberger's (1997) ICM statistic $\widehat{\mathcal{I}}_n \equiv \int_0^1 \hat{z}_n^2(\lambda)\mu(d\lambda)$ is based on their Theorem 7 critical value upper bound $\lim_{n\to\infty} P(\widehat{\mathcal{I}}_n \geq u_\alpha \int_0^1 v_n^2(\lambda)\mu(d\lambda)) \leq \alpha$, where $v_n^2(\lambda)$ $= \exp\{2\lambda^2\}$ satisfies $\sup_{\lambda\in[0,1]} |\hat{v}_n^2(\lambda) - v_n^2(\lambda)| \xrightarrow{p} 0$, and $\{u_{.01}, u_{.05}, u_{.10}\} = \{6.81,\ 4.26,\ 3.23\}$. We use a uniform measure $\mu(\lambda) = \lambda$ since this promotes the highest weighted average local power for alternatives near H_0 (Andrews and Ploberger, 1994; Boning and Sowell, 1999). Under H_1^L we have $\{\hat{z}_n(\lambda)\} \Rightarrow^* \{z(\lambda) + b\exp\{\lambda^2\}\}$ for some zero mean Gaussian process $\{z(\lambda)\}$ with *almost surely* uniformly continuous sample paths, and $\int_0^1 v_n^2(\lambda)d\lambda = \int_0^1 \exp\{2\lambda^2\}d\lambda = 2.3645$. This yields local asymptotic power:

$$\text{ICM: } P\left(\int_0^1 \left\{z(\lambda) + b\exp\{\lambda^2\})\right\}^2 d\lambda > c_\alpha^{(icm)}\right) \text{ where } c_\alpha^{(icm)} \equiv 2.3645 \times u_\alpha.$$

Asymptotically valid critical values can be easily computed for the present experiment by mimicking the steps below, in which case PVOT, average, supremum, and ICM tests are essentially identical. We are, however, interested in how well Bierens and Ploberger's (1997) solution to the problem of non-standard inference compares to existing methods.

Local power is computed as follows. We draw R samples $\{\epsilon_{i,t}, x_{i,t}\}_{t=1}^T$, $i = 1, ..., R$, of iid random variables $(\epsilon_{i,t}, x_{i,t})$ from $N(0,1)$, and draw iid $\lambda_{*,i}$, $i = 1, ..., R$, from a uniform distribution on Λ. Then $\{\mathcal{Z}_{T,i}(\lambda)\} \equiv \{1/\sqrt{T}\sum_{t=1}^T \epsilon_{i,t}\exp\{\lambda x_{i,t} - \lambda^2\}\}$ becomes a draw from the limit process $\{\mathcal{Z}(\lambda)\}$ as $T \to \infty$. We draw $R = 100,000$ samples of size $T = 100,000$, and compute $\mathcal{T}_{T,i}^{(PVOT)}(b) \equiv$

$\int_0^1 I(\bar{F}_{1,0}(\{\mathcal{Z}_{T,i}(\lambda) + b\exp\{\lambda^2\}\}^2) < \alpha)d\lambda$, $\mathcal{T}_{T,i}^{(ave)}(b) \equiv \int_0^1 \{\mathcal{Z}_{T,i} + b\exp\{\lambda^2\}\}^2 d\lambda$ and $\mathcal{T}_{T,i}^{(\sup)}(b) \equiv \sup_{\lambda\in[0,1]}\{\mathcal{Z}_{T,i}(\lambda) + b\exp\{\lambda^2\}\}^2$ and $\mathcal{T}_{T,i}^{(rand)}(b) \equiv \{\mathcal{Z}_{T,i}(\lambda_{*,i}) + b\exp\{\lambda_{*,i}^2\}\}^2$. The critical values $\{c_\alpha^{(ave)}, c_\alpha^{(\sup)}\}$ are the $1 - a$ quantiles of $\{\mathcal{T}_{T,i}^{(ave)}(0), \mathcal{T}_{T,i}^{(\sup)}(0)\}_{i=1}^R$. In the ICM case $\{z_{T,i}(\lambda)\} \equiv \{1/\sqrt{T}\sum_{t=1}^T \epsilon_{i,t}\exp\{\lambda x_{i,t}\}\}$ becomes a draw from $\{z(\lambda)\}$ as $T \to \infty$, hence we compute $\mathcal{T}_{T,i}^{(icm)}(b) \equiv \int_0^1 \{z_{T,i} + b\exp\{\lambda^2\}\}^2 d\lambda$. Local power is $1/R\sum_{i=1}^R I(\mathcal{T}_{T,i}^{(\cdot)}(b) > c_\alpha^{(\cdot)})$. Integrals are computed by the midpoint method based on the discretization $\lambda \in \{.001, .002, ..., .999, 1\}$, hence there are 1000 points ($\lambda = 0$ is excluded because power is trivial in that case).

Figure D.1 in Hill (2020a) contains local power plots at level $\alpha = .05$ over drift parameters $b \in [0, 2]$ and $b \in [0, 7]$. Notice that under the null $b = 0$ each test, except ICM, achieves power of nearly exactly .05 (PVOT, average and supremum are .0499, and randomized is .0511), providing numerical verification that the correct critical value for the PVOT test at level α is simply α. The ICM critical value upper bound leads to an under sized test with asymptotic size .0365.

Second, local power is virtually identical across PVOT, random, average and supremum tests. This is logical since the underlying PV test is consistent on any compact Λ outside of a measure zero subset, it has non-trivial local power, and local power is asymptotic. Since the average test has the highest weighted average power aimed at alternatives near the null (Andrews and Ploberger, 1994, eq. (2.5)), we have evidence that PVOT test power is at the highest possible level. The randomized test has slightly lower power for deviations far from the null $b \geq 2.5$ ostensibly because for large b larger values of λ lead to a higher power test, while the randomized λ may be small. Finally, ICM power is lower near the null $b \in (0, 1.5]$ since these alternatives are most difficult to detect, and the test is conservative, but power is essentially identical to the remaining tests for drift $b \geq 1.5$.

4 Examples

In addition to the test of omitted nonlinearity above, we provide two more examples of PVOT tests: a test of functional form when some parameters may be weakly identified, and a test of GARCH effects.

4.1 PVOT Test of Functional Form with Possible Weak Identification

This example showcases a unique advantage of the PVOT test: it allows for robust bootstrap inference in the presence of weak identification *and* a consistent test. Conversely, test statistic functionals like the supremum $\sup_{\lambda\in\Lambda}\mathcal{T}_n(\lambda)$ and average $\int_\Lambda \mathcal{T}_n(\lambda)\mu(d\lambda)$ cannot be validly bootstrapped asymptotically when weak identification is possible (see Hill, 2020b), and $\sup_{\lambda\in\Lambda} p_n(\lambda)$ need not be consistent. The following is based on ideas developed in Hill (2020b).

We work with the following model:

$$y_t = \zeta' x_t + \beta' g(x_t, \pi) + \epsilon_t = f(\theta, x_t) + \epsilon_t \text{ where } x_t \in \mathbb{R}^{k_x} \text{ and } \theta = \left[\zeta', \beta', \pi'\right]' \in \Theta, \tag{11}$$

where g is a known function, and $E[\epsilon_t] = 0$ and $E[\epsilon_t^2] \in (0, \infty)$ for unique $\theta_0 \in \Theta$ and compact Θ.

We want to test $H_0 : E[y_t|x_t] = f(\theta_0, x_t)$ $a.s.$ against $H_1 : \sup_{\theta \in \Theta} P(E[y_t|x_t] = f(\theta_0, x_t)) < 1$. If $\beta_0 \neq 0$ then π_0 is not identified. If there is local drift $\beta_0 = \beta_n \to 0$ with $\sqrt{n}||\beta_n|| \to [0, \infty)$, then estimators of π_0 have random probability limits, and estimators for θ_0 have nonstandard limit distributions. The literature on consistent specification testing generally assumes strong identification (e.g. Bierens, 1982; White, 1989; Bierens, 1990; Hong and White, 1995; de Jong, 1996; Bierens and Ploberger, 1997; Hill, 2008), while the weak identification literature presumes model correctness $E[y_t|x_t] = f(\theta_0, x_t)$ $a.s.$ (e.g. Andrews and Cheng, 2012, 2013, 2014). See Hill (2020b) for more references. Hill (2020b) allows for both weak identification *and* model mis-specification.

Hill (2020b) proposes a modified Conditional Moment [CM] test statistic and bootstrap procedure, both to account for possible weak identification. We only highlight definitions and the main result here. Define

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

$$\widehat{\mathcal{H}}_n = \frac{1}{n}\sum_{t=1}^{n} d_{\theta,t}(\omega(\hat{\beta}_n), \hat{\pi}_n) d_{\theta,t}(\omega(\hat{\beta}_n), \hat{\pi}_n)' \text{ where } \omega(\beta) \equiv \begin{cases} \beta / \|\beta\| & \text{if } \beta \neq 0 \\ 1_{k_\beta} / \left\|1_{k_\beta}\right\| & \text{if } \beta = 0 \end{cases}$$

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

The CM statistic is:

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

The test statistic is similar to those in Bierens (1990) and Stinchcombe and White (1998). The scale $\hat{\upsilon}_n(\hat{\theta}_n, \lambda)$, however, has been altered by dividing by $||\beta||$ in order to avoid a singular Hessian matrix under semi-strong identification $\beta_0 = 0$ and $\sqrt{n}||\beta_n|| \to \infty$ (cf. Andrews and Cheng, 2012, Section 3.5).

Technical results are derived under two overlapping identification cases: under case $\mathcal{C}(i, b)$ there is $\beta_n \to \beta_0 = 0$ and $\sqrt{n}\beta_n \to b$ where $b \in (\mathbb{R} \cup \{\pm\infty\})^{k_\beta}$; and under case $\mathcal{C}(ii, \omega_0)$, $\beta_n \to \beta_0$ 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 non-identified. 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. Cf. Andrews and Cheng (2012). In order to conserve space, we say "weak identification" to mean weak and non-identification, and "strong identification" to mean semi-strong or strong identification

Under strong identification $\mathcal{C}(ii, \omega_0)$, $\{\mathcal{T}_n(\lambda) : \lambda \in \Lambda\}$ converges weakly to a chi-squared process. Under weak identification $\mathcal{C}(i, b)$ the limit process is non-standard with nuisance parameter λ, and other nuisance parameters h containing b and distribution nuisance parameters (e.g. π_0 and $E[\epsilon_t^2]$). See Hill (2020b, Theorem 4.2 and Section 5). Let $\{\mathcal{T}(\lambda) : \lambda \in \Lambda\}$ denote either limit process.

Test statistic transforms like $\sup_{\lambda \in \Lambda} \mathcal{T}_n(\lambda)$ and $\int_\Lambda \mathcal{T}_n(\lambda)\mu(d\lambda)$ cannot be consistently bootstrapped or simulated if weak identification is possible. The reason is a consistent estimate of the covariance kernel for $\{\mathcal{T}(\lambda) : \lambda \in \Lambda\}$ is required, which depends on π_0. The latter cannot be consistently estimated under the weak identification case $\mathcal{C}(i, b)$ (Andrews and Cheng, 2012). Invalidity of the bootstrap is easily demonstrated by simulation: see Hill (2020b), and see Section 5.2 below.

Hill (2020b) therefore takes a different approach by bootstrapping a p-value $p_n(\lambda)$ for $\mathcal{T}_n(\lambda)$ that is consistent for the asymptotic p-value, under any degree of (non)identification. The key steps involve computing (or bootstrapping) the asymptotic p-value under strong identification, wild bootstrapping the p-value under weak identification using $\mathcal{M}$ bootstrapped samples, and then combining the two in a way that promotes valid inference asymptotically under any degree of identification.[5] Let $\hat{p}_{n,\mathcal{M}}(\lambda)$ be the resulting combined bootstrapped p-value (see Hill, 2020b, Sections 5 and 6).

Define the PVOT $\hat{\mathcal{P}}_{n,\mathcal{M}}(\alpha) \equiv \int_\Lambda I(\hat{p}_{n,\mathcal{M}}(\lambda) < \alpha)d\lambda$. The test rejects H_0 when $\hat{\mathcal{P}}_{n,\mathcal{M}}(\alpha) > \alpha$. The PVOT test has the correct asymptotic level and is consistent. See Hill (2020b, Theorem 6.3) for a proof of the following result.

Theorem 4.1. *Let $\mathcal{M} = \mathcal{M}_n \to \infty$ as $n \to \infty$. Under regularity conditions presented in Hill (2020b, Theorem 6.3), if H_0 is true then $\lim_{n\to\infty} P(\hat{\mathcal{P}}_{n,\mathcal{M}}(\alpha) > \alpha) \leq \alpha$, and otherwise $P(\hat{\mathcal{P}}_{n,\mathcal{M}}(\alpha) > \alpha) \to 1$.*

Remark 9. As stated above, there does not exist a valid bootstrap method for handling *test statistic* functionals like the average and supremum. The bootstrap method developed in Hill (2020b) is only valid for computing an approximate p-value for the non-smoothed $\mathcal{T}_n(\lambda)$ that is asymptotically consistent for the asymptotic p-value (Hill, 2020b, Theorem 6.2). The practitioner is therefore left with smoothing such a p-value approximation $\hat{p}_{n,\mathcal{M}}(\lambda)$. The supremum $\sup_{\lambda\in\Lambda} \hat{p}_{n,\mathcal{M}}(\lambda)$, however, promotes a conservative test that is not consistent. Even though $\hat{p}_{n,\mathcal{M}}(\lambda) \xrightarrow{p} 0 \; \forall \lambda \in \Lambda/S$ where S has Lebesgue measure zero, as long as there exists $\lambda \in \Lambda$ such that a Type II error occurs, i.e. $\hat{p}_{n,\mathcal{M}}(\lambda) \xrightarrow{p}$

[5] Hill (2020b) uses the *least favorable* and *identification category selection* constructions from Andrews and Cheng (2012) as the basis for p-value combinations. Andrews and Cheng (2012) use those notions for critical value combinations under assumed model correctness and without a nuisance parameter under a specific hypothesis.

$(0,1]$, then $\sup_{\lambda\in\Lambda}\hat{p}_{n,\mathcal{M}}(\lambda)\xrightarrow{p}(0,1]$ and the sup-p-value test is inconsistent. Conversely, the PVOT test with $\hat{p}_{n,\mathcal{M}}(\lambda)$ is both consistent and immune to weak identification.

4.2 PVOT Test of GARCH Effects

We want to test the hypothesis that a process does not have GARCH effects. Consider a stationary GARCH(1,1) model (Bollerslev, 1986; Nelson, 1990):

$$\begin{aligned} &y_t=\sigma_t\epsilon_t \text{ where } \epsilon_t \text{ is iid, } E[\epsilon_t]=0,\ E[\epsilon_t^2]=1, \text{ and } E\left|\epsilon_t\right|^r<\infty \text{ for } r>4 \\ &\sigma_t^2=\omega_0+\delta_0 y_{t-1}^2+\lambda_0\sigma_{t-1}^2 \text{ where } \omega_0>0,\ \delta_0,\lambda_0\in[0,1), \text{ and } E\left[\ln\left(\delta_0\epsilon_t^2+\lambda_0\right)\right]<0. \end{aligned} \tag{12}$$

Under H_0: $\delta_0=0$ if the starting value is $\sigma_0^2=\tilde{\omega}=\omega_0/(1-\lambda_0)>0$ then $\sigma_1^2=\omega_0+\lambda_0\omega_0/(1-\lambda_0)=\tilde{\omega}$ and so on under H_0, hence $\sigma_t^2=\tilde{\omega}\ \forall t\geq 0$ which means there are no GARCH effects. In this case the σ_{t-1}^2 marginal effect λ_0 is not identified. Further, $\delta_0,\lambda_0\geq 0$ must be maintained during estimation to ensure a positive conditional variance, and because this includes a boundary value, QML asymptotics are non-standard (Andrews, 1999, 2001).

Let $\theta=[\omega,\delta,\lambda]$, and define the parameter subset $\pi=[\omega,\delta]'\in\Pi\equiv[\iota_\omega,u_\omega]\times[0,1-\iota_\delta]$ for tiny $(\iota_\omega,\iota_\delta)>0$ and some $u_\omega>0$. Express the volatility process as $\sigma_t^2(\pi,\lambda)=\omega+\delta y_{t-1}^2+\lambda\sigma_{t-1}^2(\pi,\lambda)$ for an imputed $\lambda\in\Lambda\equiv[0,1-\iota_\lambda]$ and tiny $\iota_\lambda>0$. Denote the unrestricted QML estimator of π_0 for a given $\lambda\in\Lambda$: $\hat{\pi}_n(\lambda)=[\hat{\omega}_n(\lambda),\hat{\delta}_n(\lambda)]'\equiv\arg\min_{\pi\in\Pi}1/n\sum_{t=1}^n\{\ln(\sigma_t^2(\pi,\lambda))+y_t^2/\sigma_t^2(\pi,\lambda)\}$. Andrews' (1999) test statistic is:

$$\mathcal{T}_n(\lambda)=n\hat{\delta}_n^2(\lambda). \tag{13}$$

Theorem 4.2. *Let $\{y_t\}$ be generated by process (12). Assumption 1 applies where $\mathcal{T}(\lambda)=(\max\{0,\mathcal{Z}(\lambda)\})^2$, and $\{\mathcal{Z}(\lambda)\}$ is a zero mean Gaussian process with a version that has* almost surely *uniformly continuous sample paths, and covariance function $E[\mathcal{Z}(\lambda_1)\mathcal{Z}(\lambda_2)]=(1-\lambda_1^2)(1-\lambda_2^2)/(1-\lambda_1\lambda_2)$.*

A simulation procedure can be used to approximate the asymptotic p-value (cf. Andrews, 2001). Draw $\widetilde{\mathcal{M}}\in\mathbb{N}$ samples of iid standard normal random variables $\{Z_{j,i}\}_{j=1}^{\widetilde{\mathcal{R}}}$, $i=1,...,\widetilde{\mathcal{M}}$, and compute $\mathfrak{Z}_{\widetilde{\mathcal{R}},i}(\lambda)\equiv(1-\lambda^2)\sum_{j=0}^{\widetilde{\mathcal{R}}}\lambda^j Z_{j,i}$ and $\mathcal{T}_{\widetilde{\mathcal{R}},i}(\lambda)\equiv(\max\{0,\mathfrak{Z}_{\widetilde{\mathcal{R}},i}(\lambda)\})^2$. Notice $\mathfrak{Z}_{\widetilde{\mathcal{R}}}(\lambda)\equiv(1-\lambda^2)\sum_{j=0}^{\widetilde{\mathcal{R}}}\lambda^j Z_j$ is zero mean Gaussian with the same covariance function as $\mathcal{Z}(\lambda)$ when $\widetilde{\mathcal{R}}=\infty$, hence $\{\mathcal{T}_{\infty,i}(\lambda):\lambda\in\Lambda\}$ is an independent draw from the limit process $\{\mathcal{T}(\lambda):\lambda\in\Lambda\}$. The p-value approximation is $\hat{p}_{\widetilde{\mathcal{R}},\widetilde{\mathcal{M}},n}(\lambda)\equiv 1/\widetilde{\mathcal{M}}\sum_{i=1}^{\widetilde{\mathcal{M}}}I(\mathcal{T}_{\widetilde{\mathcal{R}},i}(\lambda)>\mathcal{T}_n(\lambda))$. Since we can choose $\widetilde{\mathcal{M}}$ and $\widetilde{\mathcal{R}}$ to be arbitrarily large, we can make $\hat{p}_{\widetilde{\mathcal{R}},\widetilde{\mathcal{M}},n}(\lambda)$ close to the asymptotic p-value by the Glivenko-Cantelli theorem. Now compute the PVOT $\mathcal{P}^*_{\widetilde{\mathcal{R}},\widetilde{\mathcal{M}},n}(\alpha)\equiv\int_\Lambda I(\hat{p}_{\widetilde{\mathcal{R}},\widetilde{\mathcal{M}},n}(\lambda)<\alpha)d\lambda$.

Theorem 4.3. *Let $\{y_t\}$ be generated by the process in (12), and let $\{\widetilde{\mathcal{R}}_n, \widetilde{\mathcal{M}}_n\}_{n\geq 1}$ be sequences of positive integers, $\widetilde{\mathcal{R}}_n \to \infty$ and $\widetilde{\mathcal{M}}_n \to \infty$. If H_0: $\delta_0 = 0$ is true then $\lim_{n\to\infty} P(\mathcal{P}^*_{\widetilde{\mathcal{R}}_n,\widetilde{\mathcal{M}}_n,n}(\alpha) > \alpha)$ $\in (0, \alpha]$. Otherwise if $\delta_0 > 0$ then $P(\mathcal{P}^*_{\widetilde{\mathcal{R}}_n,\widetilde{\mathcal{M}}_n,n}(\alpha) > \alpha) \to 1$.*

Remark 10. Under H_0, $h(\mathcal{T}_n(\lambda)) \overset{d}{\to} h(\mathcal{T}(\lambda))$ for mappings $h : \mathbb{R} \to \mathbb{R}$, continuous *a.e.*, by exploiting theory in Andrews (2001, Section 4). The relevant simulated p-value is $\hat{p}^{(h)}_{\widetilde{\mathcal{R}},\widetilde{\mathcal{M}},n} \equiv 1/\widetilde{\mathcal{M}} \sum_{i=1}^{\widetilde{\mathcal{M}}} I(h(\mathcal{T}_{\widetilde{\mathcal{R}},i}(\lambda)) > h(\mathcal{T}_n(\lambda)))$. Arguments used to prove Theorem 4.3 easily lead to a proof that $\hat{p}^{(h)}_{\widetilde{\mathcal{R}},\widetilde{\mathcal{M}},n}$ is consistent for the corresponding asymptotic p-value.

5 Simulation Study

We perform three Monte Carlo experiments concerning tests of functional form with and without the possibility of weak identification, and GARCH effects. The same discretized Λ is used for PVOT and bootstrap p-value tests, and integrals are discretized using the midpoint method. Wild bootstrapped p-values are computed with $R = 1000$ samples of iid standard normal random variables $\{z_{t,i}\}_{t=1}^n$. Sample sizes are $n \in \{100, 250, 500\}$ and $10,000$ samples $\{y_t\}_{t=1}^n$ are independently drawn in each case. Nominal levels are $\alpha \in \{.01, .05., .10\}$.

5.1 Test of Functional Form

We work with a threshold process in which all parameters are strongly identified.

Step-Up Samples $\{y_t\}_{t=1}^n$ are drawn from one of four data generating processes: linear $y_t = 2x_t + \epsilon_t$ or quadratic $y_t = 2x_t + .1x_t^2 + \epsilon_t$, where $\{x_t, \epsilon_t\}$ are iid standard normal random variables; and AR(1) $y_t = .9x_t + \epsilon_t$ or Self-Exciting Threshold AR(1) $y_t = .9x_t - .4x_t I(x_t > 0) + \epsilon_t$, where $x_t = y_{t-1}$ and ϵ_t is iid standard normal. In the time series cases we draw $2n$ observations with starting values $y_1 = \epsilon_1$ and retain the last n observations. Now write $\sum$ for sample summations: for iid data $\sum = \sum_{t=1}^n$ and for time series $\sum = \sum_{t=2}^n$. The estimated model is $y_t = \beta x_t + \epsilon_t$, and we test $H_0 : E[y_t|x_t] = \beta_0 x_t$ *a.s.* for some β_0.

We compute $\mathcal{T}_n(\lambda)$ in (7) with logistic $F(\Psi(x_t)) = (1 + \exp\{\Psi(x_t)\})^{-1}$ and $\Psi(x_t) = \arctan(x_t^*)$, where $x_t^* \equiv x_t - 1/n \sum x_t$. Write $F_t(\lambda) = F(\lambda\Psi(x_t))$, let $\hat{\beta}_n$ be the least squares estimator, and define $\hat{z}_n(\lambda) \equiv 1/n^{1/2} \sum(y_t - \hat{\beta}_n x_t)F_t(\lambda)$. Then $\mathcal{T}_n(\lambda) \equiv \hat{z}_n^2(\lambda)/\hat{v}_n^2(\lambda)$ with $\hat{v}_n^2(\lambda) \equiv 1/n \sum(y_t - \hat{\beta}_n x_t)^2 \hat{w}_{n,t}^2(\lambda)$, where $\hat{w}_{n,t}(\lambda) \equiv F_t(\lambda) - \hat{b}_n(\lambda)'\hat{A}_n^{-1}x_t$, $\hat{b}_n \equiv 1/n \sum x_t F_t(\lambda)$ and $\hat{A}_n \equiv 1/n \sum x_t x_t'$ (see White, 1989, cf. Bierens, 1990). It is straightforward to show Assumption 2.a,b,c,e holds, and $\sup_{\lambda\in\Lambda} |\hat{v}_n^2(\lambda) - v^2(\lambda)| \overset{p}{\to} 0$ by arguments used to prove Lemma C.1 in Hill (2020a). By Theorem 3.3, weak convergence (9) therefore applies, and $\mathcal{T}_n(\lambda)$ is pointwise asymptotically $\chi^2(1)$ under H_0.

Tests We perform four tests. First, the PVOT over $\Lambda = [.0001, 1]$ based on the asymptotic p-value for $\mathcal{T}_n(\lambda)$. The discretized set is $\Lambda_n \equiv \{.0001 + 1/(\varpi n), .0001 + 2/(\varpi n), ..., .0001 + \bar{\imath}_n(\varpi)/(\varpi n)\}$ where $\bar{\imath}_n(\varpi) \equiv \text{argmax}\{1 \leq i \leq \varpi n : i \leq .9999\varpi n\}$, with a coarseness parameter $\varpi = 100$. We can use a much smaller ϖ if the sample size is large enough (e.g. $\varpi = 10$ when $n = 250$, or $\varpi = 1$ when $n \geq 500$), but in general small ϖn leads to over-rejection of H_0. Second, we use $\mathcal{T}_n(\lambda_*)$ with a uniformly randomized $\lambda_* \in \Lambda$ and an asymptotic p-value. Third, $\sup_{\lambda \in \Lambda_n} \mathcal{T}_n(\lambda)$ and $\int_{\Lambda_n} \mathcal{T}_n(\lambda)\mu(d\lambda)$ with uniform measure $\mu(\lambda)$, and wild bootstrapped p-values. Fourth, Bierens and Ploberger's (1997) ICM $\widehat{\mathcal{I}}_n \equiv \int_{\Lambda_n} \hat{z}_n^2(\lambda)\mu(d\lambda)$ with uniform $\mu(\lambda)$, and the critical value upper bound $c_\alpha \int_\Lambda \hat{v}_n^2(\lambda)\mu(d\lambda)$, where $\{c_{.01}, c_{.05}, c_{.10}\} = \{6.81, 4.26, 3.23\}$ (Bierens and Ploberger, 1997, Section 6).

Results Rejection frequencies for $\alpha \in \{.01, .05, .10\}$ are reported in Table 1. The ICM test tends to be under sized, which is expected due to the critical value upper bound. Randomized, average and supremum tests have accurate empirical size for iid data, but exhibit size distortions for time series data when $n \in \{100, 250\}$. The PVOT test has relatively sharp size in nearly every case, but is slightly over-sized for time series data when $n = 100$.

All tests except the supremum test have comparable power, while the ICM test has low power at the 1% level. The supremum test has the lowest power, although its local power was essentially identical to the average and PVOT tests for a similar test of omitted nonlinearity (see Section 3.3). In the time series case, however, PVOT power when $n = 100$ is lower than all other tests, except the supremum test in general and the ICM test at level $\alpha = .01$. PVOT rejection frequencies are $\{.135, .206, .645\}$ for tests at levels $\{.01, .05, .10\}$, while randomized, average, supremum and ICM power are $\{.135, .592, .846\}$, $\{.062, .412, .726\}$, $\{.021, .209, .561\}$ and $\{.004, .643, .866\}$ respectively. These discrepancies, however, vanish when $n \in \{250, 500\}$. The ICM test has dismal power at the 1% level when $n \leq 250$ and much lower power than all other tests when $n = 500$, but comparable or better power at levels 5% and 10%. In summary, across cases the various tests are comparable; supremum test power is noticeably lower in many cases; and the PVOT test generally exhibits fewer size distortions, and competitive or high power in nearly every case.

Of particular note, the accuracy of PVOT size provides further evidence that the PVOT asymptotic critical value is identically α. Finally, when $n = 100$ the PVOT test took on average .0085 minutes (.51 seconds), while the bootstrapped average or supremum test took 8.07 minutes on average. The 1000-fold increase is due to the number of bootstrap samples. This demonstrates the PVOT test computational convenience, arising entirely from its asymptotic critical value (upper bound) being the test level α.

5.2 Test of Functional Form with Weak Identification

We now work with a Smooth Transition Autoregression [STAR], allowing for weak identification. The following summarize the monte carlo study in Hill (2020b).

Step-Up The data are drawn from:

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

where ϵ_t is iid $N(0,1)$. If $\varpi_0 = 0$ then y_t is a Logistic STAR process and the null hypothesis is true. If $\beta_n \to 0$ too quickly then π_0 cannot be identified and estimation asymptotics are non-standard. We use $\zeta_0 = .6$, $\pi_0 = 0$ and $\varpi_0 \in \{0, .03, .3\}$, the latter allowing for weak and strong degrees of deviation from the null. We use $\beta_n \in \{.3, .3/\sqrt{n}, 0\}$ representing strong identification, weak identification with $\sqrt{n}\beta_n = .3$ and $\beta_n \to \beta_0 = 0$, and non-identification with $\beta_n = \beta_0 = 0$.

Let $\iota = 10^{-10}$. The estimated parameters satisfy $\beta_n \in \mathcal{B}^*$, $\zeta_0 \in \mathcal{Z}^*(\beta)$ and $\pi_0 \in \Pi^*$. 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 $\Theta \equiv \mathcal{B} \times \mathcal{Z}(\beta) \times \Pi$, which ensures stationarity (see Bhattacharya and Lee, 1995, Theorem 1).

We draw 100 start values uniformly on Θ and estimate $\theta_0 = [\zeta_0, \beta_0, \pi_0]'$ by least squares for each start value, resulting in $\{\hat{\theta}_{n,i}\}_{i=1}^{100}$. The final $\hat{\theta}_n$ minimizes the least squares criterion over $\{\hat{\theta}_{n,i}\}_{i=1}^{100}$.[6] We also require $\hat{\sigma}_n^2 = 1/n \sum_{t=2}^n (y_t - \hat{\zeta}_n y_{t-1} - \hat{\beta}_n y_{t-1}(1 + \exp\{-10(y_{t-1} - \hat{\pi}_n)\})^{-1})^2$. Notice $\hat{\sigma}_n^2 \overset{p}{\to} \sigma^2$ under mild conditions and any degree of (non)identification: if $\hat{\beta}_n \overset{p}{\to} 0$ fast enough then the non-standard limit properties of $\hat{\pi}_n$ are irrelevant (see Hill, 2020b, Theorem 4.1 and Remark 7).

The test weight $\mathcal{F}(u) = 1/(1 + \exp\{u\})$, and $\mathcal{F}(\lambda'\Psi(x_t))$ uses the bounded one-to-one transform $\Psi(x) = \text{atan}(x)$ (e.g. Bierens, 1990, p. 1445, 1453). The parameter space is $\Lambda = [1, 5]$, discretized as Λ_n with endpoints $\{1, 5\}$ and equal increments with n elements (e.g. $\Lambda_{100} = \{1, 1.04, 1.08, ..., 5)$.

Tests We perform eleven tests. The first five are not robust to weak identification: (*i*) uniformly randomize λ^* on Λ, 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, 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.

[6] Computation is performed using Matlab R2016. An analytic gradient is used for optimization. The criterion tolerance for ceasing iterations is $1e^{-8}$, and the maximum number of allowed iterations is $20,000$.

The final six tests are robust based on the bootstrapped p-value procedure in Hill (2020b). We compute $\mathcal{T}_n(\lambda^*)$ using (vi) the plug-in least-favorable [LF] and (vii) plug-in Identification Category Selection Type 1 [ICS-1] p-values from (Hill, 2020b, Sections 5 and 6); $\sup_{\lambda \in \Lambda_n} p_n(\lambda)$ using (vii) the plug-in LF and (ix) plug-in ICS-1 p-values; and PVOT using (x) the plug-in LF and (xi) plug-in ICS-1 p-values. See Hill (2020b, Section 7) for details on p-value computation for the present experiment.

Results Table 2 contains rejection frequencies. All tests are fairly comparable under strong identification $\beta_n = .3$. By construction the LF p-values are larger than the ICS-1 p-values, which are larger than the χ^2 p-values. This results in lower rejection rates even under strong identification. The sup-p-value test is conservative by construction, with comparatively smaller rejection rates.

Under weak and non-identification most non-robust tests over reject the null hypothesis, and most distortions are comparatively large. Ironically, the non-robust $\sup_{\lambda \in \Lambda_n} p_n(\lambda)$ is relatively large, which pushes that test's rejection frequencies down. While this inadvertently compensates for a potentially large size distortion, it leads to lower empirical power.

The sole test that both controls for weak identification and obtains relatively high power is the PVOT test with ICS-1 p-values. The PVOT test with LF p-values also works well, but tends to have lower power than the ICS-1 based PVOT test. This follows since the LF p-values are larger than the ICS-1 p-values.

5.3 Test of GARCH Effects

Setp-Up Samples $\{y_t\}_{t=1}^n$ are drawn from a GARCH process $y_t = \sigma_t \epsilon_t$ and $\sigma_t^2 = \omega_0 + \delta_0 y_{t-1}^2 + \lambda_0 \sigma_{t-1}^2$ with parameter values $\omega_0 = 1$, $\lambda_0 = .6$, and $\delta_0 = 0$ or $.3$, where ϵ_t is iid $N(0,1)$. The initial condition is $\sigma_0^2 = \omega_0/(1-\lambda_0) = 2.5$. Simulation results are qualitatively similar for other values $\lambda_0 \in (0,1)$. Put $\Lambda = [.01, .99]$ with discretization $\Lambda_n \equiv \{.01 + 1/(\varpi n), .01 + 2/(\varpi n), ..., .01 + \bar{\imath}_n(\varpi)/(\varpi n)\}$, where $\bar{\imath}_n(\varpi) \equiv \text{argmax}\{1 \leq i \leq \varpi n : i \leq .98\varpi n\}$, with coarseness $\varpi = 1$. Hence there are $\mathcal{N}_n \approx n-1$ points in Λ_n. A finer grid based on $\varpi = 10$ or 100, for example, leads to improved empirical size at the 1% level for the PVOT test, and more severe size distortions for the supremum test. The cost, however, is computation time since a QML estimator *and* bootstrapped p-value are required for each sample. We estimate $\pi_0 = [\omega_0, \delta_0]'$ by QML for fixed $\lambda \in \Lambda_n$, with criterion $Q_n(\pi, \lambda) = \sum\{\ln \sigma_t^2(\pi, \lambda) + y_t^2/\sigma_t^2(\pi, \lambda)\}$ where $\sigma_t^2(\pi, \lambda) = \omega + \alpha y_{t-1}^2 + \lambda \sigma_{t-1}^2(\pi, \lambda)$, and $\sigma_0^2(\pi, \lambda) = \omega/(1-\lambda)$. The estimator is $\hat{\pi}_n(\lambda) = [\hat{\omega}_n(\lambda), \hat{\delta}_n(\lambda)]' = \arg\min_{\pi \in \Pi} Q_n(\pi, \lambda)$ with space $\Pi = [.001, 2] \times [0, .99]$.[7]

[7] We compute $\hat{\pi}_n(\lambda)$ using Matlab's built-in *fmincon* routine for constrained optimization, with numerical approximations for the first and second derivatives. We cease computation iterations when the numerical gradient, or the difference in the current and previous iteration of $\hat{\pi}_n(\lambda)$, is less than .0001. The initial parameter value is a uniform random uniform draw on Π.

The test statistic is $\mathcal{T}_n(\lambda) = n\hat{\delta}_n(\lambda)^2$, and the p-value approximation $\hat{p}_{\widetilde{\mathcal{R}},\widetilde{\mathcal{M}},n}(\lambda)$ is computed by the method in Section 4.2 with $\widetilde{\mathcal{M}} = 10,000$ simulated samples of size $\widetilde{\mathcal{R}} = 25,000$.

Tests We handle the nuisance parameter λ by uniformly randomizing it on Λ; computing the PVOT; and computing $\sup_{\lambda \in \Lambda} \mathcal{T}_n(\lambda)$ and $\int_{\Lambda} \mathcal{T}_n(\lambda)\mu(d\lambda)$, along with corresponding simulation-based bootstrapped p-values $\hat{p}^{(\cdot)}_{\widetilde{\mathcal{R}},\widetilde{\mathcal{M}},n}$ detailed in Remark 10.

Results Consult Table 3 for simulation results. The randomized test under rejects the null, and has lower size adjusted power than the remaining tests. Andrews' (2001) proposed supremum test is highly over-sized, resulting in relatively low size adjusted power. The best tests in terms of size and size adjusted power are the PVOT and average tests. The average test tends to under reject the null at each level for sample sizes $n \in \{100, 250\}$, and the PVOT test tends to over reject the null at the 1% level for $n \in \{100, 250\}$. Recall the average test has the highest weighted average power for alternatives near the null (Andrews and Ploberger, 1994), hence the PVOT test performs on par with, or is slightly better than, an optimal test (depending on n and α). Finally, the PVOT size performance suggests the asymptotic critical value is α. The PVOT, average and supremum tests are roughly equal in terms of computational cost due to the simulation procedure required for computing the p-value. See Remark 10.

6 Conclusion

Hill and Aguilar (2013) and Hill (2012) develop the p-value occupation time [PVOT] to smooth over a trimming tuning parameter. The idea is extended here to tests when a nuisance parameter is present under the alternative, and complete asymptotic theory is developed for the first time. We show in a likelihood setting that the PVOT is a point estimate of the weighted average rejection probability of the PV test, evaluated under the null. The average is weighted over a local alternative drift parameter and the nuisance parameter. By construction, a critical value upper bound for the PVOT test is the nominal significance level α, making computation and interpretation very simple, and much easier to perform than standard transforms like the average or supremum since these typically require a bootstrapped p-value. If the original test is consistent then so is the PVOT test. Moreover, the PVOT form of smoothing naturally accepts weak identification robust p-values, while conventionally smoothed test statistics cannot be consistently bootstrapped under weak identification. Indeed, evidently only the PVOT test with a weak identification robust p-value achieves both accurate level and high power. We are not aware of any other test statistic construction that allows for nuisance parameter smoothing that is both robust to weak identification *and* not conservative.

A Appendix: Proofs

Proof of Theorem 2.1 By Assumption 1, $\{\mathcal{T}_n(\lambda)\} \Rightarrow^* \{\mathcal{T}(\lambda)\}$ under H_0, a process with a version that has *almost surely* uniformly continuous sample paths, where $\mathcal{T}(\lambda)$ has a continuous distribution function $F_0(c) \equiv P(\mathcal{T}(\lambda) \leq c)$ that is not a function of λ. The continuous mapping theorem therefore yields $\{\bar{F}_0(\mathcal{T}_n(\lambda)) : \lambda \in \Lambda\} \Rightarrow^* \{\bar{F}_0(\mathcal{T}(\lambda)) : \lambda \in \Lambda\}$, where $\bar{F}_0(c) \equiv 1 - F_0(c)$ (e.g. Billingsley, 1999, Theorem 2.7). Furthermore, $\sup_{\lambda \in \Lambda} |p_n(\lambda) - \bar{F}_0(\mathcal{T}_n(\lambda))| \xrightarrow{p} 0$ by Assumption 1.b. It follows that $\{p_n(\lambda)) : \lambda \in \Lambda\} \Rightarrow^* \{\bar{F}_0(\mathcal{T}(\lambda)) : \lambda \in \Lambda\}$. By distribution continuity, $\mathcal{U}(\lambda) \equiv \bar{F}_0(\mathcal{T}(\lambda))$ is for each $\lambda \in \Lambda$ uniformly distributed on $[0,1]$. The continuous mapping theorem therefore yields $\mathcal{P}_n^*(\alpha) = \int_\Lambda I(p_n(\lambda) < \alpha) d\lambda \xrightarrow{d} \int_\Lambda I(\mathcal{U}(\lambda) < \alpha) d\lambda$. Now use Lemma A.1, below, to yield $P(\int_\Lambda I(\mathcal{U}(\lambda) < \alpha) d\lambda > \alpha) \leq \alpha$ and each remaining claim. $\mathcal{QED}$.

Lemma A.1. *Let $\{\mathcal{U}(\lambda) : \lambda \in \Lambda\}$ be a stochastic process where $\mathcal{U}(\lambda)$ is distributed uniform on $[0,1]$, and $\int_\Lambda d\lambda = 1$. Then (a) $P(\int_\Lambda I(\mathcal{U}(\lambda) < \alpha)d\lambda > \alpha) \leq \alpha$. In particular, (b) if $\mathcal{U}(\lambda) = \mathcal{U}(\lambda^*) = a.s.$ $\forall \lambda \in \Lambda$ and some $\lambda^* \in \Lambda$ then $P(\int_\Lambda I(\mathcal{U}(\lambda) < \alpha)d\lambda > \alpha) = \alpha$; and (c) if any h-tuple $\{\mathcal{U}(\lambda_1), ..., \mathcal{U}(\lambda_h)\}$ is jointly independent, $\lambda_i \neq \lambda_j$ for each $i \neq j$, and any $h \in \mathbb{N}$, then $\int_\Lambda I(\mathcal{U}(\lambda) < \alpha)d\lambda = \alpha$ a.s. hence $P(\int_\Lambda I(\mathcal{U}(\lambda) < \alpha)d\lambda > \alpha) = 0$. Finally, (d) if $P(\mathcal{U}(\lambda) < \alpha, \mathcal{U}(\tilde{\lambda}) < \alpha) > \alpha^2$ for all couplets $(\lambda, \tilde{\lambda})$ on a subset of $\Lambda \times \Lambda$ with positive measure, then $P(\int_\Lambda I(\mathcal{U}(\lambda) < \alpha)d\lambda > \alpha) > 0$.*

Remark 11. The key proof that $P(\int_\Lambda I(\mathcal{U}(\lambda) < \alpha)d\lambda > \alpha) \leq \alpha$ exploits a variation of the Bernstein inequalities. If we know $\mathcal{U}(\lambda)$ is perfectly dependent across λ, or $\{\mathcal{U}(\lambda_1), ..., \mathcal{U}(\lambda_h)\}$ are jointly independent, then the bound is exact.

Proof. Let $\mathcal{P} \equiv \int_\Lambda I(\mathcal{U}(\lambda) < \alpha)d\lambda$, where $\mathcal{P} \in [0,1]$ since $\int_\Lambda d\lambda = 1$. In order to prove (a), use Markov's inequality (cf. the Chermoff bound variation of the Bernstein inequality) to yield

$$P(\mathcal{P} > \alpha) \leq \inf_{k \geq 0} \left\{ e^{-k\alpha} E\left[e^{k\mathcal{P}} \right] \right\}.$$

Note that $E[\mathcal{P}^i] \leq E[\mathcal{P}]$ for all $i \geq 1$ due to $\mathcal{P} \in [0,1]$. Now invoke Fubini's theorem, the fact that $\mathcal{U}(\lambda)$ is uniformly distributed on $[0,1]$, and $\int_\Lambda d\lambda = 1$ to deduce:

$$E[\mathcal{P}] = E\left[\int_\Lambda I(\mathcal{U}(\lambda) < \alpha)d\lambda \right] = \int_\Lambda P(\mathcal{U}(\lambda) < \alpha)d\lambda = \alpha \int_\Lambda d\lambda = \alpha.$$

Expanding $E[e^{k\mathcal{P}}]$ around $k = 0$, and exploiting $E[\mathcal{P}^i] \leq \alpha$, yields:

$$P(\mathcal{P} > \alpha) \leq \inf_{k \geq 0} \left\{ e^{-k\alpha} E\left[e^{k\mathcal{P}} \right] \right\} = \inf_{k \geq 0} \left\{ e^{-k\alpha} \sum_{i=0}^{\infty} \frac{1}{i!} k^i E\left[\mathcal{P}^i \right] \right\} \leq \alpha \inf_{k \geq 0} \left\{ e^{-k\alpha} \sum_{i=0}^{\infty} \frac{1}{i!} k^i \right\}.$$

Since $\alpha \in [0,1]$ and therefore $e^{k(1-\alpha)} \geq 1$ $\forall k \geq 0$, trivially

$$\inf_{k \in \mathcal{K}} \{ e^{-k\alpha} \sum_{i=0}^{\infty} k^i / i! \} = \inf_{k \geq 0} \{ e^{-k\alpha} e^k \} = \inf_{k \geq 0} e^{k(1-\alpha)} = 1.$$

This proves $P(\mathcal{P} > \alpha) \leq \alpha$ as required.

Consider (b). If $P(\mathcal{U}(\lambda) = \mathcal{U}(\lambda^*)) = 1 \; \forall \lambda \in \Lambda$ and some λ^*, then $P(\mathcal{P} = \int_\Lambda I(\mathcal{U}(\lambda^*) < \alpha) d\lambda) = 1$. Hence $P(\mathcal{P} > \alpha) = P(I(\mathcal{U}(\lambda^*) < \alpha) > \alpha) = P(\mathcal{U}(\lambda^*) < \alpha)$. The latter is identically α by uniform distributedness.

Now consider (c), and assume every h-tuple $\{\mathcal{U}(\lambda_1), ..., \mathcal{U}(\lambda_h)\}$ is jointly independent for arbitrary $h \in \mathbb{N}$, and $\lambda_i \neq \lambda_j$ for each $i \neq j$. We have by Fubini's theorem

$$E[\mathcal{P}^2] = \int_{\lambda \neq \tilde{\lambda}} P(\mathcal{U}(\lambda) < \alpha) P(\mathcal{U}(\tilde{\lambda}) < \alpha) d\lambda d\tilde{\lambda} = \alpha^2.$$

Since $E[\mathcal{P}] = \alpha$ by Fubini's Theorem and uniformity of $\mathcal{U}(\lambda)$, it follows that $V[\int_\Lambda I(\mathcal{U}(\lambda) < \alpha) d\lambda] = 0$, therefore $\mathcal{P} = \alpha$ *a.s.*

Finally, for (d) if $P(\mathcal{U}(\lambda) < \alpha, \mathcal{U}(\tilde{\lambda}) < \alpha) > \alpha^2$ on a subset of $\Lambda \times \Lambda$ with positive measure, then $E[\mathcal{P}^2] > (E[\mathcal{P}])^2 = \alpha^2$. Since $E[\mathcal{P}^2] = E[\mathcal{P}^2 I(\mathcal{P}^2 > \alpha^2)] + E[\mathcal{P}^2 I(\mathcal{P}^2 \leq \alpha^2)]$, and $\mathcal{P}$ is bounded, by a variant of the second moment method $P(\mathcal{P} > \alpha) \geq (E[\mathcal{P}^2] - \alpha^2)^2 / E[\mathcal{P}^4] > 0$. $\mathcal{QED}$.

Proof of Theorem 2.2.

Claim (a). Let H_0 be false, and define the set of $\lambda's$ such that we reject the PV test for sample size n: $\Lambda_{n,\alpha} \equiv \{\lambda \in \Lambda : p_n(\lambda) < \alpha\}$. The set $\Lambda_{n,\alpha}$ is stochastic with Lebesgue measure $\mathcal{M}_{n,\alpha}$. By construction

$$\mathcal{P}_n^*(\alpha) \equiv \int_{\Lambda_{n,\alpha}} I\left(p_n(\lambda) < \alpha\right) d\lambda + \int_{\Lambda/\Lambda_{n,\alpha}} I\left(p_n(\lambda) < \alpha\right) d\lambda = \int_{\Lambda_{n,\alpha}} d\lambda.$$

Hence $\lim_{n\to\infty} P(\mathcal{P}_n^*(\alpha) > \alpha) = \lim_{n\to\infty} P(\int_{\Lambda_{n,\alpha}} d\lambda > \alpha)$. Therefore $\lim_{n\to\infty} P(\mathcal{P}_n^*(\alpha) > \alpha) > 0$ *if and only if* $\lim_{n\to\infty} P(\mathcal{M}_{n,\alpha} > \alpha) > 0$, *if and only if* $\lim_{n\to\infty} P(p_n(\lambda) < \alpha) > 0$ on some subset with measure greater than α.

Claim (b). Let Λ_α denote the set of $\lambda's$ such that $\lim_{n\to\infty} P(p_n(\lambda) < \alpha) = 1$, hence $\lim_{n\to\infty} P(p_n(\lambda) < \alpha) < 1$ on Λ/Λ_α. Then by dominated convergence $\lim_{n\to\infty} P(\mathcal{P}_n^*(\alpha) > \alpha) = \lim_{n\to\infty} P(\int_{\Lambda_\alpha} d\lambda + \int_{\Lambda/\Lambda_\alpha} I(p_n(\lambda) < \alpha) d\lambda > \alpha)$. If Λ_α has measure greater than α then $\lim_{n\to\infty} P(\mathcal{P}_n^*(\alpha) > \alpha) = 1$. $\mathcal{QED}$.

Proof of Theorem 3.2. Recall F_1 is a $\chi^2(1)$ distribution, $\bar{F}_1 \equiv 1 - F_1$, and $F_{1,v}$ is a noncentral chi-squared distribution with noncentrality v. By construction $p_n(\lambda) = \bar{F}_1(\mathcal{T}_n(\lambda))$.

In view of (9), under H_1^L it follows $p_n(\lambda) \xrightarrow{d} \bar{F}_1(\mathfrak{T}_b)$, a law on $[0,1]$ where $\mathfrak{T}_b$ is distributed $F_{1,b^2 c(\lambda)^2}$. Hence $\bar{F}_1(\mathfrak{T}_b)$ is skewed left for $b \neq 0$. Let $\mathcal{U}_b(\lambda)$ be distributed $\bar{F}_0(\mathfrak{T}_b)$. Then $\mathcal{U}_0(\lambda)$ is a uniform random variable, and in general $P(\mathcal{U}_b(\lambda) \leq a) - P(\mathcal{U}_0(\lambda) \leq a) > 0$ is monotonically increasing in b since $P(\mathcal{U}_b(\lambda) \leq a) \to 1$ is monotonic as $|b| \to \infty$ for any a.

Now, by construction $\{\mathcal{U}_b(\lambda)\}$ has *almost surely* continuous sample paths with $\mathcal{U}_b(\lambda)$ distributed $F_1(\mathfrak{T}_b)$. Hence under H_1^L by (9), and the continuous mapping theorem:

$$\mathcal{P}_n^*(\alpha) = \int_\Lambda I\left(p_n(\lambda) < \alpha\right) d\lambda \xrightarrow{d} \int_\Lambda I\left(\mathcal{U}_b(\lambda) < \alpha\right) d\lambda.$$

By construction $\int_\Lambda I(\mathcal{U}_b(\lambda) < \alpha)d\lambda \geq \int_\Lambda I(\mathcal{U}_0(\lambda) < \alpha)d\lambda$ with equality only if $b = 0$: the asymptotic occupation time of a p-value rejection $p_n(\lambda) < \alpha$ is higher under any sequence of non-trivial local alternatives $H_1^L : \beta_0 = b/n^{1/2}$, $b \neq 0$. Further, $\int_\Lambda I(\mathcal{U}_b(\lambda) < \alpha)d\lambda \to 1$ as $|b| \to \infty$. Hence as the local deviation from the null increases, the probability of a PVOT test rejection of H_1^L approaches one $\lim_{n\to\infty} P(\mathcal{P}_n^*(\alpha) > \alpha) \nearrow 1$ for any nominal level $\alpha \in [0, 1)$. $\mathcal{QED}$.

Proof of Theorem 4.2. Since the GARCH process is stationary and has an iid error with a finite fourth moment, the claim follows from arguments in Andrews (2001, Section 4.1). $\mathcal{QED}$.

Proof of Theorem 4.3. By Theorem 4.2, the limit process of $\{\mathcal{T}_n(\lambda)\}$ under H_0 is $\{\mathcal{T}(\lambda)\}$, where $\mathcal{T}(\lambda) = (\max\{0, \mathcal{Z}(\lambda)\})^2$ and $\{\mathcal{Z}(\lambda)\}$ is Gaussian with covariance $E[\mathcal{Z}(\lambda_1)\mathcal{Z}(\lambda_2)] = (1 - \lambda_1^2)(1 - \lambda_2^2)/(1 - \lambda_1\lambda_2)$. Define $\bar{F}_0(c) = P(\mathcal{T}(\lambda) \geq c)$ and $p_n(\lambda) \equiv \bar{F}_0(\mathcal{T}_n(\lambda))$, the asymptotic p-value. Define $\mathcal{D}_n \equiv \sup_{\lambda\in\Lambda} |\hat{p}_{\widetilde{\mathcal{R}}_n,\widetilde{\mathcal{M}}_n,n}(\lambda) - p_n(\lambda)|$. Theorems 2.1 and 2.2 apply by Theorem 4.2. Hence, by Lemma A.2, below, and weak convergence arguments developed in the proof of Theorem 2.1, under H_0 for some uniform process $\{\mathcal{U}(\lambda)\}$:

$$\begin{aligned}\int_\Lambda I\left(\mathcal{U}(\lambda) < \alpha\right) d\lambda \overset{d}{\leftarrow} \int_\Lambda I\left(p_n(\lambda) - \mathcal{D}_n < \alpha\right) d\lambda &\geq \int_\Lambda I\left(\hat{p}_{\widetilde{\mathcal{R}}_n,\widetilde{\mathcal{M}}_n,n}(\lambda) < \alpha\right) d\lambda \\ &\geq \int_\Lambda I\left(p_n(\lambda) + \mathcal{D}_n < \alpha\right) d\lambda \overset{d}{\to} \int_\Lambda I\left(\mathcal{U}(\lambda) < \alpha\right) d\lambda.\end{aligned}$$

Therefore $\int_\Lambda I(\hat{p}_{\widetilde{\mathcal{R}}_n,\widetilde{\mathcal{M}}_n,n}(\lambda) < \alpha)d\lambda \overset{d}{\to} \int_\Lambda I(\mathcal{U}(\lambda) < \alpha)d\lambda$. The claim now follows from the proof of Theorem 2.1 and the fact that $\{\mathcal{T}(\lambda)\}$ is weakly dependent in the sense of Lemma A.1.c. $\mathcal{QED}$.

Lemma A.2. $\sup_{\lambda\in\Lambda} |\hat{p}_{\widetilde{\mathcal{R}}_n,\widetilde{\mathcal{M}}_n,n}(\lambda) - p_n(\lambda)| \overset{p}{\to} 0$ *where* $\hat{p}_{\widetilde{\mathcal{R}}_n,\widetilde{\mathcal{M}}_n,n}(\lambda) \equiv 1/\widetilde{\mathcal{M}}_n \sum_{i=1}^{\widetilde{\mathcal{M}}_n} I(\mathcal{T}_{\widetilde{\mathcal{R}}_n,i}(\lambda) \geq \mathcal{T}_n(\lambda))$.

Proof. We first state known properties and define some terms. Assumption 1 applies to $\mathcal{T}_n(\lambda)$ by Theorem 4.2, where $\{\mathcal{T}_n(\lambda)\} \Rightarrow^* \{\mathcal{T}(\lambda)\}$, $\mathcal{T}(\lambda) = (\max\{0, \mathcal{Z}(\lambda)\})^2$, and $\{\mathcal{Z}(\lambda)\}$ is a zero mean Gaussian process with a version that has *almost surely* continuous sample paths, and covariance function $(1 - \lambda_1^2)(1 - \lambda_2^2)/(1 - \lambda_1\lambda_2)$ for $\lambda_1, \lambda_2 \in \Lambda$. Recall we have samples $\{Z_{j,i}\}_{j=1}^{\widetilde{\mathcal{R}}}$ where $Z_{j,i} \overset{iid}{\sim} N(0, 1)$, and for $(\widetilde{\mathcal{R}}, \widetilde{\mathcal{M}}) \in \mathbb{N}$:

$$\mathfrak{z}_{\widetilde{\mathcal{R}},i}(\lambda) \equiv (1 - \lambda^2) \sum_{j=1}^{\widetilde{\mathcal{R}}} \lambda^j Z_{j,i} \text{ and } \mathcal{T}_{\widetilde{\mathcal{R}},i}(\lambda) \equiv \left(\max\{0, \mathfrak{z}_{\widetilde{\mathcal{R}},i}(\lambda)\}\right)^2 \text{ for } i = 1, ..., \widetilde{\mathcal{M}}.$$

$\mathfrak{z}_\infty(\lambda)$ has the same functional Gaussian distribution as $\mathcal{Z}(\lambda)$, and therefore $(\max\{0, \mathfrak{z}_\infty(\lambda)\})^2$ is a random draw from the distribution of $\mathcal{T}(\lambda)$. The distribution $\bar{F}_0(c) \equiv P(\mathcal{T}(\lambda) \geq c)$ is continuous and not a function of λ under Assumption 1. Hence, the p-value is identically $p_n(\lambda) = \bar{F}_0(\mathcal{T}_n(\lambda))$. Let $\{\mathcal{T}_{1,i}(\lambda)\}_{i=1}^{\widetilde{\mathcal{M}}}$ and $\mathcal{T}_2(\lambda)$ be iid copies of $\mathcal{T}(\lambda)$, and define

$$\mathcal{T}_{\widetilde{\mathcal{R}}}^{(\widetilde{\mathcal{M}})}(\lambda) \equiv \left[\mathcal{T}_{\widetilde{\mathcal{R}},i}(\lambda), ..., \mathcal{T}_{\widetilde{\mathcal{R}},\widetilde{\mathcal{M}}}(\lambda)\right]' \text{ and } \mathcal{T}_1^{(\widetilde{\mathcal{M}})}(\lambda) \equiv \left[\mathcal{T}_{1,i}(\lambda), ..., \mathcal{T}_{1,\widetilde{\mathcal{M}}}(\lambda)\right]'.$$

The arguments in Andrews (2001, Section 4.1) for weak convergence of $\{\mathcal{T}_n(\lambda)\}$ trivially extend to

$[\mathcal{T}_{\widetilde{\mathcal{R}}_n}^{(\widetilde{\mathcal{M}})}(\lambda)', \mathcal{T}_n(\lambda)]'$ in view of independence of the individual processes, and normality and smoothness of $\mathfrak{Z}_{\widetilde{\mathcal{R}}_n,i}(\lambda)$. Specifically, there exist $\mathcal{T}_1^{(\widetilde{\mathcal{M}})}(\lambda)$ and $\mathcal{T}_2(\lambda)$ such that:

$$\left\{ \left[\begin{array}{c} \mathcal{T}_{\widetilde{\mathcal{R}}_n}^{(\widetilde{\mathcal{M}})}(\lambda) \\ \mathcal{T}_n(\lambda) \end{array} \right] : \lambda \in \Lambda \right\} \Rightarrow^* \left\{ \left[\begin{array}{c} \mathcal{T}_1^{(\widetilde{\mathcal{M}})}(\lambda) \\ \mathcal{T}_2(\lambda) \end{array} \right] : \lambda \in \Lambda \right\} \text{ as } n \to \infty \text{ for each } \widetilde{\mathcal{M}} \in \mathbb{N}.$$

Hence, by two applications of the continuous mapping theorem, for each $\widetilde{\mathcal{M}} \in \mathbb{N}$ as $n \to \infty$:

$$\begin{aligned} \left\{ \hat{p}_{\widetilde{\mathcal{R}}_n,\widetilde{\mathcal{M}},n}(\lambda) - \bar{F}_0(\mathcal{T}_n(\lambda)) : \lambda \in \Lambda \right\} &= \left\{ \frac{1}{\widetilde{\mathcal{M}}} \sum_{i=1}^{\widetilde{\mathcal{M}}} I\left(\mathcal{T}_{\widetilde{\mathcal{R}}_n,i}(\lambda) \geq \mathcal{T}_n(\lambda) \right) - \bar{F}_0(\mathcal{T}_n(\lambda)) : \lambda \in \Lambda \right\} \\ &\Rightarrow^* \left\{ \frac{1}{\widetilde{\mathcal{M}}} \sum_{i=1}^{\widetilde{\mathcal{M}}} I\left(\mathcal{T}_{1,i}(\lambda) \geq \mathcal{T}_2(\lambda) \right) - \bar{F}_0(\mathcal{T}_2(\lambda)) : \lambda \in \Lambda \right\} \end{aligned}$$

and

$$\sup_{\lambda \in \Lambda} \left| \hat{p}_{\widetilde{\mathcal{R}}_n,\widetilde{\mathcal{M}},n}(\lambda) - \bar{F}_0(\mathcal{T}_n(\lambda)) \right| \xrightarrow{d} \sup_{\lambda \in \Lambda} \left| \frac{1}{\widetilde{\mathcal{M}}} \sum_{i=1}^{\widetilde{\mathcal{M}}} I\left(\mathcal{T}_{1,i}(\lambda) \geq \mathcal{T}_2(\lambda) \right) - \bar{F}_0(\mathcal{T}_2(\lambda)) \right| \text{ as } n \to \infty.$$

The proof is complete if we show

$$\sup_{\lambda \in \Lambda} \left| \frac{1}{\widetilde{\mathcal{M}}} \sum_{i=1}^{\widetilde{\mathcal{M}}} I(\mathcal{T}_{1,i}(\lambda) \geq \mathcal{T}_2(\lambda)) - \bar{F}_0(\mathcal{T}_2(\lambda)) \right| \xrightarrow{p} 0 \text{ as } \widetilde{\mathcal{M}} \to \infty, \tag{A.1}$$

since this means $\sup_{\lambda \in \Lambda} |\hat{p}_{\widetilde{\mathcal{R}}_n,\widetilde{\mathcal{M}},n}(\lambda) - \bar{F}_0(\mathcal{T}_n(\lambda))|$ can be made arbitrarily close to zero in probability by choice of $\widetilde{\mathcal{M}}$. Note that by construction $\bar{F}_0(\mathcal{T}_2(\lambda)) = E[I(\mathcal{T}_{1,i}(\lambda) \geq \mathcal{T}_2(\lambda))|\mathcal{T}_2(\lambda)]$ since $\mathcal{T}_{1,i}(\lambda)$ and $\mathcal{T}_2(\lambda)$ are iid copies of $\mathcal{T}(\lambda)$. We therefore derive a uniform LLN for

$$\mathcal{I}_i(\lambda) \equiv I\left(\mathcal{T}_{1,i}(\lambda) \geq \mathcal{T}_2(\lambda) \right) - E\left[I\left(\mathcal{T}_{1,i}(\lambda) \geq \mathcal{T}_2(\lambda) \right) | \mathcal{T}_2(\lambda) \right].$$

Since $(\mathcal{T}_{1,i}(\lambda), \mathcal{T}_2(\lambda))$ are iid copies of $\mathcal{T}(\lambda)$, it follows $E[\bar{F}_0(\mathcal{T}_2(\lambda))] = P(\mathcal{T}_{1,i}(\lambda) \geq \mathcal{T}_2(\lambda))$ hence:

$$E\left[\mathcal{I}_i(\lambda)\right] = P\left(\mathcal{T}_{1,i}(\lambda) \geq \mathcal{T}_2(\lambda) \right) - E\left[\bar{F}_0(\mathcal{T}_2(\lambda)) \right] = P\left(\mathcal{T}_{1,i}(\lambda) \geq \mathcal{T}_2(\lambda) \right) - P\left(\mathcal{T}_{1,i}(\lambda) \geq \mathcal{T}_2(\lambda) \right) = 0.$$

Second, $1/\widetilde{\mathcal{M}} \sum_{i=1}^{\widetilde{\mathcal{M}}} \mathcal{I}_i(\lambda) \xrightarrow{p} 0$ as $\widetilde{\mathcal{M}} \to \infty$ pointwise on Λ since $\mathcal{I}_i(\lambda)$ is iid, and $E[\mathcal{I}_i(\lambda)] = 0$.

It remains to demonstrate $\{\mathcal{I}_i(\lambda) : \lambda \in \Lambda\}$ is stochastically equicontinuous: $\forall (\epsilon, \eta) > 0$ there exists $\delta > 0$ such that (see, e.g., Pollard 1984, and Billingsley 1999, Chap. 7):

$$P\left(\sup_{\lambda, \tilde{\lambda} \in \Lambda : ||\lambda - \tilde{\lambda}|| \leq \delta} \left| \frac{1}{\widetilde{\mathcal{M}}} \sum_{i=1}^{\widetilde{\mathcal{M}}} \left\{ \mathcal{I}_i(\lambda) - \mathcal{I}_i(\tilde{\lambda}) \right\} \right| > \eta \right) < \varepsilon.$$

The function $\mathcal{I}_i : \Lambda \rightarrow [-1, 1]$ is not continuous. We therefore adapt arguments developed in Arcones and Yu (1994, proof of Theorem 2.1 and Lemma 2.1), which requires the notion of the *V-C subgraph* class of functions, denoted $\mathcal{V}(\mathcal{C})$. See Vapnik and Červonenkis (1971) and Dudley (1978, Section 7), and see Pollard (1984, Chap. II.4) for the closely related *polynomial discrimination* class. We use the following well known properties: $\mathcal{V}(\mathcal{C})$ contains continuous functions and the indicator function; $\mathcal{V}(\mathcal{C})$ contains linear combinations of $\mathcal{V}(\mathcal{C})$ functions; and $\mathcal{V}(\mathcal{C})$ transforms of $\mathcal{V}(\mathcal{C})$ functions are in $\mathcal{V}(\mathcal{C})$. Cf. Vapnik and Červonenkis (1971), Dudley (1978, Section 7) and Pollard (1990).

By using the approach of Arcones and Yu (1994), we may show that $1/\widetilde{\mathcal{M}} \sum_{i=1}^{\widetilde{\mathcal{M}}} \mathcal{I}_i(\lambda)$ is stochastically equicontinuous. $\mathcal{T}_{1,i}(\lambda)$ and $\mathcal{T}_2(\lambda)$ are, respectively, versions of $(\max\{0, \mathfrak{Z}_{1,\infty,i}(\lambda))^2$ and $(\max\{0, \mathfrak{Z}_{2,\infty}(\lambda))^2$, where $\mathfrak{Z}_{1,\infty,i}(\lambda)$ and $\mathfrak{Z}_{2,\infty}(\lambda)$ are independent copies of $\mathfrak{Z}_\infty(\lambda)$, and $\mathfrak{Z}_\infty(\lambda) \equiv (1 - \lambda^2) \sum_{j=0}^{\infty} \lambda^j Z_j$ is zero mean Gaussian with the same covariance function as $\mathcal{Z}(\lambda)$. By construction $\mathfrak{Z}_\infty(\lambda)$ is continuous in λ, hence it lies in $\mathcal{V}(\mathcal{C})$. Further, $(\max\{0, \cdot)^2$ lies in $\mathcal{V}(\mathcal{C})$. Therefore $(\max\{0, \mathfrak{Z}_\infty(\lambda))^2$ lies in $\mathcal{V}(\mathcal{C})$, which implies $\mathcal{T}_{1,i}(\lambda)$ and $\mathcal{T}_2(\lambda)$ have versions that lie in $\mathcal{V}(\mathcal{C})$. Hence $\mathcal{T}_{1,i}(\lambda) - \mathcal{T}_2(\lambda)$ has a version in $\mathcal{V}(\mathcal{C})$. Therefore $I(\mathcal{T}_{1,i}(\lambda) - \mathcal{T}_2(\lambda) \geq 0)$ has a version in $\mathcal{V}(\mathcal{C})$. Moreover, the continuous transform $\bar{F}_0(\mathcal{T}_2(\lambda))$ lies in $\mathcal{V}(\mathcal{C})$. Hence the difference $\mathcal{I}_i(\lambda) \equiv I(\mathcal{T}_{1,i}(\lambda) \geq \mathcal{T}_2(\lambda)) - \bar{F}_0(\mathcal{T}_2(\lambda))$ lies in $\mathcal{V}(\mathcal{C})$. This, and boundedness of $\mathcal{I}_i(\lambda)$, imply that the covering numbers with respect to the L_p-metric satisfy, for any $p > 2$, $\mathcal{N}(\varepsilon, \Lambda, || \cdot ||_p) < a\varepsilon^{-b}$ for all $\varepsilon \in (0, 1)$ and some $a, b > 0$ that may depend on p (e.g. Lemma 7.13 in Dudley, 1978, and Lemma II.25 in Pollard, 1984). Further, $\mathcal{I}_i(\lambda)$ is uniformly bounded and iid. Therefore $\{\mathcal{I}_i(\lambda) : \lambda \in \Lambda\}$ is stochastically equicontinuous by adapting the proof of Lemma 2.1 in Arcones and Yu (1994): see especially Arcones and Yu (1994, eq. (2.13)). $\mathcal{QED}$.

Bibliography

Aguilar, M., and J. B. Hill (2015): "Robust Score and Portmanteau Tests of Volatility Spillover," *Journal of Econometrics*, 184, 37–61.

Aguirre-Torres, V., and A. R. Gallant (1983): "The Null and Non-null Asymptotic Distribution of the Cox Test for Multivariate Nonlinear Regression Alternatives and a New Distribution Free Cox Test," *Journal of Econometrics*, 21, 5–33.

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

——— (1993): "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica*, 61, 821–856.

——— (1999): "Estimation When a Parameter is on a Boundary," *Econometrica*, 67, 1341–1383.

——— (2001): "Testing when a Parameter is on the Boundary of the Maintained Hypothesis," *Econometrica*, 69, 683–734.

Andrews, D. W. K., and X. Cheng (2012): "Estimation and Inference with Weak, Semi-Strong and Strong Identification," *Econometrica*, 80, 2153–2211.

——— (2013): "Maximum Likelihood Estimation and Uniform Inference with Sporadic Identification Failure," *Journal of Econometrics*, 173, 36–56.

——— (2014): "GMM Estimation and Uniform Subvector Inference with Possible Identification Failure," *Econometric Theory*, 30, 287–333.

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

——— (1995): "Admissibility of the Likelihood Ratio Test When a Nuisance Parameter is Present Only Under the Alternative," *Annals of Statistics*, 23, 1609–1629.

Arcones, M. A., and B. Yu (1994): "Central Limit Theorems for Empirical U-Processes of Stationary Mixing Sequences," *Journal of Theoretical Probability*, 7, 47–71.

Bandi, F. M., and P. C. B. Phillips (2007): "A Simple Approach to the Parametric Estimation of Potentially Nonstationary Diffusions," *Journal of Econometrics*, 137, 354–395.

Bartkiewicz, K., A. Jakubowski, T. Mikosch, and O. Wintenberger (2010): "Stable Limits for Sums of Dependent Infinite Variance Random Variables," *Probability Theory and Related Fields*, 150, 337–372.

Bhattacharya, R., and C. Lee (1995): "On Geometric Ergodicity of Nonlinear Autoregressive Models," *Statistics and Probability Letters*, 22, 311–315.

Bierens, H. J. (1982): "Consistent Model Specification Tests," *Journal of Econometrics*, 20, 105–134.

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

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

Billingsley, P. (1999): *Convergence of Probability Measures*. Wiley, 2nd edn.

Bollerslev, T. (1986): "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

Boning, B. W., and F. Sowell (1999): "Optimality for the Integrated Conditional Moment Test," *Econometric Theory*, 15, 710–718.

Chernoff, H., and 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.

Davies, R. (1977): "Hypothesis Testing when a Nuisance Parameter is Present Only under the Alternative," *Biometrika*, 64, 247–254.

——— (1987): "Hypothesis Testing when a Nuisance Parameter is Present Only under the Alternative," *Biometrika*, 74, 33–43.

de Jong, R. M. (1996): "On the Bierens Test under Data Dependence," *Journal of Econometrics*, 72, 1–32.

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

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

Dufour, J.-M. (1997): "Some Impossibility Theorems in Econometrics With Applications to Structural and Dynamic Models," *Econometrica*, 65, 1365–1387.

ENGLE, R., D. LILIEN, AND R. ROBINS (1987): "Estimating Time Varying Risk Premia in the Term Structure: The ARCH-M Model," *Econometrica*, 55, 391–407.

GALLANT, A. R. (1977): "Testing a Nonlinear Regression Specification: A Nonregular Case," *Journal of the Americal Statistical Society*, 72, 523–530.

——— (1981): "On the Bias in Flexible Functional Forms and an Essentially Unbiased Form: The Fourier Flexible Form," *Journal of Econometrics*, 15, 211–245.

——— (1984): "The Fourier Flexible Form," *American Journal of Agricultural Economics*, 66, 204–208.

GALLANT, A. R., AND G. H. GOLUB (1984): "Imposing curvature restrictions on flexible functional forms," *Journal of Econometrics*, 26, 295–321.

GHYSELS, E., J. B. HILL, AND K. MOTEGI (2016): "Testing for Granger Causality with Mixed Frequency Sata," *Journal of Econometrics*, 192, 207–230.

GHYSELS, E., P. SANTA-CLARA, AND R. VALKANOV (2004): "The MIDAS Touch: Mixed Data Sampling Regression," *Working Paper, UNC and UCLA*.

GINE, E., AND J. ZINN (1984): "Some Limit Theorems for Empirical Processes," *Annals of Probability*, 12, 929–989.

HANSEN, B. E. (1996): "Inference When a Nuisance Parameter Is Not Identified Under the Null Hypothesis," *Econometrica*, 64, 413–430.

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.

——— (2012): "Heavy-Tail and Plug-In Robust Consistent Conditional Moment Tests of Functional Form," in *Festschrift in Honor of Hal White*, ed. by X. Chen, and N. Swanson, pp. 241–274. Springer: New York.

——— (2020a): "Supplemental Material for 'A Smoothed P-Value Test When There is a Nuisance Parameter under the Alternative'," Dept. of Economics, University of North Carolina.

——— (2020b): "Weak-Identification Robust Wild Bootstrap applied to a Consistent Model Specification Test," *Econometric Theory*, in press.

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

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

HONG, Y., AND H. WHITE (1995): "Consistent Specification Testing Via Nonparametric Series Regression," *Econometrica*, 63, 1133–1159.

KING, M., AND T. SHIVELY (1993): "Locally Optimal Testing When a Nuisance Parameter is Present Only Under the Alternative," *Review of Economics and Statistics*, 75, 1–7.

LEHMANN, E. L. (1994): *Testing Statistical Hypotheses*. Chapman and Hall, New York, 2nd edn.

LIU, R. Y. (1988): "Bootstrap Procedures under some Non-I.I.D. Models," *Annals of Statistics*, 16, 1696–1708.

Nelson, D. (1990): "Stationarity and Persistence in the GARCH(1,1) Model," *Econometric Theory*, 6, 318–334.

Nyblom, J. (1989): "Testing for the Constancy of Parameters over Time," *Journal of the American Statistical Association,*, 84, 223–230.

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

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

Shao, X. (2010): "The Dependent Wild Bootstrap," *Journal of the American Statistical Association*, 105, 218–235.

Silvapulle, M. (1996): "A Test in the Presence of Nuisance Parameters," *Journal of the American Statistical Association*, 91, 1690–1693.

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

Terasvirta, T. (1994): "Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models," *Journal of the American Statistical Association*, 89, 208 218.

Vapnik, V. K., and 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.

Wald, A. (1943): "Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations Is Large," *Transactions of the American Mathematical Society*, 12, 1–19.

White, H. (1989): "An Additional Hidden Unit Test for Neglected Nonlinearity in Multilayer Feedforward Networks," in *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.

Table 1: Functional Form Test Rejection Frequencies

iid data: linear vs. quadratic										
		$n = 100$			$n = 250$			$n = 500$		
Hyp[a]	Test	1%	5%	10%	1%	5%	10%	1%	5%	10%
H_0	sup-p_n [b]	.008[c]	.058	.108	.000	.039	.094	.009	.043	.091
	sup-$\mathcal{T}_n$ [d]	.004	.037	.097	.008	.041	.083	.019	.058	.096
	aver-$\mathcal{T}_n$	.014	.057	.116	.007	.040	.088	.018	.071	.109
	rand-$\mathcal{T}_n$ [e]	.014	.056	.117	.011	.045	.094	.021	.059	.109
	ICM[f]	.001	.033	.086	.001	.014	.075	.003	.062	.086
	PVOT[g]	.013	.056	.116	.010	.044	.092	.014	.063	.108
H_1	sup-p_n	.042	.162	.258	.137	.337	.473	.339	.597	.695
	sup-$\mathcal{T}_n$	.051	.156	.251	.160	.331	.512	.354	.539	.743
	aver-$\mathcal{T}_n$	.051	.211	.316	.193	.377	.576	.412	.643	.776
	rand-$\mathcal{T}_n$	.051	.221	.316	.212	.392	.586	.404	.668	.798
	ICM	.001	.149	.329	.043	.330	.606	.163	.678	.809
	PVOT	.058	.224	.320	.232	.391	.604	.404	.614	.783

time series data: AR vs. SETAR										
		$n = 100$			$n = 250$			$n = 500$		
Hyp	Test	1%	5%	10%	1%	5%	10%	1%	5%	10%
H_0	sup-p_n	.022	.075	.158	.008	.052	.113	.020	.064	.116
	sup-$\mathcal{T}_n$	.001	.003	.039	.002	.012	.036	.003	.052	.124
	aver-$\mathcal{T}_n$	.002	.022	.082	.002	.013	.066	.008	.072	.132
	rand-$\mathcal{T}_n$	.021	.113	.193	.001	.03	.114	.018	.082	.143
	ICM[f]	.002	.058	.132	.000	.030	.066	.005	.038	.089
	PVOT[g]	.016	.076	.145	.011	.047	.115	.016	.061	.114
H_1	sup-p_n	.108	.596	.845	.925	1.00	1.00	1.00	1.00	1.00
	sup-$\mathcal{T}_n$	.021	.209	.561	.685	1.00	1.00	1.00	1.00	1.00
	aver-$\mathcal{T}_n$	.062	.412	.726	.888	1.00	1.00	1.00	1.00	1.00
	rand-$\mathcal{T}_n$	.135	.592	.846	.960	1.00	1.00	1.00	1.00	1.00
	ICM	.004	.643	.866	.108	.928	1.00	.712	1.00	1.00
	PVOT	.135	.647	.883	.957	1.00	1.00	1.00	1.00	1.00

a. H_0 is $E[\epsilon|x] = 0$. b. sup-p_n is the $\sup_{\lambda \in \Lambda} p_n(\lambda)$ test. and ave-$\mathcal{T}_n$ tests are based on a wild bootstrapped p-value. c. Rejection frequency at the given level. Empirical power is not size-adjusted. d. sup-$\mathcal{T}_n$ e. rand-$\mathcal{T}_n$ is an asymptotic χ^2 test based on $\mathcal{T}_n(\lambda)$ with randomized λ on [0,1]. f. The ICM test is based on critical value upper bounds in Bierens and Ploberger (1997). g. PVOT: *p-value occupation time* test.

Table 2: A. 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
	Non-Identification: $\beta_n = \beta_0 = 0$								
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
PVOT LF	.015	.061	.117	.122	.322	.415	.740	.911	.936
PVOT ICS-1	.015	.057	.116	.253	.464	.570	.847	.939	.954

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 randomized 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 $\kappa_n = \ln(\ln(n))$.

Table 2: B. 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
	Non-Identification: $\beta_n = \beta_0 = 0$								
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
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

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 randomized 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 $\kappa_n = \ln(\ln(n))$.

Table 2: C. 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
	Non-Identification: $\beta_n = \beta_0 = 0$								
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
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

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 randomized 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 $\kappa_n = \ln(\ln(n))$.

Table 3: GARCH Effects Test Rejection Frequencies

	$n=100$			$n=250$			$n=500$		
Test	1%	5%	10%	1%	5%	10%	1%	5%	10%
No GARCH Effects (empirical size)[a]									
sup-p_n [b]	.000[c]	.000	.000	.000	.000	.000	.000	.000	.000
sup-$\mathcal{T}_n$ [d]	.160	.198	.248	.148	.188	.224	.241	.294	.321
ave-$\mathcal{T}_n$	.004	.032	.052	.005	.031	.059	.008	.053	.107
rand-$\mathcal{T}_n$ [e]	.004	.004	.012	.007	.017	.027	.003	.028	.038
PVOT[f]	.015	.059	.096	.019	.059	.091	.015	.063	.111
GARCH Effects (empirical power)									
sup-p_n	.006	.014	.017	.000	.010	.017	.003	.011	.015
sup-$\mathcal{T}_n$	.848	.934	.934	.976	.979	.988	1.00	1.00	1.00
ave-$\mathcal{T}_n$	.733	.891	.904	.974	.978	.986	1.00	1.00	1.00
rand-$\mathcal{T}_n$	.446	.555	.633	.756	.818	.846	.873	.923	.935
PVOT	.788	.914	.914	.975	.988	.988	1.00	1.00	1.00
GARCH Effects (size adjusted power)									
sup-p_n	.006	.014	.017	.000	.010	.017	.003	.011	.015
sup-$\mathcal{T}_n$	.698	.786	.786	.838	.841	.864	.769	.756	.779
ave-$\mathcal{T}_n$	.739	.909	.952	.979	.997	1.00	1.00	.997	.993
rand-$\mathcal{T}_n$	.452	.601	.721	.759	.851	.919	.880	.945	.997
PVOT	.774	.902	.902	.966	.979	.997	.995	.987	.989

a. The GARCH volatility process is $\sigma_t^2 = \omega_0 + \delta_0 y_{t-1}^2 + \lambda_0 \sigma_{t-1}^2$ with initial condition $\sigma_t^2 = \omega_0/(1-\lambda_0))$. The null hypothesis is no GARCH effects $\delta_0 = 0$, and under the alternative $\delta_0 = .3$. In all cases the true $\lambda_0 = .6$. b. sup-p_n is the $\sup_{\lambda \in \Lambda} p_n(\lambda)$ test. c. sup-$\mathcal{T}_n$ and ave-$\mathcal{T}_n$ tests are based on a wild bootstrapped p-value. d. Rejection frequency at the given significance level. e. rand-$\mathcal{T}_n$ is an asymptotic χ^2 test based on $\mathcal{T}_n(\lambda)$ with randomized λ on [.01,.99]. f. PVOT: *p-value occupation time* test.