Generalized Runs Tests for the IID Hypothesis*

JIN SEO CHO
School of Economics
Yonsei University
Email: jinseocho@yonsei.ac.kr

HALBERT WHITE
Department of Economics
University of California, San Diego
Email: hwhite@weber.ucsd.edu

First version: March, 2004. This version: July, 2010.

Abstract

We provide a family of tests for the IID hypothesis based on generalized runs, powerful against unspecified alternatives, providing a useful complement to tests designed for specific alternatives, such as serial correlation, GARCH, or structural breaks. Our tests have appealing computational simplicity in that they do not require kernel density estimation, with the associated challenge of bandwidth selection. Simulations show levels close to nominal asymptotic levels. Our tests have power against both dependent and heterogeneous alternatives, as both theory and simulations demonstrate.

Key Words IID condition, Runs test, Geometric Distribution, Gaussian process, Dependence, Structural Break.
JEL Classification C12, C23, C80.

1 Introduction

The assumption that data are independent and identically distributed (IID) plays a central role in the analysis of economic data. In cross-section settings, the IID assumption holds under pure random sampling. As Heckman (2001) notes, violation of the IID property, therefore random sampling, can indicate the presence of sample selection bias. The IID assumption is also important in time-series settings, as processes driving time series of interest are often assumed to be IID. Moreover, transformations of certain time series can be shown to be IID under specific null hypotheses. For example Diebold, Gunther, and Tay (1998) show that

***Acknowledgements** The authors are grateful to the Co-editor, Ronald Gallant, and two anonymous referees for their very helpful comments. Also, they have benefited from discussions with Robert Davies, Ai Deng, Juan Carlos Escanciano, Chirok Han, Yongmiao Hong, Estate Khmaladze, Jin Lee, Tae-Hwy Lee, Leigh Roberts, Peter Robinson, Peter Thomson, David Vere-Jones and participants at FEMES07, NZESG, NZSA, SRA, UC-Riverside, SEF and SMCS of Victoria University of Wellington, and Joint Economics Conference (Seoul National University, 2010).

to test density forecast optimality, one can test whether the series of probability integral transforms of the forecast errors are IID uniform (U[0,1]).

There is a large number of tests designed to test the IID assumption against specific alternatives, such as structural breaks, serial correlation, or autoregressive conditional heteroskedasticity. Such special purpose tests may lack power in other directions, however, so it is useful to have available broader diagnostics that may alert researchers to otherwise unsuspected properties of their data. Thus, as a complement to special purpose tests, we consider tests for the IID hypothesis that are sensitive to general alternatives. Here we exploit runs statistics to obtain necessary and sufficient conditions for data to be IID. In particular, we show that if the underlying data are IID, then suitably defined runs are IID with the geometric distribution. By testing whether the runs have the requisite geometric distribution, we obtain a new family of tests, the *generalized runs tests*, suitable for testing the IID property. An appealing aspect of our tests is their computational convenience relative to other tests sensitive to general alternatives to IID. For example, Hong and White's (2005) entropy-based IID tests require kernel density estimation, with its associated challenge of bandwidth selection. Our tests do not require kernel estimation and, as we show, have power against dependent alternatives. Our tests also have power against structural break alternatives, without exhibiting the non-monotonicities apparent in certain tests based on kernel estimators (Crainiceanu and Vogelsang, 2007; Deng and Perron, 2008).

Runs have formed an effective means for understanding data properties since the early 1940's. Wald and Wolfowitz (1940), Mood (1940), Dodd (1942), and Goodman (1958) first studied runs to test for randomness of data with a fixed percentile p used in defining the runs. Granger (1963) and Dufour (1981) propose using runs as a nonparametric diagnostic for serial correlation, noting that the choice of p is important for the power of the test. Fama (1965) extensively exploits a runs test to examine stylized facts of asset returns in US industries, with a particular focus on testing for serial correlation of asset returns. Heckman (2001) observes that runs tests can be exploited to detect sample selection bias in cross-sectional data; such biases can be understood to arise from a form of structural break in the underlying distributions.

Earlier runs tests compared the mean or other moments of the runs to those of the geometric distribution for fixed p, say 0.5 (in which case the associated runs can be computed alternatively using the median instead of the mean). Here we develop runs tests based on the probability generating function (PGF) of the geometric distribution. Previously, Kocherlakota and Kocherlakota (KK, 1986) have used the PGF to devise tests for discrete random variables having a given distribution under the null hypothesis. Using fixed values of the PGF parameter s, KK develop tests for the Poisson, Pascal-Poisson, bivariate Poisson, or bivariate Neyman type A distributions. More recently, Rueda, Pérez-Abreu, and O'Reilly (1991) study PGF-based tests for the Poisson null hypothesis, constructing test statistics as functionals of stochastic processes indexed by the PGF parameter s. Here we develop PGF-based tests for the geometric distribution with parameter p, applied to the runs for a sample of continuously distributed random variables.

We construct our test statistics as functionals of stochastic processes indexed by both the runs percentile p and the PGF parameter s. By not restricting ourselves to fixed values for p and/or s, we create the opportunity to construct tests with superior power. Further, we obtain weak limits for our statistics in situations where the distribution of the raw data from which the runs are constructed may or may not be known and where there may or may not be estimated parameters. As pointed out by Darling (1955), Sukhatme (1972), Durbin (1973), and Henze (1996), among others, goodness-of-fit (GOF) based statistics such as ours may have limiting distributions affected by parameter estimation. As we show, however, our test statistics have asymptotic null distributions that are not affected by parameter estimation under mild conditions. We also provide straightforward simulation methods to consistently estimate asymptotic critical values for our test statistics.

We analyze the asymptotic local power of our tests, and we conduct Monte Carlo experiments to explore the properties of our tests in settings relevant for economic applications. In studying power, we give particular attention to dependent alternatives and to alternatives containing an unknown number of structural breaks. To analyze the asymptotic local power of our tests against dependent alternatives, we assume a first-order Markov process converging to an IID process in probability at the rate $n^{-1/2}$, where n is the sample size, and we find that our tests have nontrivial local power. We work with first-order Markov processes for conciseness. Our results generalize to higher-order Markov processes, but that analysis is sufficiently involved that we leave it for subsequent work.

Our Monte Carlo experiments corroborate our theoretical results and also show that our tests exhibit useful finite sample behavior. For dependent alternatives, we compare our generalized runs tests to the entropy-based tests of Robinson (1991), Skaug and Tjøstheim (1996), and Hong and White (2005). Our tests perform respectably, showing good level behavior and useful, and in some cases superior, power against dependent alternatives. For structural break alternatives, we compare our generalized runs tests to Feller's (1951) and Kuan and Hornik's (1995) RR test, Brown, Durbin and Evans's (1975) RE-CUSUM test, Sen's (1980) and Ploberger, Krämer and Kontrus's (1989) RE test, Ploberger and Krämer's (1992) OLS-CUSUM test, Andrews's (1993) Sup-W test, Andrews and Ploberger's (1994) Exp-W and Avg-W tests, and Bai's (1996) M-test. These prior tests are all designed to detect a finite number of structural breaks at unknown locations. We find good level behavior for our tests and superior power against multiple breaks. An innovation is that we consider alternatives where the number of breaks grows with sample size. Our new tests perform well against such structural break alternatives, whereas the prior tests do not.

This paper is organized as follows. In Section 2, we introduce our new family of generalized runs statistics and derive their asymptotic null distributions. These involve Gaussian stochastic processes. Section 3 provides methods for consistently estimating critical values for the test statistics of Section 2. This permits us to compute valid asymptotic critical values even when the associated Gaussian processes are transformed by continuous mappings designed to yield particular test statistics of interest. We achieve this using other easily simulated Gaussian processes whose distributions are identical to those of Section 2. Section 4 studies aspects of local power for our tests. Section 5 contains Monte Carlo simulations; this also illustrates use of the simulation methods developed for obtaining the asymptotic critical values in Section 2. Section 6 contains concluding remarks. All mathematical proofs are collected in the Appendix.

Before proceeding, we introduce mathematical notation used throughout. We let $\mathbf{1}_{\{\,\cdot\,\}}$ stand for the indicator function such that $\mathbf{1}_{\{A\}}$ = 1 if the event A is true, and 0 otherwise. $\Rightarrow$ and $\rightarrow$ denote 'converge(s) weakly' and 'converge(s) to', respectively, and $\overset{d}{=}$ denotes equality in distribution. Further, $\|\cdot\|$ and $\|\cdot\|_\infty$ denote the Euclidean and uniform metrics respectively. We let $\mathcal{C}(A)$ and $\mathcal{D}(A)$ be the spaces of continuous and cadlag mappings from a set A to $\mathbb{R}$ respectively, and we endow these spaces with Billingsley's (1968, 1999) or Bickel and Wichura's (1971) metric. We denote the unit interval as $\mathbb{I} := [0, 1]$.

2 Testing the IID Hypothesis

2.1 Maintained Assumptions

We begin by collecting together assumptions maintained throughout and proceed with our discussion based on these. We first specify the data generating process (DGP) and a parameterized function whose behavior is of interest.

A1 (DGP): *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a complete probability space. For* $m \in \mathbb{N}$, $\{\mathbf{X}_t : \Omega \mapsto \mathbb{R}^m, t = 1, 2, ...\}$ *is a stochastic process on* $(\Omega, \mathcal{F}, \mathbb{P})$.

A2 (PARAMETERIZATION): *For* $d \in \mathbb{N}$, *let* $\boldsymbol{\Theta}$ *be a non-empty convex compact subset of* $\mathbb{R}^d$. *Let* $h : \mathbb{R}^m \times \boldsymbol{\Theta} \mapsto \mathbb{R}$ *be a function such that* (i) *for each* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $h(\mathbf{X}_t(\,\cdot\,), \boldsymbol{\theta})$ *is measurable; and* (ii) *for each* $\omega \in \Omega$, $h(\mathbf{X}_t(\omega), \,\cdot\,)$ *is such that for each* $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \boldsymbol{\Theta}$, $|h(\mathbf{X}_t(\omega), \boldsymbol{\theta}) - h(\mathbf{X}_t(\omega), \boldsymbol{\theta}')| \leq M_t(\omega)\|\boldsymbol{\theta} - \boldsymbol{\theta}'\|$, *where* M_t *is measurable and is* $O_{\mathbb{P}}(1)$, *uniformly in* t.

Assumption A2 specifies that $\mathbf{X}_t$ is transformed via h. The Lipschitz condition of A2(ii) is mild and typically holds in applications involving estimation. Our next assumption restricts attention to continuously distributed random variables.

A3 (CONTINUOUS RANDOM VARIABLES): *For given* $\boldsymbol{\theta}_* \in \boldsymbol{\Theta}$, *the random variables* $Y_t := h(\mathbf{X}_t, \boldsymbol{\theta}_*)$ *have continuous cumulative distribution functions* (*CDFs*) $F_t : \mathbb{R} \mapsto \mathbb{I}$, $t = 1, 2, ...$

Our main interest attaches to distinguishing the following hypotheses:

$$\mathbb{H}_0: \{Y_t : t = 1, 2, ...\} \text{ is an IID sequence;} \quad \text{vs.} \quad \mathbb{H}_1: \{Y_t : t = 1, 2, ...\} \text{ is not an IID sequence.}$$

Under $\mathbb{H}_0$, $F_t \equiv F$ (say), $t = 1, 2, ...$. We separately treat the cases in which F is known or unknown. In the latter case, we estimate F using the empirical distribution function.

We also separately consider cases in which $\boldsymbol{\theta}_*$ is known or unknown. In the latter case, we assume $\boldsymbol{\theta}_*$ is consistently estimated by $\hat{\boldsymbol{\theta}}_n$. Formally, we impose

A4 (ESTIMATOR): *There exists a sequence of measurable functions* $\{\hat{\boldsymbol{\theta}}_n : \Omega \mapsto \boldsymbol{\Theta}\}$ *such that* $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) = O_{\mathbb{P}}(1)$.

Thus, the sequence of transformed observations $\{Y_t : t = 1, 2, ...\}$ need not be observable. Instead, it will suffice that these can be estimated, as occurs when regression errors are of interest. In this case, $h(\mathbf{X}_t, \boldsymbol{\theta}_*)$ can be regarded as a representation of regression errors $X_{1t} - E[X_{1t}|X_{2t}, ..., X_{mt}]$, say. Estimated residuals then have the representation $h(\mathbf{X}_t, \hat{\boldsymbol{\theta}}_n)$. We pay particular attention to the effect of parameter estimation on the asymptotic null distribution of our test statistics.

2.2 Generalized Runs (GR) Tests

Our first result justifies popular uses of runs in the literature. For this, we provide a characterization of the runs distribution, new to the best of our knowledge, that can be exploited to yield a variety of runs-based tests consistent against departures from the IID null hypothesis.

We begin by analyzing the case in which $\boldsymbol{\theta}_*$ and F are known. We define runs in the following two steps: first, for each $p \in \mathbb{I}$, we let $T_n(p) := \{t \in \{1, ..., n\} : F(Y_t) < p\}$, $n = 1, 2, ...$. This set contains those indices whose percentiles $F(Y_t)$ are less than the given number p. That is, we first employ the probability integral transform of Rosenblatt (1952). Next, let $M_n(p)$ denote the (random) number of elements of $T_n(p)$, let $t_{n,i}(p)$ denote the ith smallest element of $T_n(p)$, $i = 1, ..., M_n(p)$, and define the *p-runs* $R_{n,i}(p)$ as

$$R_{n,i}(p) := \begin{cases} t_{n,i}(p), & i = 1; \\ t_{n,i}(p) - t_{n,i-1}(p), & i = 2, ..., M_n(p). \end{cases}$$

Thus, a p-run $R_{n,i}(p)$ is a number of observations separating data values whose percentiles are less than the given value p.

This is the conventional definition of runs found in the literature, except that F is assumed known for the moment. Thus, if the population median is known, then the conventional runs given by Wald-Wolfowitz (1940) are identical to ours with $p = 0.5$. The only difference is that we apply the probability integral transform; this enables us to later accommodate the influence of parameter estimation error on the asymptotic distribution. In Section 2.3 we relax the assumption that F is known and examine how this affects the results obtained in this section. Note that $M_n(p)/n = p + o_{\mathbb{P}}(1)$.

Conventional runs are known to embody the IID hypothesis nonparametrically; this feature is exploited in the literature to test for the IID hypothesis. For example, the Wald-Wolfowitz (1940) runs test considers the standardized number of runs, whose distribution differs asymptotically from the standard normal if the data are not IID, giving the test its power.

It is important to note that for a given n and p, n need not be an element of $T_n(p)$. That is, there may be an "incomplete" or "censored" run at the end of the data that arises because $F(Y_n) \geq p$. We omit this censored run from consideration to ensure that all the runs we analyze have an identical distribution.

To see why this is important, consider the first run, $R_{n,1}(p)$, and, for the moment, suppose that we admit censored runs (i.e., we include the last run, even if $F(Y_n) \geq p$). When a run is censored, we denote its length by $k = \emptyset$. When the original data $\{Y_t\}$ are IID, the marginal distribution of $R_{n,1}(p)$ is then

$$\mathbb{P}(R_{n,1}(p) = k) = \begin{cases} (1-p)p^k, & \text{if } k \leq n; \\ p^n, & \text{if } k = \emptyset, \end{cases} .$$

Thus, when censored runs are admitted, the unconditional distribution of $R_{n,1}(p)$ is a mixture distribution. The same is true for runs other than the first, but the mixture distributions differ due to the censoring. On the other hand, the *uncensored* run $R_{n,1}(p)$ is distributed as $\mathbb{G}_p$, the geometric distribution with parameter p. The same is also true for uncensored runs other than the first. That is, $\{R_{n,i}(p), i = 1, 2, ..., M_n(p)\}$ is the set of runs with identical distribution $\mathbb{G}_p$, as every run indexed by $i = 1, 2, \ldots, M_n(p)$ is uncensored. (The censored run, when it exists, is indexed by $i = M_n(p) + 1$. When the first run is censored, $M_n(p) = 0$.) Moreover, as we show, the uncensored runs are independent when $\{Y_t\}$ is IID. Thus, in what follows, we consider only the uncensored runs, as formally defined above. Further, we construct and analyze our statistics in such a way that values of p for which $M_n(p) = 0$ have no adverse impact on our results.

We now formally state our characterization result. For this, we let $K_{n,i}$ stand as a shorthand notation for $K_{n,i}(p, p')$, with $p' \leq p$, satisfying $K_{n,0}(p, p') = 0$, and $\sum_{j=K_{n,i-1}(p,p')+1}^{K_{n,i}(p,p')} R_{n,j}(p) = R_{n,i}(p')$. The desired characterization is as follows.

LEMMA 1: *Suppose Assumptions A1, A2(i), and A3 hold. (a) Then for each* $n = 1, 2, ..., \{Y_t, t = 1, ..., n\}$ *is IID only if the following regularity conditions ($\mathcal{R}$) hold*:

1. *for every* $p \in \mathbb{I}$ such that $M_n(p) > 0$, $\{R_{n,i}(p), i = 1, ..., M_n(p)\}$ *is IID with distribution* $\mathbb{G}_p$, *the geometric distribution with parameter* p*; and*

2. *for every* $p, p' \in \mathbb{I}$ *with* $p' \leq p$ such that $M_n(p') > 0$,

   (i) $R_{n,j}(p)$ *is independent of* $R_{n,i}(p')$ *if* $j \notin \{K_{n,i-1}+1, K_{n,i-1}+2, ..., K_{n,i}\}$*;*

(ii) *otherwise, for* $w = 1, ..., M_n(p')$, $m = 1, ..., w$, *and* $\ell = m, ..., w$,

$$\mathbb{P}(\sum_{j=1+K_{n,i-1}}^{m+K_{n,i-1}} R_{n,j}(p) = \ell, R_{n,i}(p') = w | K_{n,i-1}, K_{n,i})$$
$$= \begin{cases} \binom{\ell-1}{m-1}(1-p)^{\ell-m}(p-p')^m(1-p')^{w-(\ell+1)}p', & \text{if } \ell = m, \cdots, w-1; \\ \binom{\ell-1}{m-1}(1-p)^{\ell-m}(p-p')^{m-1}p', & \text{if } \ell = w, \end{cases}$$

(b) If $\mathcal{R}$ holds, then Y_t is identically distributed and pairwise independent.

Conditions (1) and (2) of Lemma 1(a) enable us to detect violations of IID $\{Y_t\}$ in directions that differ from the conventional parametric approaches. Specifically, by Lemma 1, alternatives to IID $\{Y_t\}$ may manifest as p-runs with the following alternative ($\mathcal{A}$) properties:

$\mathcal{A}(i)$: the p-runs have distribution $\mathbb{G}_q$, $q \neq p$;

$\mathcal{A}(ii)$: the p-runs have non-geometric distribution;

$\mathcal{A}(iii)$: the p-runs have heterogeneous distributions;

$\mathcal{A}(iv)$: the p-runs and p'-runs have dependence between $R_{n,i}(p)$ and $R_{n,j}(p')$ ($i \neq j, p' \leq p$);

$\mathcal{A}(v)$: any combination of $(i) - (iv)$.

Popularly assumed alternatives to IID data can be related to the alternatives in $\mathcal{A}$. For example, stationary autoregressive processes yield runs with geometric distribution, but for a given p, $\{R_{n,i}(p)\}$ has a geometric distribution different from $\mathbb{G}_p$ and may exhibit serial correlation. Thus stationary autoregressive processes exhibit $\mathcal{A}(i)$ or $\mathcal{A}(iv)$. Alternatively, if the original data are independent but heterogeneously distributed, then for some p, $\{R_{n,i}(p)\}$ is non-geometric or has heterogeneous distributions. This case thus belongs to $\mathcal{A}(ii)$ or $\mathcal{A}(iii)$.

To keep our analysis manageable, we focus on detecting $\mathcal{A}(i) - \mathcal{A}(iii)$ by testing the p-runs for distribution $\mathbb{G}_p$. That is, the hypotheses considered here are as follows: $\mathbb{H}_0' : \{R_{n,i}(p),\ i = 1, ..., M_n(p)\}$ is IID with distribution $\mathbb{G}_p$ for each $p \in \mathbb{I}$ such that $M_n(p) > 0$; vs. $\mathbb{H}_1' : \{R_{n,i}(p),\ i = 1, ..., M_n(p)\}$ manifests $\mathcal{A}(i), \mathcal{A}(ii)$, or $\mathcal{A}(iii)$ for some $p \in \mathbb{I}$ such that $M_n(p) > 0$. Stated more primitively, the alternative DGPs aimed at here include serially correlated and/or heterogeneous alternatives. Alternatives that violate $\mathcal{A}(iv)$ without violating $\mathcal{A}(i) - \mathcal{A}(iii)$ will generally not be detectable. Thus, our goal is different from the rank-based white noise test of Hallin, Ingenbleek, and Puri (1985) and the distribution-function based serial independence test of Delgado (1996).

Certainly, it is of interest to devise statistics specifically directed at $\mathcal{A}(iv)$ in order to test $\mathbb{H}_0$ fully against the alternatives of $\mathbb{H}_1$. Such statistics are not as simple to compute and require analysis different than those motivated by $\mathbb{H}_1'$; moreover, the Monte Carlo simulations in Section 5 show that even with attention restricted to $\mathbb{H}_1'$, we obtain well-behaved tests with power against both commonly assumed dependent and heterogeneous alternatives to IID. We thus leave consideration of tests designed specifically to detect $\mathcal{A}(iv)$ to other work.

Lemma 1(b) is a partial converse of Lemma 1(a). It appears possible to extend this to a full converse (establishing $\{Y_t\}$ is IID) using results of Jogdeo (1968), but we leave this aside here for brevity.

There are numerous ways to construct statistics for detecting $\mathcal{A}(i) - \mathcal{A}(iii)$. For example, as for conventional runs statistics, we can compare the first two runs moments with those implied by the geometric distribution. Nevertheless, this approach may fail to detect differences in higher moments. To avoid difficulties of this sort, we exploit a GOF statistic based on the PGF to test the $\mathbb{G}_p$ hypothesis. For this, let $-1 < \underline{s} < 0 < \bar{s} < 1$; for each $s \in \mathbb{S} := [\underline{s}, \bar{s}]$, define

$$G_n(p, s) := \frac{1}{\sqrt{n}} \sum_{i=1}^{M_n(p)} \left(s^{R_{n,i}(p)} - \frac{sp}{\{1 - s(1-p)\}} \right), \tag{1}$$

if $p \in (p_{\min,n},\ 1)$, and $G_n(p, s) := 0$ otherwise, where $p_{\min,n} := \min[F(Y_1), F(Y_2), \ldots, F(Y_n)]$. This is a scaled difference between the p-runs sample PGF and the $\mathbb{G}_p$ PGF.

Two types of GOF statistics are popular in the literature: those exploiting the empirical distribution function (e.g., Darling, 1955; Sukhatme, 1972; Durbin, 1973; and Henze, 1996) and those comparing empirical characteristic or moment generating functions (MGFs) with their sample estimates (e.g., Bierens, 1990; Brett and Pinkse, 1997; Stinchcombe and White 1998; Hong, 1999; and Pinkse, 1998). The statistic in (1) belongs to the latter type, as the PGF for discrete random variables plays the same role as the MGF, as noted by Karlin and Taylor (1975). The PGF is especially convenient because it is a rational polynomial in s, enabling us to easily handle the weak limit of the process G_n. Specifically, the rational polynomial structure permits us to represent this weak limit as an infinite sum of independent Gaussian processes, enabling us to straightforwardly estimate critical values by simulation, as examined in detail in Section 3. GOF tests using (1) are diagnostic, as are standard MGF-based GOF tests; thus, tests based on (1) do not tell us in which direction the null is violated. Also, like standard MGF-based GOF tests, they are not consistent against all departures from the IID hypothesis. Section 4 examines local alternatives to the null; we provide further discussion there.

Our use of G_n builds on work of Kocherlakota and Kocherlakota (KK, 1986), who consider tests for a number of discrete null distributions, based on a comparison of sample and theoretical PGFs for a given finite set of s's. To test their null distributions, KK recommend choosing s's close to zero. Subsequently, Rueda, Pérez-Abreu, and O'Reilly (1991) examined the weak limit of an analog of $G_n(\,p, \cdot\,)$ to test the IID Poisson null hypothesis. Here we show that if $\{R_{n,i}(p)\}$ is a sequence of IID p-runs distributed as $\mathbb{G}_p$ then $G_n(\,p, \cdot\,)$ obeys the functional central limit theorem; test statistics can be constructed accordingly.

Specifically, for each p, $G_n(p, \cdot) \Rightarrow \mathcal{G}(p, \cdot)$, where $\mathcal{G}(p, \cdot)$ is a Gaussian process such that for each $s, s' \in \mathbb{S}$, $E[\mathcal{G}(p, s)] = 0$, and

$$E\left[\mathcal{G}(p, s)\mathcal{G}(p, s')\right] = \frac{ss'p^2(1-s)(1-s')(1-p)}{\{1-s(1-p)\}\{1-s'(1-p)\}\{1-ss'(1-p)\}}. \tag{2}$$

This mainly follows by showing that $\{G_n(p, \cdot) : n = 1, 2, ...\}$ is tight (see Billingsley, 1999); the given covariance structure (2) is derived from $E\left[G_n(p, s)G_n(p, s')\right]$ under the null. Let $f : \mathcal{C}(\mathbb{S}) \mapsto \mathbb{R}$ be a continuous mapping. Then by the continuous mapping theorem, under the null any test statistic $f[G_n(p, \cdot)]$ obeys $f[G_n(p, \cdot)] \Rightarrow f[\mathcal{G}(p, \cdot)]$.

As Granger (1963) and Dufour (1981) emphasize, the power of runs tests may depend critically on the specific choice of p. For example, if the original data set is a sequence of independent normal variables with population mean zero and variance dependent upon index t, then selecting $p = 0.5$ yields no power, as the runs for $p = 0.5$ follow $\mathbb{G}_{0.5}$ despite the heterogeneity. Nevertheless, useful power can be delivered by selecting p different from 0.5. This also suggests that better powered runs tests may be obtained by considering numerous p's at the same time.

To fully exploit G_n, we consider G_n as a random function of both p and s, and not just $G_n(p, \cdot)$ for given p. Specifically, under the null, a functional central limit theorem ensures that

$$G_n \Rightarrow \mathcal{G} \tag{3}$$

on $\mathbb{J} \times \mathbb{S}$, where $\mathbb{J} := [\underline{p}, 1]$ with $\underline{p} > 0$, and $\mathcal{G}$ is a Gaussian process such that for each (p, s) and (p', s') with $p' \leq p$, $E\left[\mathcal{G}(p, s)\right] = 0$, and

$$E\left[\mathcal{G}(p, s)\mathcal{G}(p', s')\right] = \frac{ss'p'^2(1-s)(1-s')(1-p)\{1-s'(1-p)\}}{\{1-s(1-p)\}\{1-s'(1-p')\}^2\{1-ss'(1-p)\}}. \tag{4}$$

When $p = p'$ then the covariance structure is as in (2). Note also that the covariance structure in (4) is symmetric in both s and p, as we specify that $p' \leq p$. Without this latter restriction, the symmetry is easily seen, as the covariance then has the form

$$\frac{ss' \min[p, p']^2(1-s)(1-s')(1 - \max[p, p'])\{1 - s'(1 - \max[p, p'])\}}{\{1 - s(1 - \max[p, p'])\}\{1 - s'(1 - \min[p, p'])\}^2\{1 - ss'(1 - \max[p, p'])\}}.$$

To obtain (3) and (4), we exploit the joint probability distribution of runs associated with different percentiles p and p'. Although our statistic G_n is not devised to test for dependence between $R_{n,j}(p)$ and $R_{n,j}(p')$, verifying eq.(4) nevertheless makes particular use of the dependence structure implied by $\mathcal{A}(iv)$. This structure also makes it straightforward to devise statistics specifically directed at $\mathcal{A}(iv)$; we leave this aside here to maintain a focused presentation.

For each s, $G_n(\cdot, s)$ is cadlag, so the tightness of $\{G_n\}$ must be proved differently from that of $\{G_n(p, \cdot)\}$. Further, although $\mathcal{G}$ is continuous in p, it is not differentiable almost surely. This is because

$R_{n,i}$ is a discrete random function of p. As n increases, the discreteness of G_n disappears, but its limit is not smooth enough to deliver differentiability in p.

The weak convergence given in (3) is proved by applying the convergence criterion of Bickel and Wichura (1971, theorem 3). We verify this by showing that the modulus of continuity based on the fourth-order moment is uniformly bounded on $\mathbb{J} \times \mathbb{S}$. By taking $\underline{p} > 0$, we are not sacrificing much, as $M_n(p)$ decreases as p tends to zero, so that $G_n(p, \cdot)$ converges to zero uniformly on $\mathbb{S}$. For practical purposes, we can thus let $\underline{p}$ be quite small. We examine the behavior of the relevant test statistics in our Monte Carlo experiments of Section 5 by examining what happens when $\underline{p}$ is zero.

As before, the continuous mapping theorem ensures that, given a continuous mapping $f : \mathcal{D}(\mathbb{J} \times \mathbb{S}) \mapsto \mathbb{R}$, under the null the test statistic $f[G_n]$ obeys $f[G_n] \Rightarrow f[\mathcal{G}]$.

Another approach uses the process $G_n(\cdot, s)$ on $\mathbb{J}$. Under the null, we have $G_n(\cdot, s) \Rightarrow \mathcal{G}(\cdot, s)$, where $\mathcal{G}(\cdot, s)$ is a Gaussian process such that for each p and p' in $\mathbb{J}$ with $p' \leq p$, $E[\mathcal{G}(\cdot, s)] = 0$, and

$$E\left[\mathcal{G}(p, s)\mathcal{G}(p', s)\right] = \frac{s^2 p'^2 (1-s)^2 (1-p)\{1 - s(1-p)\}}{\{1 - s(1-p)\}\{1 - s(1-p')\}^2\{1 - s^2(1-p)\}}. \tag{5}$$

Given a continuous mapping $f : \mathcal{D}(\mathbb{J}) \mapsto \mathbb{R}$, under the null we have $f[G_n(\cdot, s)] \Rightarrow f[\mathcal{G}(\cdot, s)]$.

We call tests based on $f[G_n(p, \cdot)]$, $f[G_n(\cdot, s)]$, or $f[G_n]$ *generalized* runs tests (*GR tests*) to emphasize their lack of dependence on specific values of p and/or s. We summarize our discussion as

THEOREM 1: *Given conditions A1, A2(i), A3, and* $\mathbb{H}_0$,

(i) *for each* $p \in \mathbb{I}$, $G_n(p, \cdot) \Rightarrow \mathcal{G}(p, \cdot)$, *and if* $f : \mathcal{C}(\mathbb{S}) \mapsto \mathbb{R}$ *is continuous, then* $f[G_n(p, \cdot)] \Rightarrow f[\mathcal{G}(p, \cdot)]$;

(ii) *for each* $s \in \mathbb{S}$, $G_n(\cdot, s) \Rightarrow \mathcal{G}(\cdot, s)$, *and if* $f : \mathcal{D}(\mathbb{J}) \mapsto \mathbb{R}$ *is continuous, then* $f[G_n(\cdot, s)] \Rightarrow f[\mathcal{G}(\cdot, s)]$;

(iii) $G_n \Rightarrow \mathcal{G}$, *and if* $f : \mathcal{D}(\mathbb{J} \times \mathbb{S}) \mapsto \mathbb{R}$ *is continuous, then* $f[G_n] \Rightarrow f[\mathcal{G}]$.

The proofs of Theorem 1(i, ii, and iii) are given in the Appendix. Although Theorem 1(i and ii) follow as corollaries of Theorem 1(iii), we prove Theorem 1(i and ii) first and use these properties as lemmas in proving Theorem 1(iii). Note that Theorem 1(i) holds even when $p = 0$, because $G_n(0, \cdot) \equiv 0$, and for every s, $\mathcal{G}(0, s) \sim N(0, 0) = 0$. We cannot allow $p = 0$ in Theorem 1(iii), however, because however large n is, there is always some p close to 0 for which the asymptotics break down. This necessitates our consideration of $\mathbb{J}$ instead of $\mathbb{I}$ in (iii).

We remark that we do not specify f in order to allow researchers to form their own statistics based upon their particular interests. There are a number of popular mappings and justifications for these in the literature, especially those motivated by Bayesian interpretations. For example, Davies (1977) considers the mapping that selects the maximum of the random functions generated by nuisance parameters present only under the alternative. The motivation for this is analogous to that for the Kolmogorov (K) goodness-of-fit statistic, namely, to test non-spurious peaks of the random functions. Bierens (1990) also proposes this choice for his consistent conditional moment statistic. Andrews and Ploberger (1994) study this mapping together with others, and propose a mapping that is optimal in a well defined sense. Alternatively, Bierens (1982) and Bierens and Ploberger (1997) consider integrating the associated random functions with respect to the nuisance parameters, similar to the Smirnov (S) statistic. This is motivated by the desire to test for a zero constant mean function of the associated random functions. Below, we examine K- and S-type mappings for our Monte Carlo simulations. A main motivation for this is that the goodness-of-fit aspects of the transformed data tested via the PGF have interpretations parallel to those for the mappings used in Kolmogorov's and Smirnov's goodness-of-fit statistics.

2.3 Empirical Generalized Runs (EGR) Tests

We now consider the case in which $\boldsymbol{\theta}_*$ is known, but the null CDF of Y_t is unknown. This is a common situation when interest attaches to the behavior of raw data. As the null CDF is unknown, G_n cannot be computed. Nevertheless, we can proceed by replacing the unknown F with a suitable estimator. The empirical distribution function is especially convenient here. Specifically, for each $y \in \mathbb{R}$, we define $\widetilde{F}_n(y) := \frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{\{Y_t \leq y\}}$. This estimation requires modifying our prior definition of p-runs as follows: First, for each $p \in \mathbb{I}$, let $\widetilde{T}_n(p) := \{t \in \mathbb{N} : \widetilde{F}_n(Y_t) < p\}$, let $\widetilde{M}_n(p)$ denote the (random) number of elements of $\widetilde{T}_n(p)$, and let $\widetilde{t}_{n,i}(p)$ denote the ith smallest element of $\widetilde{T}_n(p)$, $i = 1, ..., \widetilde{M}_n(p)$. (Note that $\lfloor \widetilde{M}_n(p)/n \rfloor = p$.) We define the *empirical p-runs* as

$$\widetilde{R}_{n,i}(p) := \begin{cases} \widetilde{t}_{n,i}(p), & i = 1; \\ \widetilde{t}_{n,i}(p) - \widetilde{t}_{n,i-1}(p), & i = 2, ..., \widetilde{M}_n(p). \end{cases}$$

For each $s \in \mathbb{S}$, define

$$\widetilde{G}_n(p,s) := \frac{1}{\sqrt{n}} \sum_{i=1}^{\widetilde{M}_n(p)} \left(s^{\widetilde{R}_{n,i}(p)} - \frac{sp}{\{1 - s(1-p)\}} \right) \tag{6}$$

if $p \in (\frac{1}{n},\ 1)$, and $\widetilde{G}_n(p,s) := 0$ otherwise.

The presence of $\widetilde{F}_n$ leads to an asymptotic null distribution for $\widetilde{G}_n$ different from that for G_n. We now examine this in detail. For convenience, for each $p \in \mathbb{I}$, let $\widetilde{q}_n(p) := \inf\{x \in \mathbb{R} : \widetilde{F}_n(x) \geq p\}$, let $\widetilde{p}_n(p) := F(\widetilde{q}_n(p))$, and abbreviate $\widetilde{p}_n(p)$ as $\widetilde{p}_n$. Then (6) can be decomposed into two pieces as $\widetilde{G}_n = W_n + H_n$, where for each (p, s), $W_n(p,s) := n^{-1/2}\sum_{i=1}^{\widetilde{M}_n(p)}(s^{\widetilde{R}_{n,i}(p)} - s\widetilde{p}_n/\{1 - s(1-\widetilde{p}_n)\})$, and $H_n(p,s) := n^{-1/2}\widetilde{M}_n(p)(s\widetilde{p}_n/\{1-s(1-\widetilde{p}_n)\} - sp/\{1-s(1-p)\})$. Our next result relates W_n to the random function G_n, revealing H_n to be the contribution of the CDF estimation error.

LEMMA 2: *Given conditions A1, A2(i), A3, and* $\mathbb{H}_0$,

(i) $\sup_{(p,s)\,\in\,\mathbb{J}\times\mathbb{S}}\ |W_n(p,s)-G_n(p,s)| = o_{\mathbb{P}}(1)$;

(ii) $H_n \Rightarrow \mathcal{H}$, *where* $\mathcal{H}$ *is a Gaussian process on* $\mathbb{J}\times\mathbb{S}$ *such that for each* (p,s) *and* (p',s') *with* $p'\le p$, $E[\mathcal{H}(p,s)]=0$, *and*

$$E\left[\mathcal{H}(p,s)\mathcal{H}(p',s')\right] = \frac{ss'pp'^2(1-s)(1-s')(1-p)}{\{1-s(1-p)\}^2\{1-s'(1-p')\}^2}. \tag{7}$$

(iii) $(W_n, H_n) \Rightarrow (\mathcal{G},\mathcal{H})$, *and for each* (p,s) *and* (p',s'), $E[\mathcal{G}(p,s)\mathcal{H}(p',s')] = -E[\mathcal{H}(p,s)\mathcal{H}(p',s')]$.

Lemma 2 relates the results of Theorem 1 to the unknown distribution function case. As W_n is asymptotically equivalent to G_n (as defined in the known F case), H_n must be the additional component incurred by estimating the empirical distribution function.

To state our result for the asymptotic distribution of $\widetilde{G}_n$, we let $\widetilde{\mathcal{G}}$ be a Gaussian process on $\mathbb{J}\times\mathbb{S}$ such that for each (p,s) and (p',s') with $p'\le p$, $E[\widetilde{\mathcal{G}}(p,s)]=0$, and

$$E[\widetilde{\mathcal{G}}(p,s)\widetilde{\mathcal{G}}(p',s')] = \frac{ss'p'^2(1-s)^2(1-s')^2(1-p)^2}{\{1-s(1-p)\}^2\{1-s'(1-p')\}^2\{1-ss'(1-p)\}}. \tag{8}$$

The analog of Theorem 1 can now be given as follows.

THEOREM 2: *Given conditions A1, A2(i), A3, and* $\mathbb{H}_0$,

(i) *for each* $p\in\mathbb{I}$, $\widetilde{G}_n(p,\cdot)\Rightarrow\widetilde{\mathcal{G}}(p,\cdot)$, *and if* $f:\mathcal{C}(\mathbb{S})\mapsto\mathbb{R}$ *is continuous, then* $f[\widetilde{G}_n(p,\cdot)]\Rightarrow f[\widetilde{\mathcal{G}}(p,\cdot)]$;

(ii) *for each* $s\in\mathbb{S}$, $\widetilde{G}_n(\cdot,s)\Rightarrow\widetilde{\mathcal{G}}(\cdot,s)$, *and if* $f:\mathcal{D}(\mathbb{J})\mapsto\mathbb{R}$ *is continuous, then* $f[\widetilde{G}_n(\cdot,s)]\Rightarrow f[\widetilde{\mathcal{G}}(\cdot,s)]$;

(iii) $\widetilde{G}_n\Rightarrow\widetilde{\mathcal{G}}$, *and if* f: $\mathcal{D}(\mathbb{J}\times\mathbb{S})\mapsto\mathbb{R}$ *is continuous, then* $f[\widetilde{G}_n]\Rightarrow f[\widetilde{\mathcal{G}}]$.

We call tests based on $f[\widetilde{G}_n(p,\cdot)]$, $f[\widetilde{G}_n(\cdot,s)]$, or $f[\widetilde{G}_n]$ *empirical* generalized runs tests (*EGR tests*) to highlight their use of the empirical distribution function. We emphasize that the distributions of the GR and EGR tests differ, as the CDF estimation error survives in the limit, a consequence of the presence of the component H_n.

2.4 EGR Tests with Nuisance Parameter Estimation

Now we consider the consequences of estimating $\boldsymbol{\theta}_*$ by $\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}$ satisfying A4. As noted by Darling (1955), Sukhatme (1972), Durbin (1973), and Henze (1996), estimation can affect the asymptotic null distribution of GOF-based test statistics. Nevertheless, as we now show in detail, this turns out not to be the case here.

We elaborate our notation to handle parameter estimation. Let $\hat{Y}_{n,t} := h(\mathbf{X}_t,\hat{\boldsymbol{\theta}}_{\boldsymbol{n}})$ and let $\hat{F}_n(y) := \frac{1}{n}\sum_t^n \mathbf{1}_{\{\hat{Y}_{n,t}\le y\}}$, so that $\hat{F}_n$ is the empirical CDF of $\hat{Y}_{n,t}$. Note that we replace $\boldsymbol{\theta}_*$ with its estimate $\hat{\boldsymbol{\theta}}_n$ to accommodate the fact that $\boldsymbol{\theta}_*$ is unknown in this case. Thus, $\hat{F}_n$ contains two sorts of estimation errors: that arising from the empirical distribution and the estimation error for $\boldsymbol{\theta}_*$.

Next, we define the associated runs using the estimates $\hat{Y}_{n,t}$ and $\hat{F}_n$. For each p in $\mathbb{I}$, we now let $\hat{T}_n(p) := \{t \in \mathbb{N} : \hat{F}_n(\hat{Y}_{n,t}) < p\}$, let $\hat{M}_n(p)$ denote the (random) number of elements of $\hat{T}_n(p)$, and let $\hat{t}_{n,i}(p)$ denote the ith smallest element of $\hat{T}_n(p)$, $i = 1, ..., \hat{M}_n(p)$. (Note that $\lfloor \hat{M}_n(p)/n \rfloor = p$.) We define the *parametric empirical* p*-runs* as

$$\hat{R}_{n,i}(p) := \begin{cases} \hat{t}_{n,i}(p), & i = 1; \\ \hat{t}_{n,i}(p) - \hat{t}_{n,i-1}(p), & i = 2, ..., \hat{M}_n(p). \end{cases}$$

For each $s \in \mathbb{S}$, define $\hat{G}_n(p, s) := n^{-1/2} \sum_{i=1}^{\hat{M}_n(p)} (s^{\hat{R}_{n,i}(p)} - sp/\{1 - s(1 - p)\})$ if $p \in (\frac{1}{n},\ 1)$, and $\hat{G}_n(p, s) := 0$ otherwise. Note that these definitions are parallel to those previously given. The only difference is that we are using $\{\hat{Y}_{n,t} : t = 1, 2, ..., n\}$ instead of $\{Y_t : t = 1, 2, ..., n\}$.

To see why estimating $\boldsymbol{\theta}_*$ has no asymptotic impact, we begin by decomposing $\hat{G}_n$ as $\hat{G}_n = \ddot{G}_n + \ddot{H}_n$, where, letting $\widetilde{q}_n(p) := \inf\{y \in \mathbb{R} : \widetilde{F}_n(y) \geq p\}$ and $\widetilde{p}_n := F(\widetilde{q}_n(p))$ as above, we define $\ddot{G}_n(p, s) := n^{-1/2} \sum_{j=1}^{\hat{M}_n(p)} (s^{\hat{R}_{n,i}(p)} - s\widetilde{p}_n/\{1 - s(1 - \widetilde{p}_n)\})$, and $\ddot{H}_n(p, s) := n^{-1/2}\hat{M}_n(p)(s\widetilde{p}_n/\{1 - s(1 - \widetilde{p}_n)\} - sp/\{1 - s(1 - p)\})$. Note that this decomposition is also parallel to the previous decomposition, $\widetilde{G}_n = W_n + H_n$. Our next result extends Lemma 2.

LEMMA 3: *Given conditions A1− A4 and* $\mathbb{H}_0$,

(i) $\sup_{(p,s) \in \mathbb{J}\times\mathbb{S}} |\ddot{G}_n(p, s) - G_n(p, s)| = o_{\mathbb{P}}(1)$;

(ii) $\sup_{(p,s) \in \mathbb{I}\times\mathbb{S}} |\ddot{H}_n(p, s) - H_n(p, s)| = o_{\mathbb{P}}(1)$.

Given Lemma 2(i), it becomes evident that $\hat{G}_n = W_n + H_n + o_{\mathbb{P}}(1) = \widetilde{G}_n + o_{\mathbb{P}}(1)$, so the asymptotic distribution of $\hat{G}_n$ coincides with that of $\widetilde{G}_n$, implying that the asymptotic runs distribution is primarily determined by the estimation error associated with the empirical distribution $\hat{F}_n$ and not by the estimation of $\boldsymbol{\theta}_*$.

The intuition behind this result is straightforward. As Darling (1955), Sukhatme (1972), Durbin (1973), and Henze (1996) note, the asymptotic distribution of an empirical process, say $p \mapsto \hat{Z}_n(p) := n^{1/2}\{F(\hat{q}_n(p)) - p\}$, $p \in \mathbb{I}$, where $\hat{q}_n(p) := \inf\{y \in \mathbb{R} : \hat{F}_n(y) \geq p\}$, is affected by parameter estimation error primarily because the empirical process $\hat{Z}_n$ is constructed using the $\hat{Y}_{n,t} := h(\mathbf{X}_t, \hat{\boldsymbol{\theta}}_{\boldsymbol{n}})$ and the differentiable function F. Because h contains not $\boldsymbol{\theta}_*$ but $\hat{\boldsymbol{\theta}}_n$, the parameter estimation error embodied in $\hat{\boldsymbol{\theta}}_n$ is transmitted to the asymptotic distribution of $\hat{Z}_n$ through $\hat{q}_n$ and F. Thus, if we were to define runs as $\ddot{T}_n(p) := \{t \in \mathbb{N} : F(\hat{Y}_{n,t}) < p\}$, then their asymptotic distribution *would* be affected by the parameter estimation error. Instead, however, our runs $\{\hat{R}_{n,i}\}$ are constructed using $\hat{T}_n(p) := \{t \in \mathbb{N} : \hat{F}_n(\hat{Y}_{n,t}) < p\}$, which replaces F with $\hat{F}_n$, a step function. Variation in $\hat{\boldsymbol{\theta}}_n$ is less important in this case, whereas the estimation of F plays the primary role in determining the asymptotic runs distribution. This also implies that when $\hat{\boldsymbol{\theta}}_n$ is estimated and F is known, it may be computationally convenient to construct the runs using $\hat{F}_n$ instead of F.

The analog of Theorems 1 and 2 is:

THEOREM 3: *Given conditions A1– A4 and* $\mathbb{H}_0$,

(i) *for each* $p \in \mathbb{I}$, $\hat{G}_n(p, \cdot) \Rightarrow \widetilde{\mathcal{G}}(p, \cdot)$, *and if* $f : \mathcal{C}(\mathbb{S}) \mapsto \mathbb{R}$ *is continuous, then* $f[\hat{G}_n(p, \cdot)] \Rightarrow f[\widetilde{\mathcal{G}}(p, \cdot)]$;

(ii) *for each* $s \in \mathbb{S}$, $\hat{G}_n(\cdot, s) \Rightarrow \widetilde{\mathcal{G}}(\cdot, s)$, *and if* $f : \mathcal{D}(\mathbb{J}) \mapsto \mathbb{R}$ *is continuous, then* $f[\hat{G}_n(\cdot, s)] \Rightarrow f[\widetilde{\mathcal{G}}(\cdot, s)]$;

(iii) $\hat{G}_n \Rightarrow \widetilde{\mathcal{G}}$, *and if* f: $\mathcal{D}(\mathbb{J} \times \mathbb{S}) \mapsto \mathbb{R}$ *is continuous, then* $f[\hat{G}_n] \Rightarrow f[\widetilde{\mathcal{G}}]$.

We call tests based on $f[\hat{G}_n(p, \cdot)]$, $f[\hat{G}_n(\cdot, s)]$, or $f[\hat{G}_n]$ *parametric* empirical generalized runs tests (*PEGR tests*) to highlight their use of estimated parameters. By Theorem 3, the asymptotic null distribution of $f[\hat{G}_n]$ is identical to that of $f[\widetilde{G}_n]$, which takes $\boldsymbol{\theta}_*$ as known. We remark that $\mathbb{I}$ appears in Lemma 3(ii) and Theorem 3(i) rather than $\mathbb{J}$, as $\ddot{H}_n$ and H_n only involve the empirical distribution and not the distribution of runs. This is parallel to results of Chen and Fan (2006) and Chan, Chen, Chen, Fan, and Peng (2009). They study semiparametric copula-based multivariate dynamic models and show that their pseudo-likelihood ratio statistic has an asymptotic distribution that depends on estimating the empirical distribution but not other nuisance parameters. The asymptotically surviving $\ddot{H}_n$ in Lemma 3 reflects the asymptotic influence of estimating the empirical distribution, whereas estimating the nuisance parameters has no asymptotic impact, as seen in Theorem 3.

3 Simulating Asymptotic Critical Values

Obtaining critical values for test statistics constructed as functions of Gaussian processes can be challenging. Nevertheless, the rational polynomial structure of our statistics permits us to construct representations of $\mathcal{G}$ and $\widetilde{\mathcal{G}}$ as infinite sums of independent Gaussian random functions. Straightforward simulations then deliver the desired critical values. Given that Theorems 1, 2, and 3 do not specify the continuous mapping f, it is of interest to have methods yielding the asymptotic distributions of $\mathcal{G}$ and $\widetilde{\mathcal{G}}$ rather than $f[\mathcal{G}]$ and $f[\widetilde{\mathcal{G}}]$ for a particular mapping f, as the latter distributions are easily obtained from the methods provided here once f is specified.

To represent $\mathcal{G}$ and $\widetilde{\mathcal{G}}$, we use the Karhunen-Loève (K-L) representation (Loève, 1978, ch.11) of a stochastic process. This represents Brownian motion as an infinite sum of sine functions multiplied by independent Gaussian random coefficients. Grenander (1981) describes this representation as a complete orthogonal system (CONS) and provides many examples. For example, Krivyakov, Matynov, and Tyurin (1977) obtain the asymptotic critical values of von Mises's ω^2 statistic in the multi-dimensional case by applying this method. In econometrics, Phillips (1998) has used the K-L representation to obtain asymptotic critical values for testing cointegration. Andrews's (2001) analysis of test statistics for a GARCH(1,1) model with nuisance parameter not identified under the null also exploits a CONS representation. By theorem 2 of Jain and Kallianpur (1970), Gaussian processes with almost surely continuous paths have a CONS representation and can be approximated uniformly. We apply this result to our GR and (P)EGR test statistics; this straightforwardly delivers reliable asymptotic critical values.

3.1 Generalized Runs Tests

A fundamental property of Gaussian processes is that two Gaussian processes have identical distributions if their covariance structures are the same. We use this fact to represent $\mathcal{G}(p, \cdot\,)$, $\mathcal{G}(\,\cdot\,, s)$, and $\mathcal{G}$ as infinite sums of independent Gaussian processes that can be straightforwardly simulated.

To obtain critical values for GR tests, we can use the Gaussian process $\mathcal{Z}^*$ defined by

$$\mathcal{Z}^*(p, s) := \frac{sp(1-s)\mathcal{B}_0^0(p)}{\{1-s(1-p)\}^2} + \frac{(1-s)^2}{\{1-s(1-p)\}^2}\sum_{j=1}^{\infty} s^j \mathcal{B}_j^s\left(p^2, (1-p)^{1+j}\right), \tag{9}$$

where $\mathcal{B}_0^0$ is a Brownian bridge, and $\{\mathcal{B}_j^s : j = 1, 2, ...\}$ is a sequence of independent Brownian sheets, whose covariance structure is given by $E[\mathcal{B}_j^s(p, q)\mathcal{B}_i^s(p', q')] = \mathbf{1}_{\{i=j\}} \min[p, p'] \cdot \min[q, q']$. The arguments of $\mathcal{B}_j^s$ lie only in the unit interval, and it is readily verified that $E[\mathcal{Z}^*(p, s)\mathcal{Z}^*(p', s')]$ is identical to (4), so $\mathcal{Z}$ has the same distribution as $\mathcal{G}$.

An inconvenient computational aspect of $\mathcal{Z}^*$ is that the terms $\mathcal{B}_j^s$ require evaluation on a two dimensional square, which is computationally demanding. More convenient in this regard is the Gaussian process $\mathcal{Z}$ defined by

$$\mathcal{Z}(p, s) := \frac{sp(1-s)\mathcal{B}_0^0(p)}{\{1-s(1-p)\}^2} + \frac{(1-s)^2}{\{1-s(1-p)\}^2}\sum_{j=1}^{\infty} s^j (1-p)^{1+j}\mathcal{B}_j\left(\frac{p^2}{(1-p)^{1+j}}\right), \tag{10}$$

where $\{\mathcal{B}_j : j = 1, 2, ...\}$ is a sequence of independent standard Brownian motions independent of the Brownian bridge $\mathcal{B}_0^0$. It is straightforward to compute $E[\mathcal{Z}(p, s)\mathcal{Z}(p', s')]$. Specifically, if $p' \leq p$ then

$$E[\mathcal{Z}(p, s)\mathcal{Z}(p', s')] = \frac{ss'p'^2(1-s)(1-s')(1-p)\{1-s'(1-p)\}}{\{1-s(1-p)\}\{1-s'(1-p')\}^2\{1-ss'(1-p)\}}.$$

This covariance structure is also identical to (4), so $\mathcal{Z}$ has the same distribution as $\mathcal{G}$. The processes $\mathcal{B}_0^0$ and $\{\mathcal{B}_j\}$ are readily simulated as a consequence of Donsker's (1951) theorem or the K-L representation (Loève, 1978, ch.11), ensuring that critical values for any statistic $f[G_n]$ can be straightforwardly found by Monte Carlo methods.

Although one can obtain asymptotic critical values for p-runs test statistics $f[G_n(p,\cdot)]$ by fixing p in (9) or (10), there is a much simpler representation for $\mathcal{G}(p,\cdot)$. Specifically, consider the process $\mathcal{Z}_p$ defined by $\mathcal{Z}_p(s) := \frac{sp(1-s)(1-p)^{1/2}}{\{1-s(1-p)\}} \sum_{j=0}^{\infty} s^j(1-p)^{j/2} Z_j$, where $\{Z_j\}$ is a sequence of IID standard normals. It is readily verified that for each p, $E[\mathcal{Z}_p(s)\mathcal{Z}_p(s')]$ is identical to (2). Because $\mathcal{Z}_p$ does not involve the Brownian bridge, Brownian motions, or Brownian sheets, it is more efficient to simulate than $\mathcal{Z}(p,\cdot)$. This convenient representation arises from the symmetry of equation (4) in s and s' when $p = p'$. The fact that equation (4) is asymmetric in p and p' when $s = s'$ implies that a similar convenient representation for $\mathcal{G}(\cdot, s)$ is not available. Instead, we obtain asymptotic critical values for test statistics $f[G_n(\cdot, s)]$, by fixing s in (9) or (10).

We summarize these results as follows.

THEOREM 4: (i) *For each* $p \in \mathbb{I}$, $\mathcal{G}(p,\cdot) \overset{d}{=} \mathcal{Z}_p$, *and if* f: $\mathcal{C}(\mathbb{S}) \mapsto \mathbb{R}$ *is continuous, then* $f[\mathcal{G}(p,\cdot)] \overset{d}{=} f[\mathcal{Z}_p]$;

(ii) $\mathcal{G} \overset{d}{=} \mathcal{Z}^* \overset{d}{=} \mathcal{Z}$, *and if* f: $\mathcal{D}(\mathbb{J}\times\mathbb{S}) \mapsto \mathbb{R}$ *is continuous, then* $f[\mathcal{G}] \overset{d}{=} f[\mathcal{Z}^*] \overset{d}{=} f[\mathcal{Z}]$.

As deriving the covariance structures of the relevant processes is straightforward, we omit the proof of Theorem 4 from the Appendix.

3.2 (P)EGR Tests

For the EGR statistics, we can similarly provide a Gaussian process whose covariance structure is the same as (8) and that can be straightforwardly simulated. By Theorem 3, this Gaussian process also yields critical values for PEGR test statistics.

We begin with a representation for $\mathcal{H}$. Specifically, consider the Gaussian process $\mathcal{X}$ defined by $\mathcal{X}(p,s) := -\frac{sp(1-s)}{\{1-s(1-p)\}^2}\mathcal{B}_0^0(p)$, where $\mathcal{B}_0^0$ is a Brownian bridge as before. It is straightforward to show that when $p' \leq p$, $E[\mathcal{X}(p,s)\mathcal{X}(p',s')]$ is the same as (7), implying that this captures the asymptotic distribution of the empirical distribution estimation error H_n, which survives to the limit. The representation $\mathcal{Z}$ for $\mathcal{G}$ in Theorem 4(ii) and the covariance structure for $\mathcal{G}$ and $\mathcal{H}$ required by Lemma 2(iii) together suggest representing $\widetilde{\mathcal{G}}$ as $\widetilde{\mathcal{Z}}^*$ or $\widetilde{\mathcal{Z}}$ defined by

$$\widetilde{\mathcal{Z}}^*(p,s) := \frac{(1-s)^2}{\{1-s(1-p)\}^2}\sum_{j=1}^{\infty} s^j \mathcal{B}_j^s\left(p^2,(1-p)^{1+j}\right) \tag{11}$$

and

$$\widetilde{\mathcal{Z}}(p,s) := \frac{(1-s)^2}{\{1-s(1-p)\}^2}\sum_{j=1}^{\infty} s^j(1-p)^{1+j}\mathcal{B}_j\left(\frac{p^2}{(1-p)^{1+j}}\right) \tag{12}$$

respectively, so that $\widetilde{\mathcal{Z}}^*$ (resp. $\widetilde{\mathcal{Z}}$) is the sum of $\mathcal{Z}^*$ (resp. $\mathcal{Z}$) and $\mathcal{X}$ with the identical $\mathcal{B}_0^0$ in each. As is readily verified, (8) is the same as $E[\widetilde{\mathcal{Z}}^*(p',s')\widetilde{\mathcal{Z}}^*(p,s)]$ and $E[\widetilde{\mathcal{Z}}(p',s')\widetilde{\mathcal{Z}}(p,s)]$. Thus, simulating (11) or (12) can deliver the asymptotic null distribution of $\widetilde{G}_n$ and $\hat{G}_n$.

Similar to the previous case, the following representation is convenient when p is fixed:

$$\widetilde{\mathcal{Z}}_p(s) := \frac{sp(1-s)(1-p)^{1/2}}{\{1-s(1-p)\}} \sum_{j=0}^{\infty} \left\{ \frac{s^j(1-s) - p(1-s^{j+1})}{\{1-s(1-p)\}} \right\} (1-p)^{j/2} Z_j.$$

For fixed s, we use the representation provided by $\widetilde{\mathcal{Z}}(\cdot, s)$ or $\widetilde{\mathcal{Z}}^*(\cdot, s)$.

We summarize these results as follows.

THEOREM 5 (i) $\mathcal{H} \stackrel{d}{=} \mathcal{X}$;

(ii) *For each* $p \in \mathbb{I}$, $\widetilde{\mathcal{G}}(p, \cdot\,) \stackrel{d}{=} \widetilde{\mathcal{Z}}_p$, *and if* $f : \mathcal{C}(\mathbb{S}) \mapsto \mathbb{R}$ *is continuous, then* $f[\widetilde{\mathcal{G}}(p, \cdot\,)] \stackrel{d}{=} f[\widetilde{\mathcal{Z}}_p]$;

(iii) $\widetilde{\mathcal{G}} \stackrel{d}{=} \widetilde{\mathcal{Z}}^* \stackrel{d}{=} \widetilde{\mathcal{Z}}$, *and if* $f : \mathcal{D}(\mathbb{J} \times \mathbb{S}) \mapsto \mathbb{R}$ *is continuous, then* $f[\widetilde{\mathcal{G}}] \stackrel{d}{=} f[\widetilde{\mathcal{Z}}^*] \stackrel{d}{=} f[\widetilde{\mathcal{Z}}]$.

As deriving the covariance structures of the relevant processes is straightforward, we omit the proof of Theorem 5 from the Appendix.

4 Asymptotic Local Power

Generalized runs tests target serially correlated autoregressive processes and/or independent heterogeneous processes violating $\mathcal{A}(i) - \mathcal{A}(iii)$, as stated in Section 3. Nevertheless, runs tests are not always consistent against these processes, because just as for MGF-based GOF tests, PGF-based GOF tests cannot handle certain measure zero alternatives. We therefore examine whether the given (P)EGR test statistics have nontrivial power under specific local alternatives. To study this, we consider a first-order Markov process under which (P)EGR test statistics have nontrivial power when the convergence rate of the local alternative to the null is $n^{-1/2}$. Another motivation for considering this local alternative is to show that (P)EGR test statistics can have local power directly comparable to that of standard parametric methods. We consider first-order Markov processes for conciseness. The test can also be shown to have local power against higher-order Markov processes. Our results for first-order processes provide heuristic support for this claim, as higher-order Markov processes will generally exhibit first order dependence. A test capable of detecting true first-order Markov structure will generally be able to detect apparent first-order structure, as well. The situation is analogous to the case of autoregression, where tests for AR(1) structure are generally also sensitive to AR(p) structures, $p > 1$. We provide some additional discussion below in the simulation section.

To keep our presentation succinct, we focus on EGR test statistics in this section. We saw above that the distribution theory for EGR statistics applies to PEGR statistics. This also holds for local power analysis. For brevity, we omit a formal demonstration of this fact here.

We consider a double array of processes $\{Y_{n,t}\}$, and we let $\mathcal{F}_{n,t}$ denote the smallest σ-algebra generated by $\{Y_{n,t}, Y_{n,t-1}, ..., \}$. We suppose that for each n, $\{Y_{n,1}, Y_{n,2}, ..., Y_{n,n}\}$ is a strictly stationary and geometric ergodic first-order Markov process having transition probability distributions $\mathbb{P}(Y_{n,t+1} \leq y|\mathcal{F}_{n,t})$ with the following Lebesgue-Stieltjes differential:

$$\mathbb{H}_1^\ell : dF_n(y|\mathcal{F}_{n,t}) = dF(y) + n^{-1/2} dD(y, Y_{n,t}) \tag{13}$$

under the local alternative, where we construct the remainder term to be $o_{\mathbb{P}}(n^{-1/2})$ uniformly in y. For this, we suppose that $D(\cdot, Y_{n,t})$ is a signed measure with properties specified in A5, and that for a suitable signed measure Q with Lebesgue-Stieltjes differential dQ, $Y_{n,t}$ has marginal Lebesgue-Stieltjes differential

$$dF_n(y) = dF(y) + n^{-1/2}\{dQ(y) + o(1)\}. \tag{14}$$

We impose the following formal condition.

A5 (LOCAL ALTERNATIVE): (i) *For each* $n = 1, 2, ...$, $\{Y_{n,1}, Y_{n,2}, ..., Y_{n,n}\}$ *is a strictly stationary and geometric ergodic first-order Markov process with transition probability distributions given by eq.* (13) *and marginal distributions given by eq.* (14), *where* (ii) $D : \mathbb{R} \times \mathbb{R} \mapsto \mathbb{R}$ *is a continuous function such that* $D(\cdot, z)$ *defines a signed measure for each* $z \in \mathbb{R}$; (iii) $\sup_x |D(x, Y_{n,t})| \leq M_{n,t}$ *such that* $E[M_{n,t}] \leq \Delta < \infty$ *uniformly in* t *and* n, *and* $\lim_{y\to\pm\infty} D(y, Y_{n,t}) = 0$ *a.s.*$-\mathbb{P}$ *uniformly in* t *and* n; (iv) $\sup_y \int_{-\infty}^{\infty} |D(y, x)|dF(x) \leq \Delta$ *and* $\sup_y |\int_y^{\infty} D(y, x)dD(x, Y_{n,t})| \leq M_{n,t}$ *for all* t *and* n.

Thus, as n tends to infinity, $\{Y_{n,1}, Y_{n,2}, ...\}$ converges in distribution to an IID sequence of random variables with marginal distribution F. Note that the marginal distribution given in eq. (14) is obtained by substituting the conditional distribution of $Y_{n,t-j+1}|\mathcal{F}_{n,t-j}$ $(j = 1, 2, ...)$ into (13) and integrating with respect to the random variables other than $Y_{n,t}$. For example, $\int_{-\infty}^{\infty} D(y, z)dF(z) = Q(y)$. This implies that the properties of Q are determined by those of D. For example, $\lim_{y\to\infty} Q(y) = 0$ and $\sup_y |Q(y)| \leq \Delta$.

Our motivations for condition A5 are as follows. We impose the first-order Markov condition for conciseness. Higher-order Markov processes can be handled similarly. Assumption A5(i) also implies that $\{Y_{n,t}\}$ is an ergodic β–mixing process by theorem 1 of Davydov (1973). Next, assumptions A5(ii, iii) ensure that $F_n(\,\cdot\,|\mathcal{F}_{n,t})$ is a proper distribution for all n almost surely, corresponding to A3. Finally, assumptions A5(iii, iv) asymptotically control certain remainder terms of probabilities relevant to runs. Specifically, applying an induction argument yields that for each $k = 1, 2, ...$,

$$\mathbb{P}(Y_{n,t+1} \geq y, ..., Y_{n,t+k-1} \geq y, Y_{n,t+k} < y|\mathcal{F}_{n,t}) = p(1-p)^{k-1} + \frac{1}{\sqrt{n}} h_k(p, Y_{n,t}) + r_k(p, Y_{n,t}), \tag{15}$$

where p is a short-hand notation for $F(y)$; for each p, $h_1(p, Y_{n,t}) := C(p, Y_{n,t}) := D(F^{-1}(p), Y_{n,t})$; $h_2(p, Y_{n,t}) := w(p) - p\,C(p, Y_{n,t})$; and for $k = 3, 4, ...$, $h_k(p, Y_{n,t}) := w(p)(1-p)^{k-3}(1-(k-$

$1)p) - p(1-p)^{k-2}C(p, Y_{n,t})$, where $w(p) := \alpha(F^{-1}(p)) := \int_y^\infty D(y,x)dF(x)$. Here, the remainder term $r_k(p, Y_{n,t})$ is sequentially computed using previous remainder terms and $h_k(p, Y_{n,t})$. For example, for given p, $r_1(p, Y_{n,t}) = 0$, $r_2(p, Y_{n,t}) := n^{-1}\int_{F^{-1}(p)}^\infty D(F^{-1}(p), x)dD(x, Y_{n,t})$, and so forth. These remainder terms turn out to be $O_{\mathbb{P}}(n^{-1})$, mainly due to assumptions A5(iii, iv).

Runs distributions can also be derived from (15), with asymptotic behavior controlled by assumptions A5(iii, iv). That is, if $Y_{n,t} < y$, then the distribution of a run starting from $Y_{n,t+1}$, say $R_{n,i}(p)$, can be obtained from (15) as

$$\begin{aligned}\mathbb{P}(R_{n,i}(p) = k) &= \mathbb{P}(Y_{n,t+1} \geq y, Y_{n,t+2} \geq y, ..., Y_{n,t+k} < y | Y_{n,t} < y)\\ &= p(1-p)^{k-1} + n^{-1/2}F_n(F^{-1}(p))^{-1}h_{n,k}(p) + F_n(F^{-1}(p))^{-1}r_{n,k}(p),\end{aligned} \tag{16}$$

where, as n tends to infinity, for each k, $h_{n,k}(p) := \int_{-\infty}^{F^{-1}(p)} h_k(p,x)dF_n(x) \to h_k(p) := \int_{-\infty}^{F^{-1}(p)} h_k(p,x)$ $dF(x)$ and $r_{n,k}(p) := \int_{-\infty}^{F^{-1}(p)} r_k(p,x)dF_n(x) \to r_k(p) := \int_{-\infty}^{F^{-1}(p)} r_k(p,x)dF(x)$; and for each p, $F_n(F^{-1}(p)) \to p$ from assumptions A5(iii, iv). Further, the remainder term $r_k(p)$ is $O_{\mathbb{P}}(n^{-1})$, uniformly in p.

The local power of EGR test statistics stems from the difference between the distribution of runs given in eq. (16) and that obtained under the null. Specifically, the second component on the right-hand side (RHS) of (16) makes the population mean of G_n different from zero, so that the limiting distribution of G_n corresponding to that obtained under the null can be derived when its population mean is appropriately adjusted. This non-zero population mean yields local power for $n^{-1/2}$ local alternatives for the EGR test statistics as follows.

THEOREM 6: *Given conditions A1, A2(i), A3, A5, and* $\mathbb{H}_1^\ell$, $\widetilde{G}_n - \mu \Rightarrow \widetilde{\mathcal{G}}$, *where for each* $(p,s) \in \mathbb{J} \times \mathbb{S}$, $\mu(p,s) := ps(1-s)\{sw(p) - Q(F^{-1}(p))\}/\{1-s(1-p)\}^2 + \frac{s(1-s)}{\{1-s(1-p)\}}\int_{-\infty}^{F^{-1}(p)} C(p,z)dF(z)$.

It is not difficult to specify DGPs satisfying the condition A5. For example, an AR(1) process can be constructed so as to belong to this case. That is, if for each t, $Y_{n,t} := n^{-1/2}Y_{n,t-1} + \varepsilon_t$ and $\varepsilon_t \sim$ IID $N(0,1)$, then we can let $C(p, Y_{n,t}) = -\xi(p)Y_{n,t} + o_{\mathbb{P}}(1)$ and $w(p) = -\xi(p)^2$, where $\xi(p) := \phi[\Phi^{-1}(p)]$, and $\phi(\,\cdot\,)$ and $\Phi(\,\cdot\,)$ are the probability density function (PDF) and CDF of a standard normal random variable. This gives $\mu(p,s) = \{\xi(p)^2 s(1-s)^2\}/\{1-s(1-p)\}^2$, with $Q \equiv 0$. Because we have convergence rate $n^{-1/2}$, the associated EGR test statistics have the same convergence rate as the parametric local alternative.

We point out several implications of Theorem 6. First, if the convergence rate in (13) is lower than $1/2$, the EGR test may not have useful power; EGR tests are not powerful against every alternative to $\mathbb{H}_0'$. For EGR tests to be consistent against first-order Markov processes, the rate must be at least $1/2$. Second, the statement for first-order Markov process can be extended to further higher-order Markov processes, although we do not pursue this here for brevity. Theorem 6 therefore should be understood as a starting point for identifying Markov processes as a class of $n^{-1/2}$-alternatives. Finally, the result of Theorem 6 does not hold for every local alternative specification. Our examination of a variety of other local alternative specifications reveals cases in which EGR tests have nontrivial power at the rate $n^{-1/4}$. For example, certain independent and non-identically distributed (INID) DGPs can yield EGR test statistics exhibiting $n^{-1/4}$ rates. This rate arises because analysis of these cases requires an expansion of the conditional distribution of runs of order higher than that considered in Theorem 6. For brevity, we do not examine this further here.

5 Monte Carlo Simulations

In this section, we use Monte Carlo simulation to obtain critical values for test statistics constructed with f delivering the L_1 (S-type) and uniform (K-type) norms of its argument. We also examine level and power properties of tests based on these critical values.

5.1 Critical Values

We consider the following statistics: $\mathcal{T}^p_{1,n}(\mathbb{S}_1) := \int_{\mathbb{S}_1} |G_n(p,s)|ds$, $\mathcal{T}^p_{\infty,n}(\mathbb{S}_1) := \sup_{s\in\mathbb{S}_1} |G_n(p,s)|$, $\widetilde{\mathcal{T}}^p_{1,n}(\mathbb{S}_1) := \int_{\mathbb{S}_1} |\widetilde{G}_n(p,s)|ds$, $\widetilde{\mathcal{T}}^p_{\infty,n}(\mathbb{S}_1) := \sup_{s\in\mathbb{S}_1} |\widetilde{G}_n(p,s)|$, where $\mathbb{S}_1 := [-0.99, 0.99]$, and $p \in \{0.1, 0.3, 0.5, 0.7,\ 0.9\}$; $\mathcal{T}^s_{1,n} := \int_{\mathbb{I}} |G_n(p,s)|dp$, $\mathcal{T}^s_{\infty,n} := \sup_{p\in\mathbb{I}} |G_n(p,s)|$, $\widetilde{\mathcal{T}}^s_{1,n} := \int_{\mathbb{I}} |\widetilde{G}_n(p,s)|dp$, $\widetilde{\mathcal{T}}^s_{\infty,n} := \sup_{p\in\mathbb{I}} |\widetilde{G}_n(p,s)|$, where $s \in \{-0.5, -0.3, -0.1, 0.1, 0.3, 0.5\}$; and $\mathcal{T}_{1,n}(\mathbb{S}) := \int_{\mathbb{I}}\int_{\mathbb{S}} |G_n(p,s)|\, dsdp$, $\mathcal{T}_{\infty,n}(\mathbb{S}) := \sup_{(p,s)\in\mathbb{I}\times\mathbb{S}} |G_n(p,s)|$, $\widetilde{\mathcal{T}}_{1,n}(\mathbb{S}) := \int_{\mathbb{I}}\int_{\mathbb{S}} |\widetilde{G}_n(p,s)|\, dsdp$, $\widetilde{\mathcal{T}}_{\infty,n}(\mathbb{S}) := \sup_{(p,s)\in\mathbb{I}\times\mathbb{S}} |\widetilde{G}_n(p,s)|$, where we consider $\mathbb{S}_1 := [-0.99, 0.99]$ and $\mathbb{S}_2 := [-0.50, 0.50]$ for $\mathbb{S}$. As discussed above, these S- and K-type statistics are relevant for researchers interested in testing for non-zero constant mean function and non-spurious peaks of G_n on $\mathbb{I}\times\mathbb{S}$ in terms of $\mathcal{T}_{1,n}(\mathbb{S})$ and $\mathcal{T}_{\infty,n}(\mathbb{S})$ respectively.

Note that these test statistics are constructed using $\mathbb{I}$ instead of $\mathbb{J}$. There are two reasons for doing this. First, we want to examine the sensitivity of these test statistics to $\underline{p}$. We have chosen the extreme case to examine the levels of the test statistics. Second, as pointed out by Granger (1963) and Dufour (1981), more alternatives can be handled by specifying a larger space for p.

Theorems 4 and 5 ensure that the asymptotic null distributions of these statistics can be generated by simulating $\mathcal{Z}_p$, $\mathcal{Z}$ (or $\mathcal{Z}^*$), $\widetilde{\mathcal{Z}}_p$, and $\widetilde{\mathcal{Z}}$ (or $\widetilde{\mathcal{Z}}^*$), as suitably transformed. We approximate these using

$$\mathcal{W}_p(s) := \frac{sp(1-s)(1-p)^{1/2}}{\{1-s(1-p)\}}\sum_{j=0}^{50} s^j(1-p)^{j/2}Z_j,$$

$$\mathcal{W}(p,s) := \frac{sp(1-s)}{\{1-s(1-p)\}^2}\widetilde{\mathcal{B}}_0^0(p) + \frac{(1-s)^2}{\{1-s(1-p)\}^2}\sum_{j=1}^{40} s^j(1-p)^{1+j}\widetilde{\mathcal{B}}_j\left(\frac{p^2}{(1-p)^{1+j}}\right),$$

$$\widetilde{\mathcal{W}}_p(s) := \frac{sp(1-s)(1-p)^{1/2}}{\{1-s(1-p)\}}\sum_{j=0}^{50}\left\{s^j - \frac{p}{\{1-s(1-p)\}}\right\}(1-p)^{j/2}Z_j, \quad \text{and}$$

$$\widetilde{\mathcal{W}}(p,s) := \frac{(1-s)^2}{\{1-s(1-p)\}^2}\sum_{j=1}^{40} s^j(1-p)^{1+j}\widetilde{\mathcal{B}}_j\left(\frac{p^2}{(1-p)^{1+j}}\right),$$

respectively, where $\widetilde{\mathcal{B}}_0^0(p) := W_0(p) - pW_0(1)$, $\widetilde{\mathcal{B}}_j(x+p) := W_{x+1}(p) + \sum_{k=1}^{x} W_k(1)$ (with $x \in \mathbb{N}$, and $p \in \mathbb{I}$), and $\{W_k : k = 0,1,2,...\}$ is a set of independent processes approximating Brownian motion using the K-L representation, defined as $W_k(p) := \sqrt{2}\sum_{\ell=1}^{100}\{\sin[(\ell-1/2)\pi p]\}Z_\ell^{(k)}/\{(\ell-1/2)\pi\}$, where $Z_\ell^{(j)} \sim$ IID $N(0,1)$ with respect to ℓ and j. We evaluate these functions for $\mathbb{I}$, $\mathbb{S}_1$, and $\mathbb{S}_2$ on the grids $\{0.01, 0.02, ..., 1.00\}$, $\{-0.99, -0.98, ..., 0.98, 0.99\}$, and $\{-0.50, -0.49, ..., 0.49, 0.50\}$, respectively.

Concerning these approximations, several comments are in order. First, the domains for p and s are approximated using a relatively fine grid. Second, we truncate the sum of the independent Brownian motions at 40 terms. The jth term contributes a random component with a standard deviation of $s^j p(1-p)^{(1+j)/2}$, which vanishes quickly as j tends to infinity. Third, we approximate $\widetilde{\mathcal{B}}_j$ on the positive Euclidean line by the Brownian motion on $[0, 10, 000]$. Preliminary experiments showed the impact of these approximations to be small when $\mathbb{S}$ is appropriately chosen; we briefly discuss certain aspects of these experiments below.

Table I contains the critical values generated by 10,000 replications of these processes.

5.2 Level and Power of the Test Statistics

In this section, we compare the level and power of generalized runs tests with other tests in the literature. We conduct two sets of experiments. The first examines power against dependent alternatives. The second examines power against structural break alternatives.

To examine power against dependent alternatives, we follow Hong and White (2005) and consider the following DGPs:

- DGP 1.1: $X_t := \varepsilon_t$;
- DGP 1.2: $X_t := 0.3X_{t-1} + \varepsilon_t$;
- DGP 1.3: $X_t := h_t^{1/2}\varepsilon_t$, and $h_t = 1 + 0.8X_{t-1}^2$;
- DGP 1.4: $X_t := h_t^{1/2}\varepsilon_t$, and $h_t = 0.25 + 0.6h_{t-1} + 0.5X_{t-1}^2\mathbf{1}_{\{\varepsilon_{t-1}<0\}} + 0.2X_{t-1}^2\mathbf{1}_{\{\varepsilon_{t-1}\geq 0\}}$;
- DGP 1.5: $X_t := 0.8X_{t-1}\varepsilon_{t-1} + \varepsilon_t$;
- DGP 1.6: $X_t := 0.8\varepsilon_{t-1}^2 + \varepsilon_t$;
- DGP 1.7: $X_t := 0.4X_{t-1}\mathbf{1}_{\{X_{t-1}>1\}} - 0.5X_{t-1}\mathbf{1}_{\{X_{t-1}\leq 1\}} + \varepsilon_t$;
- DGP 1.8: $X_t := 0.8|X_{t-1}|^{0.5} + \varepsilon_t$;

- DGP 1.9: $X_t := \text{sgn}(X_{t-1}) + 0.43\varepsilon_t$;

where $\varepsilon_t \sim$ IID $N(0,1)$. Note that DGP 1.1 satisfies the null hypothesis, whereas the other DGPs represent interesting dependent alternatives. As there is no parameter estimation, we apply our EGR statistics and compare these to the entropy-based nonparametric statistics of Robinson (1991), Skaug and Tjøstheim (1996), and Hong and White (2005), denoted as R_n, ST_n, and HW_n, respectively.

We present the results in Tables II to IV. To summarize, the EGR test statistics generally show approximately correct levels, even using $\mathbb{I}$ instead of $\mathbb{J}$. We notice, however, that $\widetilde{\mathcal{T}}^s_{1,n}$ exhibits level distortion when s gets close to one. This is mainly because the number of Brownian motions in the approximation is finite, and these are defined on the finite Euclidean positive real line, [0, 10,000]. If s and p are close to one and zero respectively, then the approximation can be coarse. Specifically, the given finite number of Brownian motions may not enough to adequately approximate the desired infinite sum of Brownian motions, and the given finite domain [0, 10,000] may be too small to adequately approximate the positive Euclidean real line. For the other tests, we do not observe similar level distortions.

For the DGPs generating alternatives to $\mathbb{H}_0$, the EGR tests generally gain power as n increases. As noted by Granger (1963) and Dufour (1981), a particular selection of p or, more generally, the choice of mapping f can yield tests with better or worse power. Generally, we see that the $\widetilde{\mathcal{T}}^p_{1,n}(\mathbb{S}_1)$ (resp. $\widetilde{\mathcal{T}}^s_{1,n}$)-based tests outperform the $\widetilde{\mathcal{T}}^p_{\infty,n}(\mathbb{S}_1)$ (resp. $\widetilde{\mathcal{T}}^s_{\infty,n}$)-based tests. Similarly, the $\widetilde{\mathcal{T}}_{1,n}(\mathbb{S})$-based tests outperform the $\widetilde{\mathcal{T}}_{\infty,n}(\mathbb{S})$-based tests for both $\mathbb{S}_1$ and $\mathbb{S}_2$. Among the $\widetilde{\mathcal{T}}^p_{1,n}(\mathbb{S}_1)$-based tests, more extreme choices for p often yield better power. Also, in general, the power performances of the $\widetilde{\mathcal{T}}_{1,n}(\mathbb{S}_2)$-based tests are midway between those of the best and worst cases for the $\widetilde{\mathcal{T}}^s_{1,n}$-based tests. Apart from these observations, there is no clear-cut relation between the $\widetilde{\mathcal{T}}^p_{1,n}(\mathbb{S}_1)$-based tests and the $\widetilde{\mathcal{T}}_{1,n}(\mathbb{S}_1)$-based tests. The more powerful $\widetilde{\mathcal{T}}^p_{1,n}(\mathbb{S}_1)$-based tests dominate the $\widetilde{\mathcal{T}}_{1,n}(\mathbb{S}_1)$-based tests for DGPs 1.3-1.5, but the $\widetilde{\mathcal{T}}_{1,n}(\mathbb{S}_1)$-based tests dominate for DGPs 1.2, and 1.6-1.9.

Comparing EGR tests to the entropy-based tests, we observe three notable features. First, $\widetilde{\mathcal{T}}_{1,n}(\mathbb{S})$-based tests or $\widetilde{\mathcal{T}}^s_{1,n}(\mathbb{S}_2)$-based tests generally dominate entropy-based tests for DGP 1.2 and 1.8. Second, for DGPs 1.3, 1.6, and 1.7, entropy-based tests are more powerful than the EGR tests. Finally, for the other DGPs, the best powered EGR tests exhibit power roughly similar to that of the best powered entropy-based tests.

Such mixed results are common in the model specification testing literature, especially in non-parametric contexts where there is no generally optimal test. For example, Fan and Li (2000) compare the power properties of specification tests using kernel-based nonparametric statistics with Bierens and Ploberger's (1997) integrated conditional moment (ICM) tests. They find that these tests are complementary, with differing power depending on the type of local alternative. Similarly, the entropy-based and EGR tests can also be used as complements.

In addition, we conducted Monte Carlo simulations for higher-order Markov processes. As the results are quite similar to those in Tables III and IV, we omit them for brevity.

For structural break alternatives, we compare our PEGR tests to a variety of well-known tests. These include Feller's (1951) and Kuan and Hornik's (1995) RR test, Brown, Durbin, and Evans's (1975) RE-CUSUM test, Sen's (1980) and Ploberger, Krämer and Kontrus's (1989) RE test, Ploberger and Krämer's (1992) OLS-CUSUM test, Andrews's (1993) Sup-W test, Andrews and Ploberger's (1994) Exp-W and Avg-W tests, and Bai's (1996) M-test.[1] As these are all designed to test for a single structural break at an unknown point, they may not perform well when there are multiple breaks. In contrast, our PEGR statistics are designed to detect general alternatives to IID, so we expect these may perform well in such situations.

We consider the following DGPs for our Monte Carlo simulations. These have been chosen to provide a test bed in which the behaviors of the various tests can be clearly contrasted.

- DGP 2.1: $Y_t := Z_t + \varepsilon_t$;
- DGP 2.2: $Y_t := \exp(Z_t) + \varepsilon_t$;
- DGP 2.3: $Y_t := \mathbf{1}_{\{t>\lfloor 0.5\cdot n\rfloor\}} + \varepsilon_t$;
- DGP 2.4: $Y_t := Z_t\mathbf{1}_{\{t\leq\lfloor 0.5\cdot n\rfloor\}} - Z_t\mathbf{1}_{\{t>\lfloor 0.5\cdot n\rfloor\}} + \varepsilon_t$;
- DGP 2.5: $Y_t := Z_t\mathbf{1}_{\{t\leq\lfloor 0.3n\rfloor\}} - Z_t\mathbf{1}_{\{t>\lfloor 0.3n\rfloor\}} + \varepsilon_t$;
- DGP 2.6: $Y_t := Z_t\mathbf{1}_{\{t\leq\lfloor 0.1n\rfloor\}} - Z_t\mathbf{1}_{\{t>\lfloor 0.1n\rfloor\}} + \varepsilon_t$;
- DGP 2.7: $Y_t := \exp(Z_t)\mathbf{1}_{\{t\leq\lfloor 0.5\cdot n\rfloor\}} + \exp(-Z_t)\mathbf{1}_{\{t>\lfloor 0.5\cdot n\rfloor\}} + \varepsilon_t$;
- DGP 2.8: $Y_t := Z_t\mathbf{1}_{\{t\in K_n(0.02)\}} - Z_t\mathbf{1}_{\{t\notin K_n(0.02)\}} + \varepsilon_t$;
- DGP 2.9: $Y_t := Z_t\mathbf{1}_{\{t\in K_n(0.05)\}} - Z_t\mathbf{1}_{\{t\notin K_n(0.05)\}} + \varepsilon_t$;
- DGP 2.10: $Y_t := Z_t\mathbf{1}_{\{t\in K_n(0.1)\}} - Z_t\mathbf{1}_{\{t\notin K_n(0.1)\}} + \varepsilon_t$;
- DGP 2.11: $Y_t := Z_t\mathbf{1}_{\{t=\text{odd}\}} - Z_t\mathbf{1}_{\{t=\text{even}\}} + \varepsilon_t$;
- DGP 2.12: $Y_t := \exp(0.1\cdot Z_t)\mathbf{1}_{\{t=\text{odd}\}} + \exp(Z_t)\mathbf{1}_{\{t=\text{even}\}} + \varepsilon_t$,

where $Z_t = 0.5Z_{t-1} + u_t$; $(\varepsilon_t, u_t)' \sim$ IID $N(\mathbf{0}, \mathbf{I}_2)$; and $K_n(r) := \{t = 1, ..., n : (k-1)/r + 1 \leq t \leq k/r, k = 1, 3, 5, ...\}$.

For DGPs 2.1, 2.4–2.6, and 2.8–2.11, we use ordinary least squares (OLS) to estimate the parameters of a linear model $Y_t = \alpha + \beta Z_t + v_t$, and we apply our PEGR statistics to the prediction errors $\hat{v}_t := Y_t - \hat{\alpha} - \hat{\beta}Z_t$. For DGP 2.3, we specify the model $Y_t = \alpha + v_t$, and we apply our PEGR statistic to $Y_t - n^{-1}\sum_{t=1}^{n} Y_t$. The linear model is correctly specified for DGP 2.1, but is misspecified for DGPs 2.3–2.6

[1]We also examined Chu, Hornik, and Kuan's (1995a) ME test and Chu, Hornik, and Kuan's (1995b) RE-MOSUM and OLS-MOSUM tests. Their performance is comparable to that of the other prior tests, so for brevity we do not report those results here.

and 2.8–2.11. Thus, when DGP 2.1 is considered the null hypothesis holds, permitting an examination of the level of the tests. As the model is misspecified for DGPs 2.3–2.6 and 2.8–2.11, the alternative holds for $\hat{\upsilon}_t$, permitting an examination of power. DGPs 2.3–2.6 exhibit a single structural break at different break points, permitting us to see how the PEGR tests compare to standard structural break tests specifically designed to detect such alternatives. DGPs 2.8 through 2.11 are deterministic mixtures in which the true coefficient of Z_t depends on whether or not t belongs to a particular structural regime. The number of structural breaks increases as the sample size increases, but the proportion of breaks to the sample size is constant. Also, the break points are equally spaced. Thus, for example, there are four break points in DGP 2.8 when the sample size is 100 and and nine break points when the sample size is 200. The extreme case is DGP 2.11, in which the proportion of breaks is equal to one, and the coefficient of Z_t depends on whether or not t is even. Given the regular pattern of these breaks, this may be hard to distinguish from a DGP without a structural break.

For DGPs 2.2, 2.7, and 2.12, we use nonlinear least squares (NLS) to estimate the parameters of a nonlinear model $Y_t = \exp(\beta Z_t) + \upsilon_t$, and we apply our PEGR statistics to the prediction errors $\hat{\upsilon}_t := Y_t - \exp(\hat{\beta} Z_t)$. The situation is analogous to that for the linear model, in that the null holds for DGP 2.2, whereas the alternative holds for 2.7 and 2.12. Examining these alternatives permits an interesting comparison of the PEGR tests, designed for general use, to the RR, RE, M, OLS-CUSUM and RE-CUSUM statistics, which are expressly designed for use with linear models.

Our simulation results are presented in Tables V to VII. To summarize, the levels of the PEGR tests are approximately correct for both linear and nonlinear cases and generally improve as the sample size increases. On the other hand, there are evident level distortions for some of the other statistics, especially, as expected, for the linear model statistics with nonlinear DGP 2.2. The PEGR statistics also have respectable power. They appear consistent against our structural break alternatives, although the PEGR tests are not as powerful as the other (properly sized) break tests when there is a single structural break. This is as expected, as the other tests are specifically designed to detect a single break, whereas the PEGR test is not. As one might also expect, the power of the PEGR tests diminishes notably as the break moves away from the center of the sample. Nevertheless, the relative performance of the tests reverses when there are multiple breaks. All test statistics lose power as the proportion of breaks increases, but the loss of power for the non-PEGR tests is much faster than for the PEGR tests. For the extreme alternative DGP 2.11, the PEGR tests appear to be the only consistent tests.

We also note that, as for the dependent alternatives, the integral norm-based tests outperform the supremum norm-based tests.

Finally, we briefly summarize the results of other interesting experiments omitted from our tabulations for the sake of brevity. To further examine level properties, we applied our EGR tests (i) with $Y_t \sim$ IID $C(0,1)$, where $C(\ell,s)$ denotes the Cauchy distribution with location and scale parameters ℓ and s respectively, and (ii) with $Y_t = (u_t+1)\mathbf{1}_{\{\varepsilon_t \geq 0\}} + (u_t - 1)\mathbf{1}_{\{\varepsilon_t < 0\}}$, where $(\varepsilon_t, u_t) \sim$ IID $N(\mathbf{0}, \mathbf{I}_2)$. We consider the Cauchy process to examine whether the absence of moments in the raw data matters, and we consider the normal random mixture to compare the results with the deterministic mixture, DGP 2.10. Our experiments yielded results very similar to those reported for DGP 1.1. This affirms the claims for the asymptotic null distributions of the (P)EGR test statistics in the previous sections. To further examine the power of our (P)EGR tests, we also considered the mean shift processes analyzed by Crainiceanu and Vogelsang (2007), based on DGP 2.3. Our main motivation for this arises from the caveat in the literature that CUSUM and CUSQ tests may exhibit power functions non-monotonic in α. (See Deng and Perron (2008) for further details.) In contrast, we find that the (P)EGR test statistics do not exhibit this non-monotonicity.

6 Conclusion

The IID assumption plays a central role in economics and econometrics. Here we provide a family of tests based on generalized runs that are powerful against unspecified alternatives, providing a useful complement to tests designed to have power against specific alternatives, such as serial correlation, GARCH, or structural breaks. Relative to other tests of this sort, for example the entropy-based tests of Hong and White (2005), our tests have an appealing computational simplicity, in that they do not require kernel density estimation, with the associated challenge of bandwidth selection.

Our simulation studies show that our tests have empirical levels close to their nominal asymptotic levels. They also have encouraging power against a variety of important alternatives. In particular, they have power against dependent alternatives and heterogeneous alternatives, including those involving a number of structural breaks increasing with the sample size.

7 Appendix

7.1 Proofs

To prove our main results, we first state some preliminary lemmas. Recall that $\mathbb{J} := [\underline{p}, 1], \underline{p} > 0$, and for notational simplicity for every $p, p' \in \mathbb{I}$ with $p' \leq p$ and $M_n(p') > 0$, we let $K_{n,i}$ denote $K_{n,i}(p,p')$ such that $K_{n,0}(p,p') = 0$ and $\sum_{j=K_{n,i-1}(p,p')+1}^{K_{n,i}(p,p')} R_{n,j}(p) = R_{n,i}(p')$.

LEMMA A1: *Given A1, A2(*i*), A3, and* $\mathbb{H}_0$, *if* $s \in \mathbb{S}$, $p', p \in \mathbb{I}$, *and* $p' \leq p$ *such that* $M_n(p') > 0$, *then*

$E[\sum_{i=1}^{K_{n,1}} s^{R_{n,i}(p)}] = E[s^{R_{n,1}(p)}] \cdot E[K_{n,1}]$.

Proof of Lemma A1: For each $s, \ell \in \mathbb{N}$, $\mathbb{P}(R_{n,i}(p) = s|K_{n,1} = \ell) = \mathbb{P}(K_{n,1} = \ell|R_{n,i}(p) = s)\mathbb{P}(R_{n,i}(p) = s)/\mathbb{P}(K_{n,1} = \ell) = \mathbb{P}(K_{n,1} = \ell)\mathbb{P}(R_{n,i}(p) = s)/\mathbb{P}(K_{n,1} = \ell) = \mathbb{P}(R_{n,i}(p) = s)$, where the second equality holds because the event $\{K_{n,1} = \ell\}$ is independent of $\{R_{n,i}(p) = s\}$ under $\mathbb{H}_0$. Therefore, $E[\sum_{i=1}^{K_{n,1}} s^{R_{n,i}(p)}] = \sum_{\ell=1}^{\infty} E[\sum_{i=1}^{K_{n,1}} s^{R_{n,i}(p)}|K_{n,1} = \ell]\mathbb{P}(K_{n,1} = \ell) = E[s^{R_{n,1}(p)}] \sum_{\ell=1}^{\infty} \ell\ \mathbb{P}(K_{n,1} = \ell) = E[s^{R_{n,1}(p)}] \cdot E[K_{n,1}]$, where the second equality follows because $\mathbb{H}_0$ implies that $\{R_{n,i}(p)\}$ is IID. ■

LEMMA A2: *Given A1, A2(i), A3, and* $\mathbb{H}_0$, *if* $s \in \mathbb{S}$, $p', p \in \mathbb{J}$ *and* $p' \leq p$ *such that* $M_n(p') > 0$, *then*

$$E[K_{n,1}s^{R_{n,1}(p')}] = \frac{sp'\{1 - s(1 - p)\}}{\{1 - s(1 - p')\}^2}. \tag{17}$$

Proof of Lemma A2: We treat three distinct cases: (a) $p' = p = 1$; (b) $p' < p = 1$; and (c) $p' \leq p < 1$. (a) Let $p' = p = 1$. Then $R_{n,1}(p') = 1$ and $K_{n,1} = 1$, so that $E[K_{n,1}s^{R_{n,1}(p')}] = s$. Plugging $p' = p = 1$ into the RHS of (17) also gives s, verifying the result. (b) Next, suppose $p' < p = 1$. Then $K_{n,1} = R_{n,1}(p')$, implying that $E[K_{n,1}s^{R_{n,1}(p')}] = E[R_{n,1}(p')s^{R_{n,1}(p')}] = sp'/\{1 - s(1 - p')\}^2$, as can be verified by direct computation. This coincides with the RHS of (17) with $p = 1$. (c) Finally, let $p' \leq p < 1$. First, $E[K_{n,1}s^{R_{n,1}(p')}] = E[s^{R_{n,1}(p')}E[K_{n,1}|R_{n,1}(p')]]$, and $\mathbb{P}(K_{n,1} = \ell|R_{n,1}\ (p') = r_1') = \mathbb{P}(K_{n,1} = \ell, R_{n,1}(p') = r_1')/\mathbb{P}(R_{n,1}(p') = r_1') = \binom{r_1'-1}{\ell-1}(p - p')^{\ell-1}(1 - p)^{r_1'-\ell}/\ (1 - p')^{r_1'-1}$, where the last equality follows from the fact that $\mathbb{P}(R_{n,1}(p') = r_1') = p'(1-p')^{r_1'-1}$ and $\mathbb{P}(K_{n,1} = \ell, R_{n,1}(p') = r_1') = \binom{r_1'-1}{\ell-1}(p - p')^{\ell-1}(1 - p)^{r_1'-\ell}$. Thus, $E[K_{n,1}|R_{n,1}(p') = r_1'] = \sum_{\ell=1}^{\infty} \ell \cdot \mathbb{P}(K_{n,1} = \ell|R_{n,1}(p') = r_1') = 1+(r_1'-1)(p-p')/(1-p')$, implying that $E[K_{n,1}s^{R_{n,1}(p')}] = E[s^{R_{n,1}(p')}]+E[s^{R_{n,1}(p')}(R_{n,1}(p')-1)](p - p')/(1 - p')$. Second, we note that $E[s^{R_{n,1}(p')}(R_{n,1}(p') - 1)] = s^2(d/ds)E[s^{R_{n,1}(p')-1}] = s^2p'(1 - p')/\{1 - s(1 - p')\}^2$. Therefore, $E[K_{n,1}s^{R_{n,1}(p')}] = sp'/\{1 - s(1 - p')\} + [\{p - p'\}/\{1 - p'\}]s^2p'(1 - p')/\{1 - s(1 - p')\}^2 = sp'\{1 - s(1 - p)\}/\{1 - s(1 - p')\}^2$. This completes the proof. ■

In the special case in which $p = p'$, we have $K_{n,1} = 1$ and $E(s^{R_{n,1}(p)}) = sp/(1 - s(1 - p))$.

LEMMA A3: *Given A1, A2(i), A3, and* $\mathbb{H}_0$, *if* $s \in \mathbb{S}$, $p', p \in \mathbb{J}$, *and* $p' \leq p$ *such that* $M_n(p') > 0$, *then* $E[\sum_{i=1}^{K_{n,1}} s^{R_{n,i}(p)+R_{n,1}(p')}] = s^2p'/\{1 - s^2(1 - p)\} \cdot [\{1 - s(1 - p)\}/\{1 - s(1 - p')\}]^2$.

Proof of Lemma A3: First, we note that $E[\sum_{i=1}^{K_{n,1}} s^{R_{n,i}(p)+R_{n,1}(p')}] = E[\sum_{i=1}^{K_{n,1}} E[s^{R_{n,i}(p)+R_{n,1}(p')}|\ K_{n,1}]]$ by Lemma A8 given below, and $E[s^{R_{n,i}(p)+R_{n,1}(p')}|K_{n,1}] = E[s^{R_{n,i}(p)+\sum_{j=1}^{K_{n,1}} R_{n,j}(p)}|K_{n,1}] = E[s^{2R_{n,i}(p)} |K_{n,1}] \prod_{j=1,j\neq i}^{K_{n,1}} E[s^{R_{n,j}(p)}|K_{n,1}] = [\frac{s^2p}{\{1-s^2(1-p)\}}][\frac{sp}{\{1-s(1-p)\}}]^{K_{n,1}-1}$ by Lemma A.2, so that $E[\sum_{i=1}^{K_{n,1}} s^{R_{n,i}(p)+R_{n,1}(p')}] = E[K_{n,1}[\frac{sp}{\{1-s(1-p)\}}]^{K_{n,1}-1}][\frac{s^2p}{\{1-s^2(1-p)\}}]$. Next, $\mathbb{P}(K_{n,1} = k) = (p/p')\ [(p-p')/p]^{k-1}$

by Lemma A4 below. From this, $E[K_{n,1}[sp/\{1-s(1-p)\}]^{K_{n,1}-1}] = p'p^{-1}\sum_{k=1}^{\infty} k[s(p-p')/\{1-s(1-p)\}]^{k-1} = p'p^{-1}[\{1-s(1-p)\}/\{1-s(1-p')\}]^2$. The desired result follows by substituting this into $E[\sum_{i=1}^{K_{n,1}} s^{R_{n,i}(p)+R_{n,1}(p')}]$. ■

LEMMA A4: *Given A1, A2(i), A3, and* $\mathbb{H}_0$, *if* $p', p \in \mathbb{J}$ *and* $p' \leq p$ *such that* $M_n(p') > 0$, *then* $\mathbb{P}(K_{n,1} = k) = (p/p')[(p-p')/p]^{k-1}$ *and* $E[K_{n,1}] = p/p'$.

Proof of Lemma A4: First, $\mathbb{P}(K_{n,1} = 1) = \mathbb{P}(R_{n,1}(p') = R_{n,1}(p)) = \sum_{\ell=1}^{\infty}\mathbb{P}(R_{n,1}(p') = R_{n,1}(p) = \ell) = \mathbb{P}(F(X_1) < p') + \sum_{\ell=2}^{\infty}\mathbb{P}(F(X_1) \geq p, ..., F(X_{\ell-1}) \geq p, F(X_\ell) < p') = \sum_{\ell=1}^{\infty} p'(1-p)^{\ell-1} = p'p^{-1}$. Next, $\mathbb{P}(K_{n,1} = 2) = \mathbb{P}(R_{n,1}(p') + R_{n,2}(p') = R_{n,1}(p)) = \sum_{\ell_1=1}^{\infty}\sum_{\ell_2=1}^{\infty}\mathbb{P}(R_{n,1}(p') = \ell_1, R_{n,2}(p') = \ell_2$, and $R_{n,1}(p) = \ell_1 + \ell_2) = \sum_{\ell_1=1}^{\infty}\sum_{\ell_2=1}^{\infty}(1-p)^{\ell_1-1}(p-p')(1-p)^{\ell_2-1}p' = p'(p-p')p^{-2}$. Similarly, we obtain for an arbitrarily chosen number, say k, that $\mathbb{P}(K_{n,1} = k) = \mathbb{P}(\sum_{j=1}^{k} R_{n,j}(p') = R_{n,1}(p)) = \sum_{\ell_1=1}^{\infty} ... \sum_{\ell_k=1}^{\infty}\mathbb{P}(R_{n,1}(p') = \ell_1, R_{n,2}(p') = \ell_2, ..., R_{n,k}(p') = \ell_k$, and $R_{n,1}(p) = \sum_{j=1}^{k}\ell_j) = \prod_{j=1}^{k-1}\{\sum_{\ell_j=1}^{\infty}(1-p)^{\ell_j-1}(p'-p)\}\cdot\sum_{\ell_k=1}^{\infty}(1-p)^{\ell_k-1}p' = (p/p')[(p-p')/p]^{k-1}$.

From these, it follows directly that $E[K_{n,1}] = p/p'$. This completes the proof. ■

LEMMA A5: *Let* $p \in \mathbb{I}$ *such that* $M_n(p) > 0$. *If* $\{R_{n,i}(p)\}_{i=1}^{M_n(p)}$ *is IID with distribution* $\mathbb{G}_p$ *then, for* $m = 1, 2, ...$, *and* $\ell = m, m+1, m+2, ...$, $\mathbb{P}(\sum_{i=1}^{m} R_{n,i}(p) = \ell) = \binom{\ell-1}{m-1}(1-p)^{\ell-m}p^m$.

Proof of Lemma A5: We prove this by induction. If $m = 1$, then $\mathbb{P}(\sum_{i=1}^{m} R_{n,i}(p) = \ell) = (1-p)^{\ell-1}p$, which is the distribution of $\mathbb{G}_p$. Next, suppose that the given result holds for an arbitrary m and consider the case $m+1$. Then $\mathbb{P}(\sum_{i=1}^{m+1} R_{n,i}(p) = \ell) = \sum_{j=1}^{\ell-m-1}\mathbb{P}(\sum_{i=1}^{m} R_{n,i}(p) = \ell - j)\mathbb{P}(R_{n,m+1}(p) = j) = \sum_{j=1}^{\ell-m-1}\binom{\ell-j-1}{m-1}(1-p)^{\ell-j-m}p^m(1-p)^{j-1}p = \binom{\ell-1}{m}(1-p)^{\ell-m-1}p^{m+1}$, where the first and second equalities follow by independence and the result for m, respectively. The final equality is the desired result.

■

LEMMA A6: *Let* $p, p' \in \mathbb{I}$ *such that* $M_n(p') > 0$. *Given condition* $\mathcal{R}$ *of Lemma 1, then for* $i, k = 1, 2, ...$,

(i) *if* $\ell = i, i+1, ..., i+k-1$, $\mathbb{P}(\bigcup_{m=2}^{i+k+1-\ell}\{\sum_{j=2}^{m} R_{n,j}(p') = i+k-\ell\}) = p'$;

(ii) *when* $p > p'$,

$$\mathbb{P}(\bigcup_{m=1}^{i}\{\sum_{j=1}^{m} R_{n,j}(p) = i\}, R_{n,1}(p') = \ell) = \begin{cases} p'(1-p')^{i-1}, & \text{if } \ell = i; \\ p'(p-p')(1-p')^{\ell-2}, & \text{if } \ell = i+1, \cdots, i+k. \end{cases}$$

Proof of Lemma A6: (i) From the given condition $\mathcal{R}$, $\mathbb{P}(\bigcup_{m=2}^{i+k+1-\ell}\sum_{j=2}^{m} R_{n,j}(p') = i+k-\ell) = \sum_{m=1}^{i+k-\ell}\mathbb{P}(\sum_{j=1}^{m} R_{n,j}(p') = i+k-\ell) = \sum_{m=1}^{i+k-\ell}\binom{i+k-\ell-1}{m-1}(1-p')^{i+k-\ell-m}p'^m = p'$, where the first equality follows from the IID condition for $\{R_{n,j}(p')\}$ given in $\mathcal{R}$, and the second equality holds by Lemma A5.

(ii) First, consider the case $\ell = i$. Then, $\mathbb{P}(\bigcup_{m=1}^{i}\{\sum_{j=1}^{m} R_{n,j}(p) = i\}, R_{n,1}(p') = i) = \sum_{m=1}^{i}\binom{i-1}{m-1}(1-p)^{i-m}(p-p')^{m-1}p' = p'(1-p')^{i-1}$, where the first equality holds by the second condition in $\mathcal{R}$.

Next, we let $\ell = i+1, ..., i+k$. Then, $\mathbb{P}(\bigcup_{m=1}^{i}\{\sum_{j=1}^{m} R_{n,j}(p) = i\}, R_{n,1}(p') = \ell) = \sum_{m=1}^{i}\binom{i-1}{m-1}(1-p)^{i-m}(p-p')^{m}(1-p')^{\ell-(i+1)}p' = p'^2(p-p')(1-p')^{\ell-2}$, where the first equality holds by the second condition in $\mathcal{R}$. This completes the proof. ■

LEMMA A7: *Let $p, p' \in \mathbb{I}$ such that $M_n(p') > 0$. Given condition $\mathcal{R}$ of Lemma 1 and $p > p'$, then for $i, k = 1, 2, ...$, $\sum_{\ell=i}^{i+k-1}\mathbb{P}(\bigcup_{m=1}^{i}\{\sum_{j=1}^{m} R_{n,j}(p) = i\}, R_{n,1}(p') = \ell, \bigcup_{m=2}^{i+k+1-\ell}\{\sum_{j=2}^{m} R_{n,j}(p') = i+k-\ell\}) + \mathbb{P}(\bigcup_{m=1}^{i}\{\sum_{j=1}^{m} R_{n,j}(p) = i\}, R_{n,1}(p') = i+k) = pp'(1-p')^{i-1}$.*

Proof of Lemma A7: First, for each $\ell = i, i+1, ..., i+k-1$, we let $C_{1,\ell} := \mathbb{P}(\bigcup_{m=1}^{i}\{\sum_{j=1}^{m} R_{n,j}(p) = i\}, R_{n,1}(p') = \ell, \bigcup_{m=2}^{i+k+1-\ell}\{\sum_{j=2}^{m} R_{n,j}(p') = i+k-\ell\})$. Then, the second condition in $\mathcal{R}$ implies that $C_{1,\ell} = \mathbb{P}(\bigcup_{m=1}^{i}\{\sum_{j=1}^{m} R_{n,j}(p) = i\}, R_{n,1}(p') = \ell)\mathbb{P}(\bigcup_{m=2}^{i+k+1-\ell}\{\sum_{j=2}^{m} R_{n,j}(p') = i+k-\ell\})$. Given this, by applying the results in Lemma A6, we obtain that $C_1 := \sum_{\ell=i}^{i+k-1} C_{1,\ell} = p'^2(1-p')^{i-1} + \sum_{\ell=i+1}^{i+k-1} p'^2(p-p')(1-p')^{\ell-2} = pp'(1-p')^{i-1} - p'(p-p')(1-p')^{i+k-2}$. Next, we have $C_2 := \mathbb{P}(\bigcup_{m=1}^{i}\{\sum_{j=1}^{i} R_{n,j}(p) = i\}, R_{n,1}(p') = i+k) = \sum_{m=1}^{i}\mathbb{P}(\sum_{j=1}^{i} R_{n,j}(p) = i, R_{n,1}(p') = i+k) = \sum_{m=1}^{i}(1-p)^{i-m}(p-p')^{m}(1-p')^{k-1}p' = p'(p-p')(1-p')^{i+k-2}$. Therefore, $C_1 + C_2 = pp'(1-p')^{i-1}$. This is the desired result. ■

LEMMA A8: *Let K be a random positive integer, and let $\{X_t\}$ be a sequence of random variables such that for each $i = 1, 2, ..., E(X_i) < \infty$. Then $E(\sum_{i=1}^{K} X_i) = E(\sum_{i=1}^{K} E(X_i|K))$.*

We omit proving Lemma A.8, as it is elementary.

Before proving Lemma 1, we define several relevant notions. First, for $p \in \mathbb{I}$ with $M_n(p) > 0$, we define the *building time to i* by $B_{n,i}(p) := i - \sum_{j=1}^{U_{n,i}(p)} R_{n,j}(p)$, where $U_{n,i}(p)$ is the maximum number of runs such that $\sum_{j=1}^{w} R_{n,j}(p) < i$; i.e., $U_{n,i}(p) := \max\{w \in \mathbb{N} : \sum_{j=1}^{w} R_{n,j}(p) < i\}$. Now $B_{n,i}(p) \in \{1, 2, ..., i-1, i\}$; and if $B_{n,i}(p) = i$ then $R_{n,1}(p) \geq i$. For $p, p' \in \mathbb{I}$, $p' < p$, with $M_n(p') > 0$, we also let $W_{n,i}(p, p')$ be the number of runs $\{R_{n,i}(p')\}$ such that $\sum_{j=1}^{U_{n,i}(p)} R_{n,j}(p) = \sum_{j=1}^{W_{n,i}(p,p')} R_{n,j}(p')$.

Proof of Lemma 1: As part (A) is easy, we prove only part (B). We first show that $\mathcal{R}$ implies that the original data $\{Y_t\}$ are independent. For this, we show that for any pair of two variables, (Y_i, Y_{i+k}) $(i, k \geq 1)$, say, $\mathbb{P}(F_i(Y_i) \leq p, F_{i+k}(Y_{i+k}) \leq p') = pp'$. We partition our consideration into three cases: (a) $p = p'$; (b) $p < p'$; and (c) $p > p'$ and obtain the given equality for each case. (a) Let $p = p'$. We have $\mathbb{P}(F_i(Y_i) \leq p,\ F_{i+k}(Y_{i+k}) \leq p') = \mathbb{P}(\bigcup_{m=1}^{i}\{\sum_{j=1}^{m} R_{n,j}(p) = i\}, \bigcup_{h=m+1}^{k+m}\{\sum_{j=m+1}^{h} R_{n,j}(p) = k\}) = \sum_{m=1}^{i}\mathbb{P}(\ \{\sum_{j=1}^{m} R_{n,j}(p) = i\})\mathbb{P}(\bigcup_{h=1}^{k}\{\sum_{j=1}^{h} R_{n,j}(p) = k\}) = \sum_{m=1}^{i}\mathbb{P}(\{\sum_{j=1}^{m} R_{n,j}(p) =$

$i\}) \sum_{h=1}^{k} \mathbb{P}(\sum_{j=1}^{h} R_{n,j}(\ p) = k) = p \cdot p = p^2$, where the second and third equalities follow from the condition $\mathcal{R}$ and Lemma A5 respectively. (b) Next, suppose $p < p'$. We have $\mathbb{P}(F_i(Y_i) \leq p, F_{i+k}(Y_{i+k}) \leq p') = \sum_{h=1}^{i} \mathbb{P}(\bigcup_{m=1}^{i}\{\sum_{j=1}^{m} R_{n,j}\ (p) = i\}, \sum_{j=1}^{h} R_{n,j}(p') = i, \bigcup_{m=h+1}^{k+h}\{\sum_{j=h+1}^{m} R_{n,j}(p') = k\}) = \sum_{h=1}^{i} \mathbb{P}(\bigcup_{m=1}^{i}\{\sum_{j=1}^{m} R_{n,j}(p) = i\}, \sum_{j=1}^{h}\ R_{n,j}(p') = i) \sum_{m=1}^{k} \mathbb{P}(\sum_{j=1}^{m} R_{n,j}(p') = k)$, where the second equality follows from $\mathcal{R}$. Further, Lemma A5 implies that $\sum_{h=1}^{i} \mathbb{P}(\bigcup_{m=1}^{i}\{\sum_{j=1}^{m} R_{n,j}(p) = i\}, \sum_{j=1}^{h} R_{n,j}(p') = i) = \mathbb{P}(\bigcup_{h=1}^{i}\{\sum_{j=1}^{h} R_{n,j}(p') = i\}, \bigcup_{m=1}^{i} \sum_{j=1}^{m} R_{n,j}(p) = i) = \sum_{m=1}^{i} \mathbb{P}(\sum_{j=1}^{m} R_{n,j}(p) = i) = p$, and $\sum_{m=1}^{k} \mathbb{P}(\sum_{j=1}^{m} R_{n,j}(p') = k) = p'$. Thus, $\mathbb{P}(F_i(Y_i) \leq p, F_{i+k}(Y_{i+k}) \leq p') = pp'$. (c) Finally, let $p' < p$. We have $\mathbb{P}(F_i(Y_i) < p, F_{i+k}(Y_{i+k}) < p') = \sum_{b=1}^{i} \mathbb{P}(F_i(Y_i) < p, F_{i+k}(Y_{i+k}) < p', B_{n,i}(p) = b)$ and derive each term constituting this sum separately. We first examine the case $b = i$. Then $\mathbb{P}(F_i(Y_i) < p, F_{i+k}(Y_{i+k}) < p', B_{n,i}(p) = i) = \sum_{\ell=i}^{i+k-1} \mathbb{P}(\bigcup_{m=1}^{i}\{\sum_{j=1}^{m} R_{n,j}(p) = i\}, R_{n,1}(p') = \ell, \bigcup_{m=2}^{i+k+1-\ell}\{\sum_{j=2}^{m} R_{n,j}(p') = i + k - \ell\}) + \mathbb{P}(\bigcup_{m=1}^{i}\{\sum_{j=1}^{m} R_{n,j}(p) = i\}, R_{n,1}(p') = i + k) = pp'(1 - p')^{i-1}$, where the last equality follows from Lemma A7. Next, we consider the cases $b = 1, 2, ..., i - 1$. Then it follows that $\mathbb{P}(F_i(Y_i) < p, F_{i+k}(Y_{i+k}) < p', B_{n,i}(p) = b) = [\sum_{\ell=b}^{b+k-1} \mathbb{P}(\bigcup_{m=1+w}^{b+w}\{\sum_{j=1+w}^{m} R_{n,j}(p) = b\}, R_{n,u+1}(p') = \ell, \bigcup_{m=u+2}^{b+k+u+1-\ell}\{\sum_{j=u+2}^{m} R_{n,j}(p') = b+k-\ell\}) + \mathbb{P}(\bigcup_{m=1+w}^{b+w}\{\sum_{j=1+w}^{m} R_{n,j}(p) = b\}, R_{n,u+1}(p') = b + k)] \times \mathbb{P}(\bigcup_{m=1}^{i-b}\{\sum_{j=1}^{m} R_{n,j}(p') = i - b\})$, where w and u are short-hand notations for $W_{n,i}(p, p')$ and $U_{n,i}(p)$. Given this, we further note that $\mathbb{P}(\bigcup_{m=1}^{i-b}\{\sum_{j=1}^{m} R_{n,j}(p') = i-b\}) = \sum_{m=1}^{i-b} \mathbb{P}(\sum_{j=1}^{m} R_{n,j}(p') = i-b) = p'$ by Lemma A5; the condition $\mathcal{R}$ implies that $\mathbb{P}(\bigcup_{m=1+w}^{b+w}\{\sum_{j=1+w}^{m} R_{n,j}(p) = b\}, R_{n,u+1}(p') = b + k) = \mathbb{P}(\bigcup_{m=1}^{b}\{\sum_{j=1}^{m} R_{n,j}(p) = b\}, R_{n,u+1}(p') = b+k)$ and for $\ell = b, b+1, ..., b+k-1$, $\mathbb{P}(\bigcup_{m=1+w}^{b+w}\{\sum_{j=1+w}^{m} R_{n,j}\ (p) = b\}, R_{n,u+1}(p') = \ell, \bigcup_{m=u+2}^{b+k+u+1-\ell}\{\sum_{j=u+2}^{m} R_{n,j}(p') = b + k - \ell\}) = \mathbb{P}(\bigcup_{m=1}^{b}\{\sum_{j=1}^{m} R_{n,j}(p) = b\}, R_{n,1}(p') = \ell, \bigcup_{m=2}^{b+k+1-\ell}\ \{\sum_{j=2}^{m} R_{n,j}(p') = b + k - \ell\})$, so that $\mathbb{P}(F_i(Y_i) < p,\ F_{i+k}(Y_{i+k}) < p', B_{n,i}(p) = b) = \mathbb{P}(F_b(Y_i) < p, F_{b+k}(Y_{b+k}) < p', B_{n,b}(p) = b)p' = pp'^2(1 - p')^{b-1}$. Hence, $\mathbb{P}(F_i(Y_i) < p, F_{i+k}(Y_{i+k}) < p') = \sum_{b=1}^{i} \mathbb{P}(F_i(Y_i) < p, F_{i+k}(Y_{i+k}) < p', B_{n,i}(p) = b) = \sum_{b=1}^{i-1} pp'^2(1-p')^{b-1} + pp'(1-p')^{i-1} = pp'$. Thus, Y_i and Y_{i+k} are independent.

Next, suppose that $\{Y_t\}$ is not identically distributed. Then there is a pair, say (Y_i, Y_j), such that for some $y \in \mathbb{R}$, $p_i := F_i(y) \neq p_j := F_j(y)$. Further, for the same y, $\mathbb{P}(R_{n,(j)}(p_j) = 1) = \mathbb{P}(F_j(Y_j) \leq F_j(y)|F_{j-1}(Y_{j-1}) \leq F_{j-1}(y)) = \mathbb{P}(F_j(Y_j) \leq F_j(y)) = p_j$, where the subscript (j) denotes the (j)-th run of $\{R_{n,i}(p_j)\}$ corresponding to $F_j(y)$, and the second equality follows from the independence property just shown. Similarly, $\mathbb{P}(R_{n,(i)}(p_j) = 1) = \mathbb{P}(F_j(Y_i) \leq F_j(y)) = \mathbb{P}(Y_i \leq y) = p_i$. That is, $\mathbb{P}(R_{n,(j)}(p_j) = 1) \neq \mathbb{P}(R_{n,(i)}(p_j) = 1)$. This contradicts the assumption that $\{R_{n,i}(p)\}$ is identically distributed for all $p \in \mathbb{I}$. Hence, $\{Y_t\}$ must be identically distributed. This completes the proof. ■

Proof of Theorem 1: (i) We separate the proof into three parts. In (a), we prove weak convergence of $G_n(p, \cdot)$. In (b), we show $E\,[\mathcal{G}(p, s)] = 0$ for each $s \in \mathbb{S}$. Finally, (c) derives $E\,[\mathcal{G}(p, s)\mathcal{G}(p, s')]$.

(a) First, we show that for some $\bar{\Delta}_1 > 0$, $E[\{G_n(p,s)-G_n(p,s')\}^4] \leq \bar{\Delta}_1|s-s'|^4$. Note that for each p, $E[\{G_n(p,s)-G_n(p,s')\}^4] = (1-n^{-1})E[\{\mathcal{G}(p,s)-\mathcal{G}(p,s')\}^4]+n^{-2}E[\sum_{i=1}^{M_n(p)}\{s^{R_{n,i}(p)}-s'^{R_{n,i}(p)}+\frac{sp}{\{1-s(1-p)\}}-\frac{s'p}{\{1-s'(1-p)\}}\}^4] \leq 2E[\{\mathcal{G}(p,s)-\mathcal{G}(p,s')\}^4]+n^{-1}E[\{s^{R_{n,i}(p)}-s'^{R_{n,i}(p)}+\frac{sp}{\{1-s(1-p)\}}-\frac{s'p}{\{1-s'(1-p)\}}\}^4]$, a consequence of finite dimensional weak convergence, which follows from Lindeberg-Levy's central limit theorem (CLT) and the Cramér-Wold device. We examine each piece on the RHS separately. It is independently shown in Theorem 4(i) that $\mathcal{G}(p,\cdot) \stackrel{d}{=} \mathcal{Z}_p$. Thus, $E[\{\mathcal{G}(p,s)-\mathcal{G}(p,s')\}^4] = E[\{\mathcal{Z}_p(s)-\mathcal{Z}_p(s')\}^4]$ uniformly in p. If we let $m_p(s) := sp(1-s)(1-p)^{1/2}\{1-s(1 -p)\}^{-1}$ and $B_j(p) := (1-p)^{j/2}Z_j$ for notational simplicity, then $\mathcal{Z}_p(s) = m_p(s)\sum_{j=0}^{\infty}s^jB_j(p)$, and it follows that $\{\mathcal{Z}_p(s)-\mathcal{Z}_p(s')\}^4 = \{\mathcal{A}_p(s)[m_p(s)-m_p(s')]+m_p(s')\mathcal{B}_p(s)(s-s')\}^4$, where $\mathcal{A}_p(s) := \sum_{j=0}^{\infty}s^jB_j(p)$ and $\mathcal{B}_p(s) := \sum_{j=0}^{\infty}\sum_{k=0}^{j}s'^{j-k}s^kB_j(p)$. We can also use the mean-value theorem to obtain that for some s'' between s and s', $m_p(s)-m_p(s') = m_p'(s'')(s-s')$. Therefore, if we let $\Delta_1 := (1-\underline{s})^2(1-\widetilde{s})^{-2}$ with $\widetilde{s} := \max[|\underline{s}|,\bar{s}]$, $E[\{\mathcal{Z}_p(s)-\mathcal{Z}_p(s')\}^4] = E[\{\mathcal{A}_p(s)m_p'(s'')+m_p(s')\mathcal{B}_p(s)\}^4]|s-s'|^4 \leq \Delta_1^4\{|E[\mathcal{A}_p(s)^4]|+4|E[\mathcal{A}_p(s)^3\mathcal{B}_p(s)]|+6|E[\mathcal{A}_p(s)^2\mathcal{B}_p(s)^2]|+4|E[\mathcal{A}_p(s)\mathcal{B}_p(s)^2]|+|E[\mathcal{B}_p(s)^4]|\}|s-s'|^4$, because $\sup_{p,s}|m_p'(s)| \leq \Delta_1$ and $\sup_{p,s}|m_p(s)| \leq \Delta_1$. Some tedious algebra shows that $\sup_{(p,s)\in\mathbb{I}\times\mathbb{S}} E[\mathcal{A}_p(s)^4] \leq \Delta_2 := \frac{6}{(1-\widetilde{s}^4)^2}$, and $\sup_{(p,s)\in\mathbb{I}\times\mathbb{S}}E[\mathcal{B}_p(s)^4] \leq \Delta_3 := \frac{72}{(1-\widetilde{s}^4)^5}$, so that $E[\{\mathcal{Z}_p(s)-\mathcal{Z}_p(s')\}^4] \leq 16\Delta_1^4\Delta_3|s-s'|^4$. Using Hölder's inequality, we obtain $|E[\mathcal{A}_p(s)^3\mathcal{B}_p(s)]| \leq |E[\mathcal{A}_p(s)^4]^{3/4} E[\mathcal{B}_p(s)^4]^{1/4} \leq \Delta_2^{3/4}\Delta_3^{1/4} \leq \Delta_3$, $|E[\mathcal{A}_p(s)^2\mathcal{B}_p(s)^2]| \leq |E[\mathcal{A}_p(s)^4]^{2/4}E[\mathcal{B}_p(s)^4]^{2/4} \leq \Delta_2^{2/4}\Delta_3^{2/4} \leq \Delta_3$, and $|E[\mathcal{A}_p(s)\mathcal{B}_p(s)^3]| \leq |E[\mathcal{A}_p(s)^4]^{3/4}E[\mathcal{B}_p(s)^4]^{1/4} \leq \Delta_2^{3/4}\Delta_3^{1/4} \leq \Delta_3$, where the final inequalities follow from the fact that $\Delta_2 \leq \Delta_3$. Next, we note that $|s^{R_{n,i}(p)}-s'^{R_{n,i}(p)}| \leq R_{n,i}(p)\widetilde{s}^{R_{n,i}(p)}|s-s'|$ and $|sp/\{1-s(1-p)\}-s'p/\{1-s'(1-p)\}| \leq \frac{1}{(1-\widetilde{s})^2}|s-s'|$ with $E[R_{n,i}(p)^4\widetilde{s}^{4R_{n,i}(p)}] \leq 24(1-\widetilde{s}^4)^{-5}$. Thus, when we let $Q_{n,i} := R_{n,i}(p)\widetilde{s}^{R_{n,i}(p)}+(1-\widetilde{s})^{-2}$, it follows that $E[Q_{n,i}^4] \leq \widetilde{\Delta}_1 := 384\times(1-\widetilde{s}^4)^{-5}(1-\widetilde{s})^{-8}$, and $E[\{s^{R_{n,i}(p)}-s'^{R_{n,i}(p)}+sp/\{1-s(1-p)\}-s'p/\{1-s'(1-p)\}\}^4] \leq \widetilde{\Delta}_1|s-s'|^4$. Given this, if $\bar{\Delta}_1$ is defined by $\bar{\Delta}_1 := (32\Delta_1^4\Delta_3+\widetilde{\Delta}_1)$ then $E[|G_n(p,s)-G_n(p,s')|^4] \leq \bar{\Delta}_1|s-s'|^4$.

Second, therefore, if we let $s'' \leq s' \leq s$, $E[|G_n(p,s)-G_n(p,s')|^2|G_n(p,s')-G_n(p,s'')|^2] \leq E[|G_n(p,s)-G_n(p,s')|^4]^{1/2}E[|G_n(p,s')-G_n(p,s'')|^4]^{1/2} \leq \bar{\Delta}_1|s-s''|^4$, where the first inequality follows from Cauchy-Schwarz's inequality. This verifies condition (13.14) of Billingsley (1999). The desired result follows from these, theorem 13.5 of Billingsley (1999) and the finite dimensional weak convergence, which obtains by applying the Cramér-Wold device.

(b) Under the given conditions and the null, $E[\sum_{i=1}^{M_n(p)}s^{R_{n,i}(p)}-sp/\{1-s(1-p)\}] = E[\sum_{i=1}^{M_n(p)} E[s^{R_{n,i}(p)}-sp/\{1-s(1-p)\}|M_n(p)]] = E[\sum_{i=1}^{M_n(p)}sp/\{1-s(1-p)\}-sp/\{1-s(1-p)\}] = 0$, where the first equality follows from Lemma A.8, and the second equality follows from the fact that given $M_n(p)$, $R_{n,i}(p)$ is IID under the null.

(c) Under the given the conditions and the null, $E[G_n(p,s)G_n(p,s')] = n^{-1}E[\sum_{i=1}^{M_n(p)}E[[s^{R_{n,i}(p)}-$

$sp/\{1-s(1-p)\}][s'^{R_{n,i}(p)} - s'p/\{1-s'(1-p)\}]|M_n(p)]] = n^{-1}E[M_n(p)[ss'p/\{1-ss'(1-p)\} - ss'p^2/\{1-s(1-p)\}\{1-s'(1-p)\}]] = ss'p^2(1-s)(1-s')(1-p)/[\{1-ss'(1-p)\}\{1-s(1-p)\}\{1-s'(1-p)\}]$, where the first equality follows from Lemma A.8, and the last equality follows because $n^{-1}E[M_n(p)] = p$. Finally, it follows from the continuous mapping theorem that $f[G_n(p,\cdot)] \Rightarrow f[\mathcal{G}(p,\cdot)]$. This is the desired result.

Remark 1: (a) In addition to weak convergence, it also follows that for any $\varepsilon > 0$, there is a $\delta > 0$ such that $\limsup_{n\to\infty} \mathbb{P}(w'''_{G_n(p,\cdot)}(\delta) > \varepsilon) = 0$, where $w'''_{G_n(p,\cdot)}(\delta) := \sup_{s,s'\in\mathbb{S}} \sup_{s''\in\{|s-s'|<\delta\}} \min [\sup_p |G_n(p,s'') - G_n(p,s')|,\ \sup_p |G_n(p,s'') - G_n(p,s)|]$. This follows from the proof of theorem 3 in Bickel and Wichura (1971).

(ii) This can be proved in numerous ways. We verify the conditions of theorem 13.5 of Billingsley (1999). Our proof is separated into three parts: (a), (b), and (c). In (a), we show the weak convergence of $G_n(\cdot, s)$. In (b), we prove that for each p, $E[\mathcal{G}(p,s)] = 0$. Finally, in (c), we show that $E[\mathcal{G}(p,s)\mathcal{G}(p',s)] = s^2p'^2(1-s)^2(1-p)/\{1-s(1-p')\}^2\{1-s^2(1-p)\}$.

(a) First, for each s, we have $\mathcal{G}(1,s) \equiv 0$ as $G_n(1,s) \equiv 0$, and for any $\delta > 0$, $\lim_{p\to1} \mathbb{P}(|\mathcal{G}(p,s)| > \delta) \leq \lim_{p\to1} E(|\mathcal{G}(p,s)|^2)/\delta^2 = \lim_{p\to1} s^2p^2(1-s)^2(1-p)/\delta^2\{1-s(1-p)\}^2\{1-s^2(1-p)\} = 0$ uniformly on $\mathbb{S}$, where the inequality and equality follow from Markov's inequality and the result in (c) respectively. Thus, for each s, $\mathcal{G}(p,s) - \mathcal{G}(1,s) \Rightarrow 0$ as $p \to 1$. Second, it's not hard to show that $E[\{G_n(p,s) - G_n(p',s)\}^4] = E[\{\mathcal{G}(p,s) - \mathcal{G}(p',s)\}^4] - n^{-1}p'^{-1}E[\{\mathcal{G}(p,s) - \mathcal{G}(p',s)\}^4] + n^{-1}p'E[\{\sum_{i=1}^{K_{n,1}}(s^{R_i} - E[s^{R_i}]) - (s^{R'_1} - E[s^{R'_1}])\}^4]$ using the finite dimensional weak convergence result. We examine each piece on the RHS separately. From some tedious algebra, it follows that $E[\{\mathcal{G}(p,s) - \mathcal{G}(p',s)\}^4] = 3s^4(1-s)^4\{k_s(p)m_s(p) - 2k_s(p')m_s(p) + k_s(p')m_s(p')\}^2 \leq 3\{|\{k_s(p) - k_s(p')\}m_s(p)| + |k_s(p')\{m_s(p') - m_s(p)\}|\}^2$, where for each p, $k_s(p) := \frac{p^2}{\{1-s(1-p)\}^2}$, and $m_s(p) := \frac{1-p}{\{1-s^2(1-p)\}}$. Note that $|k_s|$, $|m_s|$, $|k'_s|$ and $|m'_s|$ are bounded by $\Delta_4 := \max[\Delta_1, \Delta_2, \Delta_3]$ uniformly in (p,s). This implies that there exists $\widetilde{\Delta}_2 > 0$ such that if n is sufficiently large enough, then $E[\{\mathcal{G}(p,s) - \mathcal{G}(p',s)\}^4] \leq \widetilde{\Delta}_2|p-p'|^2$. Some algebra implemented using `Mathematica`® shows that for some $\Delta_5 > 0$, $p'E[\{\sum_{i=1}^{K_{n,1}}(s^{R_i} - E[s^{R_i}]) - (s^{R'_1} - E[s^{R'_1}])\}^4] \leq \Delta_5 p'^{-1}|p-p'|$, so that given that $p' \geq \underline{p} > 0$, if n^{-1} is less than $|p-p'|$ then $E[\{G_n(p,s) - G_n(p',s)\}^4] \leq \bar{\Delta}_2|p-p'|^2$ for sufficiently large n, where $\bar{\Delta}_2 := \widetilde{\Delta}_2(1+\underline{p}^{-1}) + \Delta_5\underline{p}^{-1}$. Finally, for each $p'' \leq p' \leq p$, $E[\{G_n(p,s) - G_n(p',s)\}^2\{G_n(p',s) - G_n(p'',s)\}^2] \leq E[|G_n(p,s) - G_n(p',s)|^4]^{1/2}E[|G_n(p',s) - G_n(p'',s)|^4]^{1/2} \leq \bar{\Delta}_2|p-p''|^2$ by the Cauchy-Schwarz inequality. The weak convergence of $\{G_n(\cdot,s)\}$ holds by theorem 13.5 of Billingsley (1999) and finite dimensional weak convergence, which can be obtained by the Cramér-Wold device.

(b) For each p, $E[\mathcal{G}(p,\cdot)] = 0$ follows from the proof of Theorem 1(i, b).

(c) First, for convenience, for each p and p', we let M and M' denote $M_n(p)$ and $M_n(p')$ respectively, and let R_i and R_i' stand for $R_{n,i}(p)$ and $R_{n,i}(p')$. Then from the definition of $K_{n,j}$, $E[G_n(p,s)G_n(p',s)] = n^{-1}E[M'E[\sum_{i=1}^{K_{n,1}}(s^{R_i} - E[s^{R_i}])(s^{R_1'} - E[s^{R_1'}])|M, M']] = n^{-1}E[M']E[\sum_{i=1}^{K_{n,1}}(s^{R_i} - E[s^{R_i}])(s^{R_1'} - E[s^{R_1'}])] = p'E[\sum_{i=1}^{K_{n,1}} s^{R_i+R_1'} - \sum_{i=1}^{K_{n,1}} s^{R_i}E[s^{R_1'}] - K_{n,1}s^{R_1'}E[s^{R_i}] + K_{n,1}E[s^{R_i}]E[s^{R_1'}]]$, where the first equality follows from Lemma A8 since $\{K_{n,j}\}$ is IID under the null and R_j' is independent of R_ℓ, if $\ell \leq K_{n,j-1}$ or $\ell \geq K_{n,j}+1$. The second equality follows, as $\{M, M'\}$ is independent of $\{R_i, R_1 : i = 1, 2, ..., K_{n,1}\}$. Further, $E[\sum_{i=1}^{K_{nm1}} s^{R_i}] = E[s^{R_i}] \cdot E[K_{n,1}]$, $E[K_{n,1}s^{R_1'}] = sp'/\{1 - s(1-p')\} \cdot \{1 - s(1-p)\}/\{1 - s(1-p')\}$, and $E[\sum_{i=1}^{K_{n,1}} s^{R_i+R_1'}] = s^2p'/\{1 - s^2(1-p)\} \cdot [\{1 - s(1-p)\}/\{1- s(1-p')\}]^2$ by Lemmas A1 to A4. Substituting these into the above equation yields the desired result.

Remark 2: (*a*) In the proof of Theorem 1(ii-a), $\bar{\Delta}_2$ is given by $\{\widetilde{\Delta}_2(1 + \underline{p}^{-1}) + \Delta_5\underline{p}^{-1}\}$, which involves $\underline{p}$. Unless p is bounded away from zero, $\{\widetilde{\Delta}_2(1 + p^{-1}) + \Delta_5 p^{-1}\}$ is not bounded uniformly in p.

(*b*) The proof of Theorem 1(ii) also implies that for any $\varepsilon > 0$, there is a $\delta > 0$ such that $\limsup_{n\to\infty} \mathbb{P}(w''_{G_n(\,\cdot\,,s)}(\delta) > \varepsilon) = 0$, where $w''_{G_n(\,\cdot\,,s)}(\delta) := \sup_{p,p'\in\mathbb{J}} \sup_{p''\in\{|p-p'|<\delta\}} \min[|G_n(p'',s) - G_n(p',s)|, |G_n(p'',s) - G_n(p,s)|]$, implying that $\{G_n(\,\cdot\,,s)\}$ is tight. A proof is given in Billingsley (1999, pp. 141–143).

(iii) We separate the proof into two parts, (a) and (b). In (a), we prove the weak convergence of G_n, and in (b) we derive its covariance structure.

(a) In order to show the weak convergence of G_n, we exploit the moment condition in theorem 3 of Bickel and Wichura (1971, p. 1665). For this, we first let B and C be neighbors in $\mathbb{J} \times \mathbb{S}$ such that $B := (p_1, p_2] \times (s_1, s_2]$ and $C := (p_1, p_2] \times (s_2, s_3]$. Without loss of generality, we suppose that $|s_2 - s_1| \leq |s_3 - s_2|$. Second, we define $|G_n(B)| := |G_n(p_1, s_1) - G_n(p_1, s_2) - G_n(p_2, s_1) + G_n(p_2, s_2)|$, then $|G_n(B)| \leq |G_n(p_1, s_1) - G_n(p_1, s_2)| + |G_n(p_2, s_2) - G_n(p_2, s_1)|$, so that $E[|G_n(B)|^4] = E[|A_1|^4] + 4E[|A_1|^3|A_2|] + 6E[|A_1|^2|A_2|^2] + 4E[|A_1||A_2|^3] + E[|A_2|^4] \leq E[|A_1|^4] + 4E[|A_1|^4]^{3/4}E[|A_2|^4]^{1/4} + 6E[|A_1|^4]^{2/4}E[|A_2|^4]^{2/4} + 4E[|A_1|^4]^{1/4}E[|A_2|^4]^{3/4} + E[|A_2|^4]$ using Hölder's inequality, where we let $A_1 := |G_n(p_1, s_1) - G_n(p_1, s_2)|$ and $A_2 := |G_n(p_2, s_2) - G_n(p_2, s_1)|$ for notational simplicity. We already saw that $E[|A_1|^4] \leq \bar{\Delta}_1|s_1 - s_2|^4$ and $E[|A_2|^4] \leq \bar{\Delta}_1|s_1 - s_2|^4$ in the proof of Theorem 1(i). Thus, $E[|G_n(B)|^4] \leq 16\bar{\Delta}_1|s_1 - s_2|^4$. Third, we define $|G_n(C)| := |G_n(p_2, s_2) - G_n(p_2, s_3) - G_n(p_3, s_2) + G_n(p_3, s_3)|$; then $|G_n(C)| \leq |G_n(p_2, s_2) - G_n(p_3, s_2)| + |G_n(p_3, s_3) - G_n(p_2, s_3)|$. Using the same logic as above, Hölder's inequality, and the result in the proof of Theorem 1(ii), we obtain $E[|G_n(C)|^4] \leq 16\bar{\Delta}_2|p_2 - p_1|^2$ for sufficiently large n. Fourth, therefore, using Hölder's inequality, we obtain that for all sufficiently large n, $E[|B|^{4/3}|C|^{8/3}] \leq E[|B|^4]^{1/3}E[|C|^4]^{2/3} \leq \bar{\Delta}\{|s_2 - s_1|^2 \cdot |p_2 - p_1|^2\}^{2/3} \leq \bar{\Delta}\{|s_2 - s_1| \cdot |p_2 - p_1|\}^{2/3}\{|s_3 - s_2| \cdot |p_2 - p_1|\}^{2/3} = \{\bar{\Delta}^{3/4}\lambda(B)\}^{2/3}\{\bar{\Delta}^{3/4}\lambda(C)\}^{2/3}$, where $\bar{\Delta} := 16\bar{\Delta}_1^{1/3}\bar{\Delta}_2^{2/3}$, and $\lambda(\,\cdot\,)$ denotes the Lebesgue measure of the given argument. This verifies the moment condition (3) in theorem 3 of Bickel and Wichura (1971, p. 1665). Fifth, it trivially holds from the definition of G_n that

$\mathcal{G} = 0$ on $\{(p, s) \in \mathbb{J} \times \mathbb{S} : s = 0\}$. Finally, the continuity of $\mathcal{G}$ on the edge of $\mathbb{J} \times \mathbb{S}$ was verified in the proof of Theorem 1(ii). Therefore, the weak convergence of $\{G_n\}$ follows from the corollary in Bickel and Wichura (1971, p. 1664) and the finite dimensional weak convergence obtained by Lindeberg-Levy's CLT and the Cramér-Wold device.

(b) As before, for convenience, for each p and p', we let M and M' denote $M_n(p)$ and $M_n(p')$ respectively, and we let R_i and R_i' be short-hand notations for $R_{n,i}(p)$ and $R_{n,i}(p')$. Also, we let $K_{n,j}$ be as previously defined. Then, under the given conditions and the null,

$$E[G_n(p,s)G_n(p',s')] = n^{-1}E[\sum_{j=1}^{M'}\sum_{i=1}^{M}(s^{R_i} - E[s^{R_i}])({s'}^{R_j'} - E[{s'}^{R_j'}])] \tag{18}$$

$$= p'E[\sum_{i=1}^{K_{n,1}} s^{R_i}{s'}^{R_1'} - \sum_{i=1}^{K_{n,1}} E[{s'}^{R_1'}]s^{R_i} - K_{n,1}{s'}^{R_1'}E[s^{R_i}] + K_{n,1}E[s^{R_i}]E[{s'}^{R_1'}]],$$

where the first equality follows from the definition of G_n, and the second equality holds for the same reason as in the proof of Theorem 1(ii). From Lemmas A1 to A4, we have that $E[\sum_{i=1}^{K_{n,1}} s^{R_i}] = E[s^{R_i}] \cdot E[K_{n,1}]$, and $E[K_{n,1}{s'}^{R_1'}] = s'p'/\{1 - s'(1 - p')\} \cdot \{1 - s'(1 - p)\}/\{1 - s'(1 - p')\}$, and $E[\sum_{i=1}^{K_{n,1}} s^{R_i}{s'}^{R_1'}] = \{s\, s'p'\}/\{1 - ss'(1 - p)\} \cdot [\{1 - s'(1 - p)\}/\{1 - s'(1 - p')\}]^2$. Thus, substituting these into (18) gives $E[G_n(p,s)G_n(p',s')] = ss'p'^2(1 - s)(1 - s')(1 - p)\{1 - s'(1 - p)\}/[\{1-s(1-p)\}\{1-s'(1-p')\}^2\{1-ss'(1 - p)\}]$. Finally, it follows from the continuous mapping theorem that $f[G_n] \Rightarrow f[\mathcal{G}]$. ■

Proof of Lemma 2: (i) First, $\sup_{p\in\mathbb{I}} |\widetilde{p}_n(p) - p| \to 0$ almost surely by Glivenko-Cantelli. Second, $G_n \Rightarrow \mathcal{G}$ by Theorem 1(ii). Third, $(\mathcal{D}(\mathbb{J} \times \mathbb{S}) \times \mathcal{D}(\mathbb{J}))$ is a separable space. Thus, $(G_n, \widetilde{p}_n(\,\cdot\,)) \Rightarrow (\mathcal{G}, \,\cdot\,)$ by theorem 3.9 of Billingsley(1999). Fourth, $|\mathcal{G}(p, s) - \mathcal{G}(p', s')| \leq |\mathcal{G}(p, s) - \mathcal{G}(p', s)| + |\mathcal{G}(p', s) - \mathcal{G}(p', s')|$, and each term of the RHS can be made as small as desired by letting $|p - p'|$ and $|s - s'|$ tend to zero, as $\mathcal{G} \in \mathcal{C}(\mathbb{J} \times \mathbb{S})$ a.s. Finally, note that for each (p, s), $W_n(p, s) = G_n(\widetilde{p}_n(p), s)$. Therefore, $W_n - G_n = G_n(\widetilde{p}_n(\,\cdot\,), \,\cdot\,) - G_n(\,\cdot\,, \,\cdot\,) \Rightarrow \mathcal{G} - \mathcal{G} = 0$ by a lemma of Billingsley (1999, p. 151) and the four facts just shown. This implies that $\sup_{(p,s)\in\mathbb{J}\times\mathbb{S}} |W_n(p, s) - G_n(p, s)| \to 0$ in probability, as desired.

(ii) We write $\widetilde{p}_n(p)$ as $\widetilde{p}_n$ for convenience. By the mean value theorem, for some $p_n^*(p)$ (in $\mathbb{I}$) between $p \in \mathbb{J}$ and $\widetilde{p}_n$, $|[\{s\widetilde{p}_n\}/\{1 - s(1 - \widetilde{p}_n)\} - sp/\{1 - s(1 - p)\}] - \{s(1 - s)(\widetilde{p}_n - p)\}/\{1 - s(1 - p)\}^2| = 2s^2(1 - s)(\widetilde{p}_n - p)^2/\{1 - s(1 - p_n^*(p))\}^3$, where $\sup_{p\in\mathbb{J}} |p_n^*(p) - p| \to 0$ a.s. by Glivenko-Cantelli. Also,

$$\sup_{p,s} \frac{\widetilde{M}_n(p)s^2(1 - s)(\widetilde{p}_n - p)^2}{\sqrt{n}\{1 - s(1 - p_n^*(p))\}^3} \leq \frac{1}{\sqrt{n}}\left|\sup_{p} \frac{\widetilde{M}_n(p)}{n}\right| \left|\sup_{p,s} \frac{s^2(1 - s)}{\{1 - s(1 - p_n^*(p))\}^3}\right| \left|\sup_{p} n(\widetilde{p}_n - p)^2\right|,$$

where $n^{-1}\widetilde{M}_n(p)$ and $s^2(1 - s)\{1 - s(1 - p_n^*(p))\}^{-3}$ are uniformly bounded by 1 and $1/(1 - \widetilde{s})^3$ respectively, with $\widetilde{s} := \max[|\underline{s}|, \bar{s}]$; and $n(\widetilde{p}_n - p)^2 = O_{\mathbb{P}}(1)$ uniformly in p. Thus,

$$\sup_{p,s} \frac{\widetilde{M}_n(p)}{\sqrt{n}} \left| \left[\frac{s\widetilde{p}_n}{\{1 - s(1 - \widetilde{p}_n)\}} - \frac{sp}{\{1 - s(1 - p)\}}\right] - \frac{s(1 - s)\,(\widetilde{p}_n - p)}{\{1 - s(1 - p)\}^2}\right| = o_{\mathbb{P}}(1).$$

Given these, the weak convergence of H_n follows immediately, as $\sup_p |n^{-1}\widetilde{M}_n(p) - p| = o_{\mathbb{P}}(1)$, and the function of p defined by $\sqrt{n}(\widetilde{p}_n - p)$ weakly converges to a Brownian bridge, permitting application of the lemma of Billingsley (1999, p. 151). These facts also suffice for the tightness of H_n.

Next, the covariance structure of $\mathcal{H}$ follows from the fact that for each (p, s) and (p', s') with $p' \leq p$, $E[\{\widetilde{M}_n(p)s(1-s)(\widetilde{p}_n - p)\}/\sqrt{n}\{1 - s(1-p)\}^2] = 0$, and

$$E\left[\frac{\widetilde{M}_n(p)s(1-s)\,(\widetilde{p}_n - p)}{\sqrt{n}\{1-s(1-p)\}^2}\frac{\widetilde{M}_n(p')s'(1-s')(\widetilde{p}_n' - p')}{\sqrt{n}\{1-s'(1-p')\}^2}\right] = \frac{ss'pp'^2(1-s)(1-s')(1-p)}{\{1-s(1-p)\}^2\{1-s'(1-p')\}^2},$$

which is identical to $E[\mathcal{H}(p,s)\mathcal{H}(p',s')]$.

(iii) To show the given claim, we first derive the given covariance structure. For each (p, s) and (p', s'), $E[W_n(p,s)H_n(p',s')] = E[E[W_n(p,s)|X_1, ..., X_n]H_n(p',s')]$, where the equality follows because H_n is measurable with respect to the smallest σ–algebra generated by $\{X_1, ..., X_n\}$. Given this, we have $E[W_n(p,s)|X_1,...,X_n] = n^{-1/2}\sum_{i=1}^{\widetilde{M}_n(p)}[E[s^{\widetilde{R}_{n,i}(p)}|X_1,...,X_n] - s\widetilde{p}_n/\{1-s(1-\widetilde{p}_n)\}] = n^{-1/2}\sum_{i=1}^{\widetilde{M}_n(p)}[sp/\,\{1-s(1-p)\} - s\widetilde{p}_n/\{1-s(1-\widetilde{p}_n)\}] = -H_n(p,s)$. Thus, $E[E[W_n(p,s)|X_1, ..., X_n]H_n(p',s')] = -\,E[H_n(p,s)H_n(p',s')]$. Next, we have that $E[\mathcal{G}(p,s)\mathcal{H}(p',s')] = \lim_{n\to\infty} E[W_n(p, s)H_n(p',s')]$ by Lemma 2(i). Further, $E[\mathcal{H}(p,s)\mathcal{H}(p',s')] = \lim_{n\to\infty} E[H_n(p,s)H_n(p',s')]$. It follows that $E[\mathcal{G}(p,s)\mathcal{H}(p',s')] = -E[\mathcal{H}(p,s)\mathcal{H}(p',s')]$.

Next, we consider $(\widetilde{G}_n, H_n)'$ and apply example 1.4.6 of van der Vaart and Wellner (1996, p. 31) to show weak convergence. Note that $\widetilde{G}_n = W_n + H_n = G_n + H_n + o_{\mathbb{P}}(1)$, and that G_n and H_n are each tight, so $\widetilde{G}_n$ is tight, too. Further, G_n and H_n have continuous limits by Theorem 1(ii) and Lemma 2(ii). Thus, if the finite-dimensional distributions of $G_n + H_n$ have weak limits, then $\widetilde{G}_n$ must weakly converge to the Gaussian process $\widetilde{\mathcal{G}}$ with the covariance structure (8). We may apply the Lindeberg-Levy CLT to show this unless $G_n + H_n \equiv 0$ almost surely. That is, for each (p, s) with $s \neq 0$, $E[G_n(p,s) + H_n(p,s)] = 0$, and $E[\{G_n(p,s) + H_n(p,s)\}^2] = E[G_n(p,s)^2] - E[H_n(p,s)^2] + o(1) \leq E[G_n(p,s)^2] + o(1) = s^2p^2(1-s)^2(1-p)/\,\{1-s^2(1-p)\}\{1-s(1-p)\}^2 + o(1)$, which is uniformly bounded, so that for each (p, s) with $s \neq 0$, the sufficiency conditions for the Lindeberg-Levy CLT hold. The first equality above follows by applying Lemma 2(ii). If $s = 0$, then $G_n(\,\cdot, 0) + H_n(\,\cdot\,, 0) \equiv 0$, so that the probability limit of $G_n(\,\cdot, 0) + H_n(\,\cdot\,, 0)$ is zero. Given these, the finite-dimensional weak convergence of $\widetilde{G}_n$ now follows from the Cramér-Wold device.

Next, note that $\widetilde{G}_n$ is asymptotically independent of H_n because $E[\widetilde{G}_n(p,s)H_n(p',s')] = E[W_n(p,s) H_n(p',s')] + E[H_n(p,s)H_n(p',s')] = 0$ by the covariance structure given above. It follows that $(\widetilde{G}_n, H_n)' \Rightarrow (\widetilde{\mathcal{G}}, \mathcal{H})'$ by example 1.4.6 of van der Vaart and Wellner (1996, p. 31). To complete the proof, take $(\widetilde{G}_n - H_n, H_n)' = (W_n, H_n)'$, and apply the continuous mapping theorem. ■

Remark 3: Durbin (1973) shows that empirical distributions with parameter estimation error are not distribution-free using a proof similar to that of Lemma 2(i). We further exploit his proof to show that here W_n is asymptotically equivalent to G_n.

Proof of Theorem 2: (i, ii, *and* iii) The proof of Lemma 2(iii) establishes that $\widetilde{G}_n \Rightarrow \widetilde{\mathcal{G}}$. This and the continuous mapping theorem imply the given claims. ■

Proof of Lemma 3: (i) First, as shown in the proof of Lemma 3(ii) below, $\hat{F}_n(y(\,\cdot\,))$ converges to $F(y(\,\cdot\,))$ in probability uniformly on $\mathbb{I}$, where for each p, $y(p) := \inf\{x \in \mathbb{R} : F(x) \geq p\}$. Second, $G_n \Rightarrow \mathcal{G}$ by Theorem 1(ii). Third, $(\mathcal{D}(\mathbb{J}\times\mathbb{S})\times\mathcal{D}(\mathbb{J}))$ is a separable space. Therefore, it follows that $(G_n, \hat{F}_n(y(\,\cdot\,))) \Rightarrow (\mathcal{G}, F(y(\,\cdot\,)))$ by theorem 3.9 of Billingsley (1999). Fourth, $\mathcal{G} \in \mathcal{C}(\mathbb{J} \times \mathbb{S})$. Finally, $\ddot{G}_n(\,\cdot\,,\,\cdot\,) = G_n(\hat{F}_n(y(\,\cdot\,)), \cdot\,)$, so that $\ddot{G}_n(\,\cdot\,,\,\cdot\,) - G_n(\,\cdot\,,\,\cdot\,) = G_n(\hat{F}_n(y(\,\cdot\,)), \cdot\,) - G_n(\,\cdot\,,\,\cdot\,) \Rightarrow \mathcal{G} - \mathcal{G} = 0$, where the weak convergence follows from the lemma of Billingsley (1999, p. 151). Thus, $\sup_{(p,s)} |\ddot{G}_n(p,s) - G_n(p,s)| = o_{\mathbb{P}}(1)$.

(ii) First, the definition of $\ddot{H}_n(p,s)$ permits the representation $\ddot{H}_n(p,s) = \{\hat{M}_n(p)/n\}\{\sqrt{n}[s\widetilde{p}_n/\{1-s(1-\widetilde{p}_n)\} - sp/\{1-s(1-p)\}]\}$. Second, it follows that $\sqrt{n}[s\widetilde{p}_n/\{1-s(1-\widetilde{p}_n)\} - sp/\{1-s(1-p)\}] \Rightarrow -s(1-s)\mathcal{B}_0^0(p)/\{1-s(1-p)\}^2$ by Lemma 2(ii) and Theorem 5(i) below. Third, if $\hat{F}_n(\,\cdot\,)$ converges to $F(\,\cdot\,)$ in probability, then $n^{-1}\hat{M}_n(p)$ converges to p in probability uniformly in p, because for each p, $\hat{M}_n(p)$ is defined as $\sum_{t=1}^n \mathbf{1}_{\{\hat{F}_n(\hat{Y}_t)<p\}}$. Finally, these facts imply that $\sup_{(p,s)} |\ddot{H}_n(p,s) - H_n(p,s)| = o_{\mathbb{P}}(1)$ by the lemma of Billingsley (1999, p. 151); this completes the proof.

Therefore, we only have to show that $\hat{F}_n(\,\cdot\,)$ converges to $F(\,\cdot\,)$ in probability; for this we exploit Glivenko-Cantelli. That is, if for each p, $\hat{F}_n(y(p))$ converges to $F(y(p))$ in probability, then the uniform convergence follows from the properties of empirical distribution: boundedness, monotonicity, and right continuity. Thus, the pointwise convergence of $\hat{F}_n(p)$ completes the proof. We proceed as follows. First, letting $y = y(p)$ for notational simplicity, for each y and for any $\varepsilon_1 > 0$, we have $\{\omega \in \Omega : h_t(\hat{\boldsymbol{\theta}}_n) < y\} \subset \{\omega \in \Omega : h_t(\boldsymbol{\theta}_*) < y + |h_t(\hat{\boldsymbol{\theta}}_n) - h_t(\boldsymbol{\theta}_*)|\} = \{\omega \in \Omega : h_t(\boldsymbol{\theta}_*) < y + |h_t(\hat{\boldsymbol{\theta}}_n) - h_t(\boldsymbol{\theta}_*)|, |h_t(\hat{\boldsymbol{\theta}}_n) - h_t(\boldsymbol{\theta}_*)| < \varepsilon_1\} \cup \{\omega \in \Omega : h_t(\boldsymbol{\theta}_*) < y + |h_t(\hat{\boldsymbol{\theta}}_n) - h_t(\boldsymbol{\theta}_*)|, |h_t(\hat{\boldsymbol{\theta}}_n) - h_t(\boldsymbol{\theta}_*)| \geq \varepsilon_1\} \subset \{\omega \in \Omega : h_t(\boldsymbol{\theta}_*) < y + \varepsilon_1\} \cup \{|h_t(\hat{\boldsymbol{\theta}}_n) - h_t(\boldsymbol{\theta}_*)| \geq \varepsilon_1\}$. Second, for the same y and ε_1, $\{\omega \in \Omega : h_t(\hat{\boldsymbol{\theta}}_n) < y\} \supset \{\omega \in \Omega : h_t(\boldsymbol{\theta}_*) < y - \varepsilon_1\} \setminus \{\omega \in \Omega : |h_t(\boldsymbol{\theta}_*) - h_t(\hat{\boldsymbol{\theta}}_n)| > \varepsilon_1\}$. These two facts imply that $n^{-1}\sum_{t=1}^n \mathbf{1}_{\{h_t(\boldsymbol{\theta}_*)<y-\varepsilon_1\}} - n^{-1}\sum_{t=1}^n \mathbf{1}_{\{|h_t(\hat{\boldsymbol{\theta}}_n)-h_t(\boldsymbol{\theta}_*)|\geq\varepsilon_1\}} \leq n^{-1}\sum_{t=1}^n \mathbf{1}_{\{h_t(\hat{\boldsymbol{\theta}}_n)<y\}} \leq n^{-1}\sum_{t=1}^n \mathbf{1}_{\{h_t(\boldsymbol{\theta}_*)<y+\varepsilon_1\}} + n^{-1}\sum_{t=1}^n \mathbf{1}_{\{|h_t(\hat{\boldsymbol{\theta}}_n)-h_t(\boldsymbol{\theta}_*)|\geq\varepsilon_1\}}$. Thus, it follows that $n^{-1}\sum_{t=1}^n \mathbf{1}_{\{h_t(\boldsymbol{\theta}_*)<y-\varepsilon_1\}}$ and $n^{-1}\sum_{t=1}^n \mathbf{1}_{\{h_t(\boldsymbol{\theta}_*)<y+\varepsilon_1\}}$ converge to $F(y-\varepsilon_1)$ and $F(y+\varepsilon_1)$ a.s. by the SLLN and the null hypothesis. Further, for any $\delta > 0$ and $\varepsilon_2 > 0$, there is an n^* such that if $n > n^*$, $\mathbb{P}(n^{-1}\sum_{t=1}^n \mathbf{1}_{\{|h_t(\boldsymbol{\theta}_*)-h_t(\hat{\boldsymbol{\theta}}_n)|>\varepsilon_1\}} \geq \delta) \leq \varepsilon_2$. This follows because $\mathbb{P}(n^{-1}\sum_{t=1}^n \mathbf{1}_{\{|h_t(\boldsymbol{\theta}_*)-h_t(\hat{\boldsymbol{\theta}}_n)|>\varepsilon_1\}} \leq \delta) \leq (n\delta)^{-1}\sum_{t=1}^n E(\mathbf{1}_{\{|h_t(\boldsymbol{\theta}_*)-h_t(\hat{\boldsymbol{\theta}}_n)|>\varepsilon_1\}}) = (\delta n)^{-1}\sum_{t=1}^n \mathbb{P}(|h_t(\boldsymbol{\theta}_*) - h_t(\hat{\boldsymbol{\theta}}_n)| > \varepsilon_1) \leq \varepsilon_2$, where the first inequality follows from Markov's inequality, and the last inequality follows from the fact that $|h_t(\boldsymbol{\theta}_*) - h_t(\hat{\boldsymbol{\theta}}_n)| \leq M_t\|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*\| = o_{\mathbb{P}}(1)$ uniformly in t by A2 and A3(ii). It follows that for any $\varepsilon_1 > 0$, $F(y - \varepsilon_1) + o_{\mathbb{P}}(1) \leq \hat{F}_n(y) \leq F(y + \varepsilon_1) + o_{\mathbb{P}}(1)$. As ε_1 may be chosen arbitrarily small, it follows that $\hat{F}_n(y)$ converges to $F(y)$ in probability as desired. ■

Proof of Theorem 3: (i, ii, *and* iii) $\hat{G}_n = \ddot{G}_n + \ddot{H}_n = G_n + H_n + o_{\mathbb{P}}(1)$ by Lemmas 3(i) and 3(ii). Further, $G_n + H_n = \widetilde{G}_n \Rightarrow \widetilde{\mathcal{G}}$ by Theorem 2(ii). Thus, $\hat{G}_n \Rightarrow \widetilde{\mathcal{G}}$, which, together with the continuous mapping theorem, implies the desired result. ■

The following Lemmas collect together further supplementary claims needed to prove the weak convergence of the EGR test statistics under the local alternative. As before, we use the notation $p = F(y)$ for brevity and suppose that $R_{n,i}(p)$ is defined by observations starting from $Y_{n,t+1}$, unless otherwise noted.

LEMMA B1: *Given conditions A1, A2*(i)*, A3, A5, and* $\mathbb{H}_1^\ell$,

(i) *for each* y *and* $k = 2, 3, ...$, $E[\mathbf{1}_{\{Y_{n,t+k}<y\}}|\mathcal{F}_{n,t}] = F(y) + \sum_{j=1}^{k-1} n^{-\frac{j}{2}}Q_j(y) + n^{-k/2}G_k(y, Y_{n,t})$, *where for each* $j = 1, 2, ...$, $Q_j(y) := \int\int ... \int D(y, x_1)dD(x_1, x_2)...dD(x_{j-1}, x)dF(x)$, *and* $G_k(y, Y_{n,t})$ $:= \int\int ... \int D(y, x_1)dD(x_1, x_2)...dD(x_{k-2}, x_{k-1})dD(x_{k-1}, Y_{n,t})$;

(ii) *for each* y, $F_n(y) = F(y) + \sum_{j=1}^{\infty} n^{-j/2}Q_j(y)$ *and* $Q_1(y) = Q(y)$;

(iii) *for each* y *and* $k = 1, 2, ...$, $E[\widetilde{J}_{n,t+k}(y)\widetilde{J}_{n,t}(y)] = O(n^{-k/2})$, *where* $\widetilde{J}_{n,t}(y) := \mathbf{1}_{\{Y_{n,t}<y\}} - F_n(y)$;

(iv) *for each* y, $E[J_{n,t+1}(y)J_{n,t}(y)] = O(n^{-1/2})$, $E[J_{n,t+2}(y)J_{n,t}(y)] = O(n^{-1})$, *and for* $k = 3, 4, ...$, $E[J_{n,t+k}(y)J_{n,t}(y)] = O(n^{-3/2})$, *where* $J_{n,t}(y) := \mathbf{1}_{\{Y_{n,t}<y\}} - F(y) - n^{-1/2}Q(y)$.

Proof of Lemma B1: (i) This follows by applying the law of iterated expectations sequentially. First, note that $E[E[\mathbf{1}_{\{Y_{n,t+2}<y\}}|Y_{n,t+1}]|Y_{n,t}] = \int F(y) + n^{-1/2}D(y, x)dF_n(x|Y_{n,t}) = F(y) + n^{-1/2}Q_1(y) + n^{-1}\int D(y, x)dD(x, Y_{n,t})$, where the second equality follows by (13). Now consider the general case $k - 1$; from the given hypothesis, $E[\mathbf{1}_{\{Y_{n,t+k-1}<y\}}|\mathcal{F}_{n,t}] = F(y) + \sum_{j=1}^{k-2} n^{-\frac{j}{2}}Q_j(y) + n^{-(k-1)/2}G_{k-1}(y, Y_{n,t})$. Then, by the stationarity of $\{Y_{n,t}\}$, $E[\mathbf{1}_{\{Y_{n,t+k}<y\}}|\mathcal{F}_{n,t+1}] = F(y) + \sum_{j=1}^{k-2} n^{-\frac{j}{2}}Q_j(y) + n^{-(k-1)/2}G_{k-1}$ $(y, Y_{n,t+1})$. Thus, $E[\mathbf{1}_{\{Y_{n,t+k}<y\}}|\mathcal{F}_{n,t}] = F(y) + \sum_{j=1}^{k-2} n^{-\frac{j}{2}}Q_j(y) + n^{-(k-1)/2}\int G_{k-1}(y, z)dF_n(z|Y_{n,t})$ $= F(y) + \sum_{j=1}^{k-1} n^{-\frac{j}{2}}Q_j(y) + n^{-k/2}\int G_{k-1}(y, z)dD(z, Y_{n,t})$, using (13). Note that $\int G_{k-1}(y, z)dD(z,$ $Y_{n,t}) = G_k(y, Y_{n,t})$, yielding the desired result.

(ii) By the geometric ergodicity and strict stationarity assumptions, $\{Y_{n,t}\}$ is an aperiodic Harris recurrent Markov process, so that $\lim_{k\to\infty} \|E[\mathbf{1}_{\{Y_{n,t+k}<\cdot\}}|\mathcal{F}_{n,t}] - F_n(\,\cdot\,)\|_{TV} = 0$ by theorem 6.8 of Durrett (1996, p. 332), where $\|\cdot\|_{TV}$ denotes the total variation. Given the continuity assumption of F and D, this implies that for each y, $\lim_{k\to\infty} E[\mathbf{1}_{\{Y_{n,t+k}<y\}}|\mathcal{F}_{n,t}] - F_n(y) = 0$. Further, $\lim_{k\to\infty} E[\mathbf{1}_{\{Y_{n,t+k}<y\}}|\mathcal{F}_{n,t}] =$

$F(y) + \sum_{j=1}^{\infty} n^{-j/2} Q_j(y)$ by B1(i), so that $F_n(y) = F(y) + \sum_{j=1}^{\infty} n^{-j/2} Q_j(y)$. Finally, the coefficient for the term of order $n^{-1/2}$ is $\int D(y,x)dF(x)$, which therefore must be $Q(y)$.

(iii) By the definition of $\widetilde{J}_{n,t}(y)$ and Lemmas B1(i and ii), $E[\widetilde{J}_{n,t+k}(y)|\mathcal{F}_{n,t}] = n^{-k/2}\{G_k(y, Y_{n,t}) - Q_k(y)\} + o_{\mathbb{P}}(n^{-k/2})$. Therefore, $E[\widetilde{J}_{n,t+k}(y)\widetilde{J}_{n,t}(y)] = E[E[\widetilde{J}_{n,t+k}(y)|\mathcal{F}_{n,t}]\widetilde{J}_{n,t}(y)] = n^{-k/2}E[\{G_k(y, Y_{n,t}) - Q_k(y)\}\widetilde{J}_{n,t}(y)] + o(n^{-k/2}) = O(n^{-k/2})$.

(iv) From Lemma B1(i), it follows that $E[J_{n,t+1}(y)J_{n,t}(y)] = E[E[J_{n,t+1}(y)|\mathcal{F}_{n,t}]J_{n,t}(y)] = n^{-1/2} E[\{D(y, Y_{n,t}) - Q(y)\}J_{n,t}(y)] = O(n^{-1/2})$; $E[J_{n,t+2}(y)J_{n,t}(y)] = E[E[J_{n,t+2}(y)|\mathcal{F}_{n,t}]\ J_{n,t}(y)] = n^{-1}E[\int D(y,x_1)dD(x_1, Y_{n,t})J_{n,t}(y)] = O(n^{-1})$; and for $k = 3, 4, \ldots$, $E[J_{n,t+k}(y)|\mathcal{F}_{n,t}] = O_{\mathbb{P}}(n^{-3/2})$, so that $E[J_{n,t+k}(y)J_{n,t}(y)] = E[E[J_{n,t+k}(y)|\mathcal{F}_{n,t}]J_{n,t}(y)] = O(n^{-3/2})$. This is the desired result. ■

Remark 4: (a) When the range of integration is not explicitly specified in the proof of above, it should be understood that the range is from $-\infty$ to ∞.

(b) The strictly stationary and geometric ergodic Markov process assumption has a number of implications important for us. They can be summarized as follows:

1. Nummelin and Tweedie (1978) show that there is a positively valued measurable function K such that $\|F_n(\cdot|\mathcal{F}_{n,t-k}) - F_n(\cdot)\|_{TV} \leq n^{-k/2}K(Y_{n,t-k})$, where for each y, $F_n(y|\mathcal{F}_{n,t-k})$ denotes $\mathbb{P}(Y_{n,t} \leq y|\mathcal{F}_{n,t-k})$, and $||\cdot||_{TV}$.

2. Nummelin and Touminen (1982) elaborate this further and show that $K(Y_{n,t})$ is integrable. This implies that the β-mixing coefficient of $\{Y_{n,t}\}$ converges to zero geometrically. That is, $\beta_{n,k} \leq n^{-k/2}\int K(y)dF_n(y)$.

LEMMA B2: *Given conditions A1, A2(i), A3, A5, and $\mathbb{H}_1^{\ell}$, for each y,* $n^{1/2}\{\widetilde{F}_n(y) - F(y)\} \overset{\text{A}}{\sim} N(Q(y), F(y)\{1 - F(y)\})$.

Proof of Lemma B2: To show the given claim, we show that $n^{-1/2}\sum_{t=1}^{n}\widetilde{J}_{n,t}(y) \overset{\text{A}}{\sim} N(0, F(y)\{1 - F(y)\})$. This is equivalent to proving Lemma B2 because $J_{n,t}(y) - \widetilde{J}_{n,t}(y) = O_{\mathbb{P}}(n^{-1})$, from which it follows that $n^{-1/2}\sum_{t=1}^{n} J_{n,t}(y) = n^{-1/2}\sum_{t=1}^{n}\widetilde{J}_{n,t}(y) + o_{\mathbb{P}}(1)$. The asymptotic normality can be proved by theorem 5.3 of Ango Nze and Doukhan (2004). First, $E[\widetilde{J}_{n,t}(y)] = 0$ from the definition of $\widetilde{J}_{n,t}(y)$. Second, the asymptotic variance of $n^{-1/2}\sum_{t=1}^{n}\widetilde{J}_{n,t}(y)$ is obtained by Lemma B1. That is, $\bar{\sigma}_n^2 :=$ var$\{n^{-1/2}\sum_{t=1}^{n}\widetilde{J}_{n,t}(y)\} = E[\widetilde{J}_{n,t}(y)^2] + n^{-1}\sum\sum_{t\neq\tau}^{n} E[\widetilde{J}_{n,t}(y)\widetilde{J}_{n,\tau}(y)]$, where $E[\widetilde{J}_{n,t}(y)^2] = F(y)\{1 - F(y)\} + o(1)$ by the definition of $\widetilde{J}_{n,t}(y)$ and $n^{-1}\sum\sum_{t\neq\tau}^{n} E[\widetilde{J}_{n,t}(y)\widetilde{J}_{n,\tau}(y)] = 2n^{-1}\sum_{k=1}^{n-1}(n - k)E[\widetilde{J}_{n,t}(y)\widetilde{J}_{n,t+k}(y)] = o(1)$ by Lemma B1(iii). Therefore, $\bar{\sigma}_n^2 = F(y)\{1 - F(y)\} + o(1)$. Finally, $\sum_{k=1}^{\infty}\beta_{n,k} < \infty$ by Remark 4(b−2). Given this, theorem 5.3 of Ango Nze and Doukhan (2004) completes the proof. ■

Remark 5: Even if Q is identical to zero, the EGR test statistics can still have local power, arising from the estimation of the empirical distribution function. The example given below Theorem 6 belongs to this case.

In what follows, we assume that $p \in \mathbb{J}$, unless otherwise noted.

LEMMA B3: *Given conditions A1, A2(i), A3, A5, and* $\mathbb{H}_1^{\ell}$,

(i) $\sup_y |\alpha(y)| \leq \Delta$, *where* $\alpha(y) := \int_y^{\infty} D(y,x)dF(x)$;

(ii) *for each* p *and* $k = 1, 2, ...,$ (15) *holds, and* $r_k(p, Y_{n,t}) = O_{\mathbb{P}}(n^{-1})$ *uniformly in* p;

(iii) *for each* p *and* $k = 1, 2, ...,$ (16) *holds, and* $r_{n,k}(p) = O(n^{-1})$ *uniformly in* p;

(iv) *for each* $k = 1, 2, ...,$ $h_{n,k}(p) \to h_k(p)$ *and* $r_{n,k}(p) \to r_k(p)$ *uniformly in* p *a.s.* $-\mathbb{P}$,

(v) *for each* $p \in \mathbb{I}$ *such that* $p > \frac{1}{n}$, *if we let* $\widetilde{p}_n := F(\widetilde{q}_n(p))$;

$$\mathbb{P}(\widetilde{R}_{n,i}(p) = k|\widetilde{p}_n) = (1-\widetilde{p}_n)^{k-1}\widetilde{p}_n + n^{-1/2}\frac{h_{n,k}(\widetilde{p}_n)}{F_n(F^{-1}(\widetilde{p}_n))} + \frac{r_{n,k}(\widetilde{p}_n)}{F_n(F^{-1}(\widetilde{p}_n))}; \tag{19}$$

(vi) *for each* $(p,s) \in \mathbb{I} \times \mathbb{S}$ *such that* $p > \frac{1}{n}$,

$$\sqrt{n}\left(E[s^{\widetilde{R}_{n,i}(p)}|\widetilde{p}_n] - \frac{s\widetilde{p}_n}{1-s(1-\widetilde{p}_n)}\right) = \frac{\nu(\widetilde{p}_n, s)}{F_n(F^{-1}(\widetilde{p}_n))} + o_{\mathbb{P}}(1), \tag{20}$$

where $\nu(p,s) := \frac{ps^2(1-s)w(p)}{\{1-s(1-p)\}^2} + \frac{s(1-s)}{\{1-s(1-p)\}}\int_{-\infty}^{F^{-1}(p)} C(p,y)dF(y)$.

Proof of Lemma B3: (i) We note that $|\alpha(y)| \leq \int_y^{\infty}|D(y,x)|dF(x) \leq \int_{-\infty}^{\infty}|D(y,x)|dF(x) \leq \Delta$ by A5(iv).

(ii) First, from the local alternative (13), $\mathbb{P}(Y_{n,t+1} < y|\mathcal{F}_{n,t}) = \int_{-\infty}^{y} dF_n(x|Y_{n,t}) = p + n^{-1/2}C(p, Y_{n,t}) = p + n^{-1/2}h_1(p, Y_{n,t})$, where p denotes $F(y)$. Second, $\mathbb{P}(Y_{n,t+1} \geq y, Y_{n,t+2} < y|\mathcal{F}_{n,t}) = \int_y^{\infty}\{p+n^{-1/2}C(p,x)\}dF_n(x|Y_{n,t-1}) = p(1-p)+n^{-1/2}\{w(p)-p\,C(p,Y_{n,t})\}+r_2(p,Y_{n,t}) = p(1-p)+n^{-1/2}h_2(p,Y_{n,t}) + r_2(p,Y_{n,t})$, where $r_2(p,Y_{n,t}) := n^{-1}R(F^{-1}(p),Y_{n,t})$ and $R(y,Y_{n,t}) := \int_y^{\infty} D(y,x)\, dD(x,Y_{n,t})$, which is $O_{\mathbb{P}}(n^{-1})$ by A5(iv). Third, $\mathbb{P}(Y_{n,t+1} \geq y, Y_{n,t+2} \geq y, Y_{n,t+3} < y|\mathcal{F}_{n,t}) = \int_y^{\infty}\{p(1-p)+n^{-1/2}\{w(p)-p\,C(p,x)\}\}dF_n(x|Y_{n,t-1})+r_3(p,Y_{n,t}) = p(1-p)^2+n^{-1/2}\{w(p)(1-2p)-p(1-p)\,C(p,Y_{n,t})\}+r_3(p,Y_{n,t}) = p(1-p)^2+n^{-1/2}h_3(p,Y_{n,t})+r_3(p,Y_{n,t})$, where $r_3(p,Y_{n,t})$ is defined to be $n^{-1}\int_y^{\infty} R(F^{-1}(p),z)\{dF(z) + n^{-1/2}dD(z,Y_{n,t})\} - n^{-1}\{w(p)C(p,Y_{n,t}) - pR(F^{-1}(p),Y_{n,t})\}$, which is also $O_{\mathbb{P}}(n^{-1})$ uniformly in p because (i) $|w(\cdot)|$ is bounded by Δ by Lemma B3(i), and (ii) both $|C(\cdot, Y_{n,t})|$ and $|R(F^{-1}(\cdot), Y_{n,t})|$ are bounded by $M_{n,t}$ by A5(iii and iv), and $\int_y^{\infty}|R(F^{-1}(p),z)|\{dF(z) + n^{-1/2}dD(z,Y_{n,t})\} \leq M_{n,t}$ uniformly in p. Thus, applying Markov's inequality ensures that $r_3(p,Y_{n,t})$ is $O_{\mathbb{P}}(n^{-1})$ uniformly in p.

Given this, we apply the induction method to obtain the desired result. Suppose that $\mathbb{P}(Y_{n,t+1} \geq y, Y_{n,t+2} \geq y, ..., Y_{n,t+k} < y|\mathcal{F}_{n,t}) = p(1-p)^{k-1} + n^{-1/2}h_k(p,Y_{n,t}) + r_k(p,Y_{n,t})$, and $r_k(p,Y_{n,t}) =$

$O_{\mathbb{P}}(n^{-1})$ uniformly in p. Then $\mathbb{P}(Y_{n,t+1} \geq y, Y_{n,t+2} \geq y, ...Y_{n,t+k+1} < y|\mathcal{F}_{n,t}) = \int_y^{\infty}\{p(1-p)^{k-1} + n^{-1/2}h_k(p, y_{n,t+1}) + r_k(p, y_{n,t+1})\}dF_n(y_{n,t+1}|\mathcal{F}_{n,t}) = p(1-p)^k + n^{-1/2}\{w(p)(1-kp)(1-p)^{k-2} - p(1-p)^{k-1}C(p, Y_{n,t})\} + r_{k+1}(p, Y_{n,t}) = p(1-p)^k + n^{-1/2}h_{k+1}(p, Y_{n,t}) + r_{k+1}(p, Y_{n,t})$, where $k = 3, 4, ...$, and $r_{k+1}(p, Y_{n,t}) := \int_y^{\infty} r_k(p, z)\{dF(z) + n^{-1/2}dD(z, Y_{n,t})\} - n^{-1}\{(1-p)^{k-3}(1-(k-1)p)w(p)C(p, Y_{n,t}) + p(1-p)^{k-2}R(F^{-1}(p), Y_{n,t})\}$. Note that the first component in the RHS is also $O_{\mathbb{P}}(n^{-1})$ uniformly in p as F_n is a proper distribution function and because $r_k(p, Y_{n,t})$ is $O_{\mathbb{P}}(n^{-1})$ uniformly in p. Also, given that $|w(\cdot)C(\cdot, Y_{n,t})| < \Delta \cdot M_{n,t}$ and $|R(F^{-1}(\cdot), Y_{n,t})| \leq M_{n,t}$, the remainders are $O_{\mathbb{P}}(n^{-1})$ uniformly in p because $|(1-p)^{k-3}(1-(k-1)p)| < 3$ and $|p(1-p)^{k-2}| < 1$ uniformly in p and $k = 3, 4, ...$ Thus, $r_{k+1}(p, Y_{n,t})$ is $O_{\mathbb{P}}(n^{-1})$ uniformly in p, and (15) follows from the definition of h_{k+1}. These are the desired results.

(iii) By Lemma B3(ii), for each $k = 1, 2, ...$, $\mathbb{P}(R_{n,i}(p) = k) = E[\mathbb{P}(Y_{n,t+1} \geq y, Y_{n,t+2} \geq y, ...Y_{n,t+k} < y|\mathcal{F}_{n,t})|Y_{n,t} < y] = p(1-p)^{k-1} + n^{-1/2}E[h_k(p, Y_{n,t})|Y_{n,t} < y] + E[r_k(p, Y_{n,t})|Y_{n,t} < y]$. Further, using the definitions of $h_{n,k}(p)$ and $r_{n,k}(p)$, we can substitute $F_n(F^{-1}(p))^{-1}h_{n,k}(p)$ and $F_n(F^{-1}(p))^{-1}r_{n,k}(p)$ into $E[h_k(p, Y_{n,t})|Y_{n,t} < y]$ and $E[r_k(p, Y_{n,t})|Y_{n,t} < y]$ respectively, yielding (16). In addition, given that $r_k(p, Y_{n,t}) = O_{\mathbb{P}}(n^{-1})$ uniformly in p, it trivially follows from the definition of $r_{n,k}(p)$ that $r_{n,k}(p) = O(n^{-1})$ uniformly in p.

(iv) Given the definition of $h_k(\cdot, Y_{n,t})$, we have that $h_k(\cdot, Y_{n,t})$ is uniformly bounded by $3\Delta + M_{n,t}$ as $|h_1(p, Y_{n,t})| = |C(p, Y_{n,t})| \leq M_{n,t}$; $|h_2(p, Y_{n,t})| \leq |w(p)| + |pC(p, Y_{n,t})| \leq \Delta + M_{n,t}$; and for $k = 3, 4, ...$, $|h_k(p, Y_{n,t})| \leq |w(p)(1-p)^{k-3}(1-(k-1)p)| + p(1-p)^{k-2}|C(p, Y_{n,t})| \leq 3\Delta + M_{n,t}$. Therefore, $\int |h_k(p, x)|dF_n(x) \leq 3\Delta + E[M_{n,t}] \leq 4\Delta < \infty$ uniformly in n, so that $\int_{-\infty}^{y} h_k(p, x)dF_n(x) \leq 4\Delta$ uniformly in p and n. Also, A5(iv) implies that for every k, $\int |h_k(p, x)|dF(x) \leq 3\Delta + \int |D(F(y), x)|dF(x) \leq 4\Delta$, implying that $\int_{-\infty}^{y} h_k(p, x)dF(x) \leq 4\Delta$ uniformly in p, and $|\int_{-\infty}^{y} h_k(p, x)dF_n(x) - \int_{-\infty}^{y} h_k(p, x) dF(x)| = n^{-1/2}|\int_{-\infty}^{y}\int_{-\infty}^{\infty} h_k(p, x)dD(x, z)dF(z) + o_{\mathbb{P}}(1)| < 4\Delta$ uniformly in p. Thus, $|h_{n,k}(p) - h_k(p)| \to 0$ uniformly in p as n tends to infinity.

As shown in the proof for Lemma B3(ii), $r_k(p, Y_{n,t}) = O_{\mathbb{P}}(n^{-1})$, and $|r_k(p, Y_{n,t})|$ is bounded uniformly in p by an integrable positive random variable. Given this, the same argument proving $h_{n,k}(p) \to h_k(p)$ uniformly in p applies to show $r_{n,k}(p) \to r_k(p)$ uniformly in p.

(v) We can replace p in (16) with $\widetilde{p}_n$ to obtain the given result.

(vi) Using the definition of $h_{n,k}(\widetilde{p}_n)$, Lemmas B3(iii and v), and the fact that

$$\sum_{k=1}^{\infty} s^k h_{n,k}(\widetilde{p}_n) = \frac{F_n(\widetilde{q}_n)}{\widetilde{p}_n}\left[\nu(\widetilde{p}_n, s) + \frac{s(1-s)(\widetilde{p}_n - F_n(\widetilde{q}_n))}{F_n(\widetilde{q}_n)\{1 - s(1-\widetilde{p}_n)\}}\int_{-\infty}^{F^{-1}(\widetilde{p}_n)} C(\widetilde{p}_n, x)dF_n(x)\right]$$
$$= \nu(\widetilde{p}_n, s) + o_{\mathbb{P}}(1),$$

we obtain that $n^{1/2}(E[s^{\widetilde{R}_{n,i}(p)}|\widetilde{p}_n] - \widetilde{p}_n s/\{1 - s(1-\widetilde{p}_n)\}) = \nu(\widetilde{p}_n, s)/F_n(F^{-1}(\widetilde{p}_n)) + o_{\mathbb{P}}(1)$. This completes the proof. ■

LEMMA B4: *Given conditions A1, A2*(i), *A3, A5, and* $\mathbb{H}_1^\ell$, *if* $R_{n,i}(p)$ *is defined by observations starting from* $Y_{n,t+1}$ *and* $p > \frac{1}{n}$, *then*

(i) *for each* p *and* $k = 1, 2, ...,$

$$\mathbb{P}(R_{n,i}(p) = k|\mathcal{F}_{n,t}) = p(1-p)^{k-1} + n^{-1/2}p^{-1}h_k(p) + p^{-1}r_k(p) + n^{-1}b_k(p, Y_{n,t-1}) + o_{\mathbb{P}}(n^{-1}), \quad (21)$$

where $b_k(p, Y_{n,t-1}) := p^{-1}\int_{-\infty}^{y} h_k(p,x)dD(x, Y_{n,t-1}) - p^{-2}h_k(p)D(y, Y_{n,t-1})$;

(ii) *for each* p *and* $k = 1, 2, ..., \mathbb{P}(R_{n,i}(p) = k) = p(1-p)^{k-1} + n^{-1/2}p^{-1}h_k(p) + p^{-1}r_k(p) + n^{-1}b_k(p) + o(n^{-1})$, *where* $b_k(p) := \int b_k(p,z)dF(z)$.

Proof of Lemma B4: (i) Assuming that $R_{n,i}(p)$ is defined by observations starting from $Y_{n,t+1}$, we have $\mathbb{P}(R_{n,i}(p) = k|\mathcal{F}_{n,t}) = \mathbb{P}(Y_{n,t+1} \geq y, ..., Y_{n,t+k-1} \geq y, Y_{n,t+k} < y|\bar{\mathcal{F}}_{n,t-1})$, where $\bar{\mathcal{F}}_{n,t-1} := \sigma(Y_{n,t} < y, Y_{n,t-1}, Y_{n,t-2}, ...)$. Given this, we can use (15) to obtain

$$\begin{aligned}\mathbb{P}(R_{n,i}(p) = k|\mathcal{F}_{n,t}) = p(1-p)^{k-1} &+ n^{-1/2}F_n(y|Y_{n,t-1})^{-1}\int_{-\infty}^{y} h_k(p,x)dF_n(x|Y_{n,t-1}) \\ &+ F_n(y|Y_{n,t-1})^{-1}\int_{-\infty}^{y} r_k(p,x)dF_n(x|Y_{n,t-1}) \quad (22)\end{aligned}$$

using the facts that $E[h_k(p, Y_{n,t})|\bar{\mathcal{F}}_{n,t-1}] = F_n(y|Y_{n,t-1})^{-1}\int_{-\infty}^{y} h_k(p,x)dF_n(x|Y_{n,t-1})$ and $E[r_k\ (p, Y_{n,t})|\bar{\mathcal{F}}_{n,t-1}] = F_n(y|Y_{n,t-1})^{-1}\int_{-\infty}^{y} r_k(p,x)dF_n(x|Y_{n,t-1})$. Using a Taylor expansion yields $F_n(y|Y_{n,t-1})^{-1} = 1/F(y) - n^{-1/2}D(y, Y_{n,t-1})/F(y)^2 + o_{\mathbb{P}}(n^{-1/2})$, and substituting (13) and this into (22) yields the desired result.

(ii) We note the fact that $\mathbb{P}(R_{n,i}(p) = k) = E[\mathbb{P}(R_{n,i}(p) = k|\mathcal{F}_{n,t})]$. Thus, $\mathbb{P}(R_{n,i}(p) = k) = p(1-p)^{k-1} + n^{-1/2}p^{-1}h_k(p) + p^{-1}r_k(p) + n^{-1}\int b_k(p,z)dF_n(z) + o(n^{-1}) = p(1-p)^{k-1} + n^{-1/2}p^{-1}h_k(p) + p^{-1}r_k(p) + n^{-1}\int b_k(p,z)dF(z) + o(n^{-1})$ by Lemma B1(ii). ■

LEMMA B5: *Given conditions A1, A2*(i), *A3, A5, and* $\mathbb{H}_1^\ell$, *if* $R_{n,i}(p)$ *is defined by observations starting from* $Y_{n,t+1}$ *and* $p > \frac{1}{n}$, *then*

(i) *for each* p *and* $k, m = 1, 2, ..., \mathbb{P}(R_{n,i}(p) = k|\mathcal{F}_{n,t-m}) - \mathbb{P}(R_{n,i}(p) = k) = n^{-\frac{m+1}{2}}B_{k,m}(p, Y_{n,t-m}) + o_{\mathbb{P}}(n^{-\frac{m+1}{2}})$, *where* $B_{k,1}(p, Y_{n,t-1}) := b_k(p, Y_{n,t-1}) - b_k(p)$, *and for* $m = 2, 3, ..., B_{k,m}(p, Y_{t-m}) := \int ... \int b_k\ (p,z)dD(z,x_1)...dD(x_{m-2}, Y_{n,t-m}) - n^{-\frac{m+1}{2}}\int ... \int b_k(p,z)dD(z,x_1)...dD(x_{m-2},x)dF(x)$;

(ii) *for each* p *and* $k, \ell, m = 1, 2, ..., \mathbb{P}(R_{n,i}(p) = k|R_{n,i-m}(p) = \ell) = \mathbb{P}(R_{n,i}(p) = k) + O_{\mathbb{P}}(n^{-(m+1)/2})$;

(iii) *for each* p *and* $k, \ell, m = 1, 2, ..., \mathbb{P}(R_{n,i}(p) = k, R_{n,i-m}(p) = \ell) = \mathbb{P}(R_{n,i}(p) = k)\mathbb{P}(R_{n,i-m}(p) = \ell) + O(n^{-(m+1)/2})$.

Proof of Lemma B5: (i) For $m = 1$, the given result trivially follows from Lemmas B4(i and ii). To show the claim for $m > 1$, we let $c_k(p, Y_{n,t-1}) := b_k(p, Y_{n,t-1}) + o_{\mathbb{P}}(1)$ for notational simplicity, where the $o_{\mathbb{P}}(1)$ term is the last piece multiplied by n in the RHS of (21). Then $\mathbb{P}(R_{n,i}(p) = k|\mathcal{F}_{n,t-m}) = p(1-p)^{k-1} + n^{-1/2}p^{-1}h_k(p) + p^{-1}r_k(p) + n^{-1}E[c_k(p, Y_{n,t-1})|\mathcal{F}_{n,t-m}]$ and $\mathbb{P}(R_{n,i}(p) = k) = p(1-p)^{k-1} + n^{-1/2}p^{-1}h_k(p) + p^{-1}r_k(p) + n^{-1}E[c_k(p, Y_{n,t-1})]$. Thus, $\mathbb{P}(R_{n,i}(p) = k|\mathcal{F}_{n,t-m}) - \mathbb{P}(R_{n,i}(p) = k) = n^{-1}\{E[c_k(p, Y_{n,t-1})|\mathcal{F}_{n,t-m}] - E[c_k(p, Y_{n,t-1})]\}$. Given this, applying Lemma B1(i) and the fact that $c_k(p, Y_{n,t-1}) = b_k(p, Y_{n,t-1}) + o_{\mathbb{P}}(1)$ yields that $\mathbb{P}(R_{n,i}(p) = k|\mathcal{F}_{n,t-m}) - \mathbb{P}(R_{n,i}(p) = k) = n^{-\frac{m+1}{2}} \int ... \int b_k(p, z)dD(z, x_1)...dD(x_{m-2}, Y_{n,t-m}) - n^{-\frac{m+1}{2}} \int ... \int b_k(p, z)dD(z, x_1)...dD(x_{m-2}, x)$ $dF(x) + o_{\mathbb{P}}(n^{-\frac{m+1}{2}}) = n^{-\frac{m+1}{2}} B_{k,m}(p, Y_{t-m}) + o_{\mathbb{P}}\,(n^{-\frac{m+1}{2}})$. This is the desired result.

(ii) Given that $R_{n,i}(p)$ is defined by the observations starting from $Y_{n,t+1}$, it follows that $Y_{n,t} < y$. Given this, we first consider $m = 1$. Then, $\{R_{n,i-1}(p) = \ell\} = \{Y_{n,t} < y, Y_{n,t-1} \geq y, ..., Y_{n,t-\ell-1} \geq y, Y_{n,t-\ell} < y\} \subset \bar{\mathcal{F}}_{n,t-1} := \sigma(Y_{n,t} < y, Y_{n,t-1}, Y_{n,t-2}, ...)$. Next, we suppose that $m > 1$. Then, the latest observation to define $R_{n,i-m}(p)$ is $Y_{n,t-m}$, which is obtained when $R_{n,i-1} = ... = R_{n,i-m+1} = 1$. Therefore, $\{R_{n,i-m}(p) = \ell\} \subset \bar{\mathcal{F}}_{n,t-m} := \sigma(Y_{n,t} < y, Y_{n,t-m}, Y_{n,t-m}, ...)$. Thus, for any m, $\{R_{n,i-m}(p) = \ell\} \subset \bar{\mathcal{F}}_{n,t-m}$. Finally, we apply Lemma B5(i) to obtain the desired result.

(iii) This trivially follows from Lemma B5(ii) and the fact that $\mathbb{P}(R_{n,i}(p) = k, R_{n,i-m}(p) = \ell) = \mathbb{P}(R_{n,i}(p) = k|R_{n,i-m}(p) = \ell)\mathbb{P}(R_{n,i-m}(p) = \ell)$. ■

LEMMA B6: *Given conditions A1, A2(i), A3, A5, and $\mathbb{H}_1^\ell$, for each* (p, s),

(i) $E[W_n(p, s)|\widetilde{p}_n] = \nu(p, s) + o_{\mathbb{P}}(1)$;

(ii) $E[H_n(p, s)|\widetilde{p}_n] = -ps(1-s)Q(F^{-1}(p))/\{1 - s(1-p)\}^2 + o_{\mathbb{P}}(1)$;

(iii) $E[W_n(p, s)^2|\widetilde{p}_n] = p^2s^2(1-p)(1-s)^2/\{1 - s(1-p)\}^2\{1 - s^2(1-p)\} + \nu(p, s)^2 + o_{\mathbb{P}}(1)$;

(iv) $E[H_n(p, s)^2|\widetilde{p}_n] = p^3s^2(1-p)(1-s)^2/\{1 - s(1-p)\}^4 + p^2s^2(1-s)^2Q(F^{-1}(p))^2/\ \{1 - s(1-p)\}^4 + o_{\mathbb{P}}(1)$.

Proof of Lemma B6: (i) Lemma B3(vi) implies that $E[W_n(p, s)|\widetilde{p}_n] = p\{F_n(\widetilde{q}_n(p))^{-1}\nu(\widetilde{p}_n, s)\} + o_{\mathbb{P}}(1) = \nu(p, s) + o_{\mathbb{P}}(1)$, where the first equality follows because $\widetilde{M}_n(p)/n = \lfloor pn \rfloor/n$, $\widetilde{q}_n(p) \to F^{-1}(p)$ a.s.$-\mathbb{P}$, as the ergodic theorem implies that $\widetilde{p}_n \to p$ a.s.$-\mathbb{P}$.

(ii) For some $\bar{p}_n$ between $\widetilde{p}_n$ and p,

$$H_n(p, s) = \frac{s(1-s)\widetilde{M}_n(p)}{\{1 - s(1-\bar{p}_n)\}^2} \frac{\{\widetilde{p}_n - p\}}{\sqrt{n}} = -\frac{ps(1-s)\sqrt{n}[\widetilde{F}_n(\widetilde{q}_n(p)) - F(\widetilde{q}_n(p))]}{\{1 - s(1-\bar{p}_n)\}^2} + o_{\mathbb{P}}(1) \qquad (23)$$

by the mean-value theorem. This implies that applying Lemma B2 and the ergodic theorem yields $E[H_n(p, s)|\widetilde{p}_n] = -ps(1-s)Q(F^{-1}(p))/\{1 - s(1-p)\}^2 + o_{\mathbb{P}}(1)$, which ensures that $\widetilde{q}_n(p) \to F^{-1}(p)$ a.s.$-\mathbb{P}$.

(iii) We decompose $W_n(p, s)^2$ into $W_n(p, s)^2 \equiv \widetilde{K}_n(p, s) + \widetilde{L}_n^{(1)}(p, s) + \widetilde{L}_n^{(2)}(p, s)$, where $\widetilde{K}_n(p, s) := n^{-1}\sum_{i=1}^{\widetilde{M}_n(p)}(s^{\widetilde{R}_{n,i}(p)} - s\widetilde{p}_n/\{1 - s(1-\widetilde{p}_n)\})^2$, $\widetilde{L}_n^{(1)}(p, s) = n^{-1}\sum\sum_{|i-j|=1}^{\widetilde{M}_n(p)}(s^{\widetilde{R}_{n,i}(p)} - s\widetilde{p}_n/\{1 - s(1 -$

$\widetilde{p}_n)\})(s^{\widetilde{R}_{n,j}(p)} - s\widetilde{p}_n/\{1-s(1-\widetilde{p}_n)\})$, and $\widetilde{L}_n^{(2)}(p,s) = n^{-1}\sum\sum_{|i-j|>1}^{\widetilde{M}_n(p)}(s^{\widetilde{R}_{n,i}(p)} - s\widetilde{p}_n/\{1-s(1-\widetilde{p}_n)\})(s^{\widetilde{R}_{n,j}(p)} - s\widetilde{p}_n/\{1-s(1-\widetilde{p}_n)\})$, and examine the asymptotic behavior of each component. Note that $E[\widetilde{K}_n(p,s)|\widetilde{p}_n] = n^{-1}\sum_{i=1}^{\widetilde{M}_n(p)} E[(s^{\widetilde{R}_{n,i}(p)} - s\widetilde{p}_n/\{1-s(1-\widetilde{p}_n)\})^2|\widetilde{p}_n]$ and $n^{-1}\widetilde{M}_n(p) = \widetilde{p}_n + o_{\mathbb{P}}(1)$. From this and (16), we obtain $E[\widetilde{K}_n(p,s)|\widetilde{p}_n] = \widetilde{p}_n^2 s^2(1-\widetilde{p}_n)(1-s)^2/[\{1-s(1-\widetilde{p}_n)\}^2\{1-s^2(1-\widetilde{p}_n)\}] + o_{\mathbb{P}}(1)$. Next, $E[\widetilde{L}_n^{(1)}(p,s)|\widetilde{p}_n] = 2n^{-1}(\widetilde{M}_n(p)-1)E[(s^{\widetilde{R}_{n,i}(p)} - s\widetilde{p}_n/\{1-s(1-\widetilde{p}_n)\})(s^{\widetilde{R}_{n,i+1}(p)} - s\widetilde{p}_n/\{1-s(1-\widetilde{p}_n)\})|\widetilde{p}_n]$. We use Lemma B5($iii$) to obtain $E[\widetilde{L}_n^{(1)}(p,s)|\widetilde{p}_n] = 2\widetilde{p}_n E[s^{\widetilde{R}_{n,i}(p)} - s\widetilde{p}_n/\{1-s(1-\widetilde{p}_n)\}|\widetilde{p}_n]^2 + O_{\mathbb{P}}(n^{-1}) = 2n^{-1}\widetilde{p}_n F_n(\widetilde{q}_n(p))^{-2}\nu(\widetilde{p}_n,s)^2 + O_{\mathbb{P}}(n^{-1}) = o_{\mathbb{P}}(1)$. Finally, $E[\widetilde{L}_n^{(2)}(p,s)|\widetilde{p}_n]$ $= n^{-1}\sum\sum_{|i-j|>1}^{\widetilde{M}_n(p)} E[(s^{\widetilde{R}_{n,i}(p)} - s\widetilde{p}_n/\{1-s(1-\widetilde{p}_n)\})(s^{\widetilde{R}_{n,i+1}(p)} - s\widetilde{p}_n/\{1-s(1-\widetilde{p}_n)\})|\widetilde{p}_n]$; using Lemma B5($iii$) then implies $E[\widetilde{L}_n^{(2)}(p,s)|\widetilde{p}_n] = n^{-1}(\widetilde{M}_n(p)-1)(\widetilde{M}_n(p)-2)\{E[s^{\widetilde{R}_{n,i}(p)} - s\widetilde{p}_n/\{1-s(1-\widetilde{p}_n)\}|\widetilde{p}_n]^2 + o_{\mathbb{P}}(n^{-1})\}$, so that $E[\widetilde{L}_n^{(2)}(p,s)|\widetilde{p}_n] = \widetilde{p}_n^2 F_n(\widetilde{q}_n(p))^{-2}\nu(\widetilde{p}_n,s)^2 + o_{\mathbb{P}}(1)$. Thus, adding together $E[\widetilde{K}_n(p,s)|\widetilde{p}_n]$, $E[\widetilde{L}_n^{(1)}(p,s)|\widetilde{p}_n]$, and $E[\widetilde{L}_n^{(2)}(p,s)|\widetilde{p}_n]$ and using $\widetilde{p}_n \to p$ a.s.$-\mathbb{P}$ (by the ergodic theorem), we have $E[W_n(p,s)^2|\widetilde{p}_n] = p^2 s^2(1-p)(1-s)^2/[\{1-s(1-p)\}^2\{1-s^2(1-p)\}] + \nu(p,s)^2 + o_{\mathbb{P}}(1)$.

(iv) This directly follows from Lemma B2 and (23) and by letting $p = F(y)$. ■

LEMMA B7: *Given conditions A1, A2(i), A3, A5, and $\mathbb{H}_1^\ell$, for each (p,s),*

(i) $W_n(p,s) - \widetilde{W}_n(p,s) = \nu(p,s) + o_{\mathbb{P}}(1)$, *where* $\widetilde{W}_n(p,s) := n^{-1/2}\sum_{i=1}^{\widetilde{M}_n(p)}(s^{\widetilde{R}_{n,i}(p)} - E[s^{\widetilde{R}_{n,i}(p)}|\widetilde{p}_n])$;

(ii) $E[S_{n,i}(p,s)|\mathcal{F}_{n,t-m}] = O_{\mathbb{P}}(n^{-\frac{m+1}{2}})$, *where* $S_{n,i}(p,s) := s^{R_{n,i}(p)} - E[s^{R_{n,i}(p)}]$, *and* $R_{n,i}(p)$ *is the run defined by observations starting from* $Y_{n,t+1}$.

Proof of Lemma B7: (i) We exploit Lemma B3(vi) to show the given result. Note that $W_n(p,s) - \widetilde{W}_n(p,s) = n^{-1/2}\widetilde{M}_n(p)(E[s^{\widetilde{R}_{n,i}(p)}|\widetilde{p}_n] - \widetilde{p}_n s/\{1-s(1-\widetilde{p}_n)\})$, and $n^{-1/2}\widetilde{M}_n(p) = pn^{1/2} + o_{\mathbb{P}}(1)$. Therefore, (20) implies that $W_n(p,s) - \widetilde{W}_n(p,s) = p\nu(\widetilde{p}_n,s)/F_n(F^{-1}(\widetilde{p}_n)) + o_{\mathbb{P}}(1) = \nu(p,s) + o_{\mathbb{P}}(1)$, where the last equality follows because $F^{-1}(\widetilde{p}_n) = F^{-1}(p) + o_{\mathbb{P}}(1) = y + o_{\mathbb{P}}(1)$ by the ergodic theorem and the fact that $F_n(y) = F(y) + o(1) = p + o(1)$ when y is such that $p = F(y)$.

(ii) By the definition of $E[S_{n,i}(p,s)|\mathcal{F}_{n,t-m}]$, $E[S_{n,i}(p,s)|\mathcal{F}_{n,t-m}] = \sum_{k=1}^\infty s^k\{\mathbb{P}(R_{n,i}(p) = k|\mathcal{F}_{n,t-m}) - \mathbb{P}(R_{n,i}(p) = k)\}$. We note that $|s| < 1$, and $\mathbb{P}(R_{n,i}(p) = k|\mathcal{F}_{n,t-m}) - \mathbb{P}(R_{n,i}(p) = k) = n^{-\frac{m+1}{2}}B_{k,m}(p, Y_{n,t-m}) + o_{\mathbb{P}}(n^{-\frac{m+1}{2}})$ by Lemma B5(i). This completes the proof. ■

LEMMA B8: *Given conditions A1, A2(i), A3, A5, and $\mathbb{H}_1^\ell$, for each (p,s),* $\widetilde{G}_n(p,s) \overset{\text{A}}{\sim} N(\mu(p,s), s^2p^2(1-s)^4(1-p)^2/[\{1-s(1-p)\}^4\{1-s^2(1-p)\}])$.

Proof of Lemma B8: To show the given claim, we partition our proof into three pieces. First, we obtain the asymptotic population mean of $\widetilde{G}_n(p,s)$. Second, we derive the asymptotic variance of $\widetilde{G}_n(p,s)$. Finally, we derive the asymptotic distribution.

First, we note that $E[\widetilde{G}_n(p,s)|\widetilde{p}_n] = E[W_n(p,s)|\widetilde{p}_n] + E[H_n(p,s)|\widetilde{p}_n]$. Given this, Lemmas B6($i$ and ii) ensure that $E[\widetilde{G}_n(p,s)|\widetilde{p}_n] = \nu(p,s) - ps(1-s)Q(F^{-1}(p))/\{1-s(1-p)\}^2 + o_{\mathbb{P}}(1) = \mu(p,s) + o_{\mathbb{P}}(1)$. Thus, the asymptotic population mean of $\widetilde{G}_n(p,s)$ must be $\mu(p,s)$.

Second, we apply Lemmas B6(i and iii) to show that the asymptotic variance of $W_n(p,s)$ satisfies

$$E[W_n(p,s)^2|\widetilde{p}_n] - E[W_n(p,s)|\widetilde{p}_n]^2 = \frac{p^2s^2(1-p)(1-s)^2}{\{1-s(1-p)\}^2\{1-s^2(1-p)\}} + o_{\mathbb{P}}(1). \tag{24}$$

Lemmas B6(ii and iv) imply that

$$E[H_n(p,s)^2|\widetilde{p}_n] - E[H_n(p,s)|\widetilde{p}_n]^2 = \frac{p^3s^2(1-p)(1-s)^2}{\{1-s(1-p)\}^4} + o_{\mathbb{P}}(1). \tag{25}$$

Further, we have $E[W_n(p,s)|\mathcal{F}_{n,n}] = -H_n(p,s) + \mu(p,s)$, so that $E[W_n(p,s)H_n(p,s)|\widetilde{p}_n] = -E[H_n(p,s)^2|\widetilde{p}_n] + \mu(p,s)E[H_n(p,s)|\widetilde{p}_n] + o_{\mathbb{P}}(1)$. Thus,

$$\begin{aligned} &2\{E[W_n(p,s)H_n(p,s)|\widetilde{p}_n] - E[W_n(p,s)|\widetilde{p}_n]E[H_n(p,s)|\widetilde{p}_n]\} \\ &\qquad = -2\{E[H_n(p,s)^2|\widetilde{p}_n] - E[H_n(p,s)|\widetilde{p}_n]^2\} + o_{\mathbb{P}}(1). \end{aligned} \tag{26}$$

Given these, and adding together (24), (25) and (26), we obtain $E[\widetilde{G}_n(p,s)^2|\widetilde{p}_n] - E[\widetilde{G}_n(p,s)|\widetilde{p}_n]^2 = s^2p^2(1-s)^4(1-p)^2/\{1-s(1-p)\}^4\{1-s^2(1-p)\} + o_{\mathbb{P}}(1)$, which is the desired asymptotic variance.

Third, the asymptotic normality of $H_n(p,s)$ follows from Lemma B2 and (23); the asymptotic covariance between $H_n(p,s)$ and $W_n(p,s)$ is given in (26). Thus, if the asymptotic normality of $W_n(p,s)$ is proved, then the asymptotic normality of $\widetilde{G}_n(p,s)$ is obtained. For this, we consider $\widetilde{W}_n(p,s)$, which is $W_n(p,s) - \nu(p,s) + o_{\mathbb{P}}(1)$ by Lemma B7(i), so that these have the same limiting distribution. This can be shown by corollary 2 of Herrndorf (1984). First, from the definition of $S_{n,i}(p,s)$, $E[S_{n,i}(p,s)] = 0$, and $E[S_{n,i}(p,s)^2] < \infty$ because $|s| < 1$. Therefore, $\{S_{n,i}(p,s)\}$ satisfies the condition (1.1) of Herrndorf (1984). Second, Lemmas B6(i and iii) show that $E[\widetilde{W}_n(p,s)^2] = E[W_n(p,s)^2] - E[W_n(p,s)]^2 + o_{\mathbb{P}}(1) = p^2s^2(1-p)(1-s)^2/[\{1-s(1-p)\}^2\ \{1-s^2(1-p)\}] + o_{\mathbb{P}}(1)$, which is greater than zero. Thus, (1.2) of Herrndorf (1984) holds. Third, (1.3) of Herrndorf (1984) trivially holds by the stationarity condition. Finally, the size of α-mixing coefficient is bounded by β-mixing coefficient because $2\alpha_{n,k} \leq \beta_{n,k}$, and $\beta_{n,k} \leq n^{-k/2}\int K(y)dF_n(y)$ by Remark 4($a-2$). Thus, $\alpha_{n,k} = O(n^{-k/2})$. Further, $|s| < 1$, so that for any $\kappa > 2$, $E[S_{n,i}^{\kappa}] < \infty$, implying that (1.3) of Herrndorf (1984) holds. Therefore, the asymptotic normality of $\widetilde{W}_n(p,s)$ follows by corollary 2 of Herrndorf (1984). ■

Remark 6: (a) The given weak convergence in Lemma B8 can be also proved using the CLT, exploiting the β-mixing coefficients.

(b) The given weak convergence in Lemma 8 can be further extended to weak convergence involving multivariate random variables on multiple elements in $\mathbb{J} \times \mathbb{S}$ by the Cramér-Wold device. We do not show this for brevity.

LEMMA B9: *Given conditions A1, A2(i), A3, A5, and* $\mathbb{H}_1^\ell$,

(i) $W_n - \nu \Rightarrow \mathcal{G}$; *and*

(ii) $H_n - (\mu - \nu) \Rightarrow \mathcal{H}$.

Proof of Lemma B9: (i) We can use theorem 3 of Bickel and Wichura (1973) as before to establish this. Computing the modulus of continuity based on the fourth moment is straightforward, but a tedious task. We omit this for brevity because it is almost identical to the proof of Theorem 1 (i, ii, and iii).

(ii) Note that for each (p, s), $\mu(p, s) - \nu(p, s) = -ps(1 - s)Q(F^{-1}(p))/\{1 - s(1 - p)\}^2$ and $H_n(p, s) = -ps(1 - s)\sqrt{n}\{\widetilde{F}_n(\widetilde{q}_n(p)) - F(\widetilde{q}_n(p))\}/\{1 - s(1 - \bar{p}_n)\}^2 + o_{\mathbb{P}}(1)$ by (23). Thus, if we can show that $\sqrt{n}\{\widetilde{F}_n - F\}$ is tight, then the desired result follows by Lemma B2. We complete the proof by applying theorem 5.3 of Ango Nze and Doukhan (2004), which says that if $\beta_{n,k} = O(k^{-1}(\log k)^{-a})$ for some $a > 2$, then the tightness of $\sqrt{n}\{\widetilde{F}_n - F\}$ follows. If we let $a = 3$ then $\beta_{n,k} = o(k^{-1}(\log k)^{-a})$ for any $n > 1$ by Remark 4(b-2). ■

Remark 7: (a) For brevity, we omit deriving the asymptotic covariance structure of W_n and H_n under H_1^ℓ as this can be obtained in a manner similar to that obtaining the asymptotic variance of $\widetilde{G}_n(p, s)$.

(b) Given the fact that $\mathcal{G}$ and $\mathcal{H}$ are in $\mathcal{C}(\mathbb{J} \times \mathbb{S})$, they are tight, so that lemma 1.3.8(ii) of van der Vaart and Wellner (1996, p. 21) implies that W_n and H_n are tight.

Proof of Theorem 6: Given the weak convergence in Lemma B8, the desired result follows by the tightness implied by Lemma B9(i) (see Remark 6(b)) and the fact that $(W_n, H_n)'$ is tight by lemma 1.4.3 of van der Vaart and Wellner (1996, p. 30). ■

7.2 Other Test Statistics

In this section, we provide formulae for the other test statistics used in our Monte Carlo experiments.

The following statistics are used for DGPs 2.1, 2.4–2.6, and 2.8–2.11:

$RE_n := \max_k \frac{k}{\hat{\sigma}_n n} \max[\widetilde{\alpha}_k, \widetilde{\beta}_k]$;
$RR_n := \max_k \frac{k}{\hat{\sigma}_n n} \max[\widetilde{\alpha}_k, \widetilde{\beta}_k] - \min_k \frac{k}{\hat{\sigma}_n n} \max[\widetilde{\alpha}_k, \widetilde{\beta}_k]$;
$RECUSUM_n := \max_k \frac{1}{\widetilde{\sigma}_n \sqrt{n-1}} |\sum_{t=2}^k v_t|$;
$OLSCUSUM_n := \max_k \frac{1}{\hat{\sigma}_n \sqrt{n}} |\sum_{t=1}^k e_t|$;
$M_n := \max_k \sup_z |\frac{k}{n}(1 - \frac{k}{n})\sqrt{n}(k^{-1} \sum_{t=1}^k \mathbf{1}_{\{e_t \le z\}} - (n - k)^{-1} \sum_{t=k+1}^n \mathbf{1}_{\{e_t \le z\}})|$;
$\text{Sup}W_n := \sup_{k_1 \le k \le k_2} W_n(k)$;
$\text{Exp}W_n := \ln\{\frac{1}{k_2 - k_1 + 1} \sum_{k=k_1}^{k_2} \exp[0.5 W_n(k)]\}$;
$\text{Avg}W_n := \frac{1}{k_2 - k_1 + 1} \sum_{k=k_1}^{k_2} W_n(k)$,

where

$[\widetilde{\alpha}_k, \widetilde{\beta}_k]' = \mathbf{M}_n^{1/2}[\hat{\alpha}_{1,k} - \hat{\alpha}_{1,n}, \hat{\beta}_{1,k} - \hat{\beta}_{1,n}]'$; $\mathbf{M}_n := \mathbf{Z}'\mathbf{Z}$;

$\mathbf{Z}$ is the $n \times 2$ matrix of the regressors;

$(\hat{\alpha}_{1,k}, \hat{\beta}_{1,k})$ is the OLS estimator using the first k observations;

$(\hat{\alpha}_{2,k}, \hat{\beta}_{2,k})$ is the OLS estimator using the $(k+1)$-th to n-th observations;

$e_t := Y_t - \hat{\alpha}_{1,n} - \hat{\beta}_{1,n}Z_t$; $\hat{\sigma}_n^2 := (n-2)^{-1}\sum_{t=1}^n e_t^2$;

$v_t := Y_t - \hat{\alpha}_{1,t-1} - \hat{\beta}_{1,t-1}Z_t$; $\widetilde{\sigma}_n^2 := (n-2)^{-1}\sum_{t=3}^n v_t^2$;

$W_n(k) := \{(n-2)\hat{\sigma}_n^2 - (n-3)\ddot{\sigma}_n^2(k)\}/\ddot{\sigma}_n^2(k)$; $k_1 = \lfloor 0.15n \rfloor$; $k_2 = \lfloor 0.85n \rfloor$;

$\hat{w}_{1,t}(k) := Y_t - \hat{\alpha}_{1,k} - \hat{\beta}_{1,k}Z_t$; $\hat{w}_{2,t}(k) := Y_t - \hat{\alpha}_{2,k} - \hat{\beta}_{2,k}Z_t$; and

$\ddot{\sigma}_n^2(k) := (n-3)^{-1}\{\sum_{t=1}^k \hat{w}_{1,t}(k)^2 + \sum_{t=k+1}^n \hat{w}_{2,t}(k)^2\}$.

The following statistics are used for DGPs 2.2, 2.7, and 2.12:

$RE_n := \max_k \frac{k}{\hat{\sigma}_n\sqrt{n}}|\widetilde{\beta}_k|$;

$RR_n := \max_k \frac{k}{\hat{\sigma}_n\sqrt{n}}|\widetilde{\beta}_k| - \min_k \frac{k}{\hat{\sigma}_n\sqrt{n}}|\widetilde{\beta}_k|$;

$RECUSUM_n := \max_k \frac{1}{\widetilde{\sigma}_n\sqrt{n-1}}|\sum_{t=2}^k v_t|$;

$NLSCUSUM_n := \max_k \frac{1}{\hat{\sigma}_n\sqrt{n}}|\sum_{t=1}^k e_t|$;

$M_n := \max_k \sup_z |\frac{k}{n}(1-\frac{k}{n})\sqrt{n}(k^{-1}\sum_{t=1}^k \mathbf{1}_{\{e_t \le z\}} - (n-k)^{-1}\sum_{t=k+1}^n \mathbf{1}_{\{e_t \le z\}})|$;

Sup$W_n := \sup_{k_1 \le k \le k_2} W_n(k)$;

Exp$W_n := \ln\{\frac{1}{k_2-k_1+1}\sum_{k=k_1}^{k_2}\exp[0.5W_n(k)]\}$;

Avg$W_n := \frac{1}{k_2-k_1+1}\sum_{k=k_1}^{k_2} W_n(k)$,

where

$\widetilde{\beta}_k = \widetilde{M}_n^{1/2}[\hat{\beta}_{1,k} - \hat{\beta}_{1,n}]$; $\widetilde{M}_n := \sum_{t=1}^n \exp(2Z_t\hat{\beta}_{1,n})Z_t^2$;

$\hat{\beta}_{1,k}$ is the NLS estimator using the first k observations;

$\hat{\beta}_{2,k}$ is the NLS estimators using the $(k+1)$-th to n-th observations;

$e_t := Y_t - \exp(\hat{\beta}_{1,n}Z_t)$; $\hat{\sigma}_n^2 := (n-1)^{-1}\sum_{t=1}^n e_t^2$;

$v_t := Y_t - \exp(\hat{\beta}_{1,t-1}Z_t)$; $\widetilde{\sigma}_n^2 := (n-1)^{-1}\sum_{t=2}^n v_t^2$;

$W_n(k) := n\{\hat{\beta}_{1,k} - \hat{\beta}_{2,k}\}\{n\hat{V}_{1,k}/k + n\hat{V}_{2,k}/(n-k)\}^{-1}\{\hat{\beta}_{1,k} - \hat{\beta}_{2,k}\}$; and

$\hat{V}_{1,k}$ and $\hat{V}_{2,k}$ are the variance estimators of $\hat{\beta}_{1,k}$ and $\hat{\beta}_{2,k}$ respectively.

References

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

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

ANDREWS, D. AND PLOBERGER,W. (1994): "Optimal Tests When a Nuisance Parameter is Present only under the Alternative," *Econometrica*, 62, 1383–1414.

ANGO NZE, P. AND DOUKHAN, P. (2004): "Weak Dependence Models and Applications to Econometrics," *Econometric Theory*, 20, 995–1045.

BAI, J. (1996): "Testing for Parameter Constancy in Linear Regressions: an Empirical Distribution Function Approach," *Econometrica*, 64, 597–622.

BICKEL, P. AND WICHURA, M. (1971): "Convergence Criteria for Multiparameter Stochastic Processes and Some Applications," *The Annals of Mathematical Statistics*, 42, 1656–1670.

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

BIERENS, H. (1990): "A Consistent Conditional Moment Test of Functional Form," *Econometrica*, 58, 1443–1458.

BIERENS, H. AND PLOBERGER, W. (1997): "Asymptotic Theory of Integrated Conditional Moment Tests," *Econometrica*, 65, 1129–1151.

BILLINGSLEY, P. (1968, 1999): *Convergence of Probability Measures*. New York: Wiley.

BRETT, C. AND PINKSE, J. (1997): "Those Taxes Are All over the Map! A Test for Spatial Independence of Municipal Tax Rates in British Columbia," *International Regional Science Review*, 20, 131–151.

BROWN, R., DURBIN, J. AND EVANS, J. (1975): "Techniques for Testing the Constancy of Regression Relationships over Time," *Journal of the Royal Statistical Society*, Series B, 37, 149–163.

CHU, J., HORNIK, K., AND KUAN, C.-M. (1995a): "MOSUM Tests for Parameter Constancy," *Biometrika*, 82, 603–617.

CHU, J., HORNIK, K., AND KUAN, C.-M. (1995b): "The Moving-Estimates Test for Parameter Stability," *Econometric Theory*, 11, 699–720.

CRAINICEANU, C. AND VOGELSANG, T. (2007): "Nonmonotonic Power for Tests of a Mean Shift in a Time Series," *Journal of Statistical Computation and Simulation*, 77, 457-476.

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

DAVYDOV, Y. (1973): "Mixing Conditions for Markov Chains," *Theory of Probability and Its Applications*, 18, 312–328.

DARLING, D. (1955): "The Cramér–Smirnov Test in the Parametric Case," *The Annals of Mathematical Statistics*, 28, 823–838.

DENG, A. AND PERRON, P. (2008): "A Non-local Perspective on the Power Properties of the CUSUM and CUSUM of Squares Tests for Structural Change," *Journal of Econometrics*, 142, 212-240.

DIEBOLD, F., GUNTHER, T., AND TAY, A. (1998): "Evaluating Density Forecasts with Applications to Financial Risk Management," *International Economic Review*, 76, 967–974.

DODD, E. (1942): "Certain Tests for Randomness Applied to Data Grouped into Small Sets," *Econometrica*, 10, 249–257.

DONSKER, M. (1951): "An Invariance Principle for Certain Probability Limit Theorems," *Memoirs of American Mathematics Society*, 6.

DUFOUR, J.-M. (1981): "Rank Tests for Serial Dependence," *Journal of Time Series Analysis*, 2, 117–128.

DURBIN, J. (1973): "Weak Convergence of the Sample Distribution Function When Parameters are Estimated," *The Annals of Statistics*, 1, 279–290.

DURRETT, R. (1996): *Probability: Theory and Examples*, Belmont: Duxbury Press.

FAMA, E. (1965): "The Behavior of Stock Market Prices," *The Journal of Business*, 38, 34–105.

FAN, Y. AND LI, Q. (2000): "Kernel-Based Tests Versus Bierens' ICM Tests," *Econometric Theory*, 16, 1016–1041.

FELLER, W. (1951): "The Asymptotic Distribution of the Range of Sums of Independent Random Variables," *The Annals of Mathematical Statistics*, 22, 427–432.

GOODMAN, L. (1958): "Simplified Runs Tests and Likelihood Ratio Tests for Markov Chains," *Biometrika*, 45, 181–97.

GRANGER, C. (1963): "A Quick Test for Serial Correlation Suitable for Use with Nonstationary Time Series," *Journal of the American Statistical Association*, 58, 728–736.

GRENANDER, U. (1981): *Abstract Inference*. New York: Wiley.

HECKMAN, J. (2001): "Micro Data, Heterogeneity, and the Evaluation of Public Policy: Nobel Lecture," *The Journal of Political Economy*, 109, 673–746.

HENZE, N. (1996): "Empirical Distribution Function Goodness-of-Fit Tests for Discrete Models," *Canadian Journal of Statistics*, 24, 81–93.

HERRNDORF, N. (1984): "A Functional Central Limit Theorem for Weakly Dependent Sequences of Random Variables," *The Annals of Probability*, 12, 141–153.

HONG, Y. (1999): "Hypothesis Testing in Time Series via the Empirical Characteristic Function: A Generalized Spectral Density Approach," *Journal of the American Statistical Association*, 84, 1201–1220.

HONG, Y. AND WHITE, H. (2005): "Asymptotic Distribution Theory for Nonparametric Entropy Measures of Serial Dependence," *Econometrica*, 73, 837–901.

JAIN, N. AND KALLIANPUR, G. (1970): "Norm Convergent Expansions for Gaussian Processes in Banach Spaces," *Proceedings of the American Mathematical Society*, 25, 890–895.

JAIN, N. AND MARCUS, M. (1975): "Central Limit Theorem for $C(S)$–valued Random Variables," *Journal of Functional Analysis*, 19, 216–231.

JOGDEO, K. (1968): "Characterizations of Independence in Certain Families of Bivariate and Multivariate Distributions," *Annals of Mathematical Statistics*, 39, 433-441.

KARLIN, S. AND TAYLOR, H. (1975): *A First Course in Stochastic Processes*. San Diego: Academic Press.

KOCHERLAKOTA, S. AND KOCHERLAKOTA, K. (1986): "Goodness-of-Fit Tests for Discrete Distributions," *Communications in Statistics–Theory and Methods*, 15, 815–829.

KRIVYAKOV, E., MARTNOV, G., AND TYURIN, Y. (1977): "On the Distribution of the ω^2 Statstics in the Multi-Dimensional Case," *Theory of Probability and Its Applications*, 22, 406–410.

KUAN, C.-M. AND HORNIK, K. (1995): "The Generalized Fluctuation Test: a Unifying View," *Econometric Review*, 14, 135–161.

LOÈVE, M. (1978): *Probability Theory II*. New York: Springer-Verlarg.

MOOD, A. (1940): "The Distribution Theory of Runs," *The Annals of Mathematical Statistics*, 11, 367–392.

NUMMELIN, E. AND TUOMINEN, P. (1982): "Geometric Ergodicity of Harris Recurrent Markov Chains with Applications to Renewal Theory," *Stochastic Processes and Their Applications*, 12, 187–202.

NUMMELIN, E. AND TWEEDIE, R. (1978): "Geometric Ergodicity and R-positivity for General Markov Chains," *Annals of Probability*, 6, 404–420.

PHILLIPS, P. (1998): "New Tools for Understanding Spurious Regressions," *Econometrica*, 66, 1299–1326.

PINKSE, J. (1998): "Consistent Nonparametric Testing for Serial Independence," *Journal of Econometrics*, 84, 205–231.

PLOBERGER, W. AND KRÄMER, W. (1992): "The CUSUM Test with OLS Residuals," *Econometrica*, 60, 271–285.

PLOBERGER, W., KRÄMER, W., AND KONTRUS, K. (1989): "A New Test for Structural Stability in the Linear Regression Model," *Journal of Econometrics*, 40, 307–318.

ROBINSON, P. (1991) "Consistent Nonparametric Entropy-Based Testing," *Review of Economic Studies*, 58, 437–453.

RUEDA, R., PÉREZ-ABREU, V. AND O'REILY, F. (1991): "Goodness-of-Fit Test for the Poisson Distribution Based on the Probability Generating Function," *Communications in Statistics–Theory and Methods* , 20, 3093–3110.

SEN, P. (1980): "Asymptotic Theory of Some Tests for a Possible Change in the Regressional Slope Occurring at an Unknown Time Point," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 52, 203–218.

SKAUG, H. AND TJØSTHEIM, D. (1996): "Measures of Distance between Densities with Application to Testing for Serial Independence," *Time Series Analysis in Memory of E. J. Hannan*, ed. P. Robinson and M. Rosenblatt. New York: Springer-Verlag, 363–377.

STINCHCOMBE, M. AND WHITE, H. (1998): "Consistent Specification Testing with Nuisance Parameters Present Only under the Alternative," *Econometric Theory*, 14, 295–324.

SUKHATME, P. (1972): "Fredholm Determinant of a Positive Definite Kernel of a Special Type and Its Applications," *The Annals of Mathematical Statistics*, 43, 1914–1926.

VAN DER VAART, A. AND WELLNER, J. (1996): *Weak Convergence and Empirical Processes with Applications to Statistics*. New York: Springer-Verlag.

WALD, A. AND WOLFOWITZ, J. (1940): "On a Test Whether Two Samples are from the Same Population," *Annals of Mathematical Statistics*, 2, 147–162.

Table I: ASYMPTOTIC CRITICAL VALUES OF THE TEST STATISTICS

Statistics \ Level		1%	5%	10%	Statistics \ Level		1%	5%	10%
$\mathcal{T}^p_{1,n}(\mathbb{S}_1)$	$p = 0.1$	0.2474	0.1914	0.1639	$\widetilde{\mathcal{T}}^p_{1,n}(\mathbb{S}_1)$	$p = 0.1$	0.2230	0.1727	0.1420
	0.3	0.5892	0.4512	0.3828		0.3	0.5004	0.3842	0.3225
	0.5	0.8124	0.6207	0.5239		0.5	0.6413	0.4886	0.4092
	0.7	0.9007	0.6841	0.5763		0.7	0.6065	0.4632	0.3889
	0.9	0.7052	0.5329	0.4478		0.9	0.3066	0.2356	0.1973
$\mathcal{T}^p_{\infty,n}(\mathbb{S}_1)$	$p =$0.1	0.7483	0.5683	0.4750	$\widetilde{\mathcal{T}}^p_{\infty,n}(\mathbb{S}_1)$	$p =$0.1	0.7454	0.5677	0.4750
	0.3	1.3517	1.0237	0.8582		0.3	1.3239	1.0069	0.8441
	0.5	1.6846	1.2818	1.0728		0.5	1.5909	1.1990	1.0091
	0.7	1.7834	1.3590	1.4000		0.7	1.5019	1.1360	0.9617
	0.9	1.3791	1.0486	0.8839		0.9	0.7912	0.6028	0.5060
$\mathcal{T}^s_{1,n}$	$s =$ -0.5	0.3114	0.2439	0.2152	$\widetilde{\mathcal{T}}^s_{1,n}$	$s =$ -0.5	0.2197	0.1785	0.1587
	-0.3	0.1698	0.1330	0.1164		-0.3	0.1124	0.0905	0.0803
	-0.1	0.0514	0.0402	0.0351		-0.1	0.0313	0.0254	0.0225
	0.1	0.0466	0.0361	0.0315		0.1	0.0262	0.0210	0.0187
	0.3	0.1246	0.0957	0.0836		0.3	0.0631	0.0504	0.0446
	0.5	0.1769	0.1356	0.1183		0.5	0.0780	0.0625	0.0552
$\mathcal{T}^s_{\infty,n}$	$s =$ -0.5	0.8091	0.6885	0.6160	$\widetilde{\mathcal{T}}^s_{\infty,n}$	$s =$ -0.5	0.6229	0.5319	0.4799
	-0.3	0.4349	0.3650	0.3267		-0.3	0.3107	0.2668	0.2412
	-0.1	0.1282	0.1072	0.0960		-0.1	0.0864	0.0734	0.0670
	0.1	0.1124	0.0939	0.0840		0.1	0.0714	0.0604	0.0547
	0.3	0.2898	0.2416	0.2154		0.3	0.1718	0.1454	0.1306
	0.5	0.3962	0.3304	0.2948		0.5	0.2175	0.1869	0.1684
$\mathcal{T}_{1,n}(\mathbb{S}_1)$		0.4571	0.3547	0.3124	$\widetilde{\mathcal{T}}_{1,n}(\mathbb{S}_1)$		0.3080	0.2440	0.2187
$\mathcal{T}_{\infty,n}(\mathbb{S}_1)$		2.2956	1.9130	1.7331	$\widetilde{\mathcal{T}}_{\infty,n}(\mathbb{S}_1)$		2.0411	1.7101	1.5615
$\mathcal{T}_{1,n}(\mathbb{S}_2)$		0.1219	0.0955	0.0836	$\widetilde{\mathcal{T}}_{1,n}(\mathbb{S}_2)$		0.0725	0.0590	0.0523
$\mathcal{T}_{\infty,n}(\mathbb{S}_2)$		0.8091	0.6885	0.6160	$\widetilde{\mathcal{T}}_{\infty,n}(\mathbb{S}_2)$		0.6229	0.5319	0.4799

Table II: LEVEL SIMULATION AT 5% LEVEL (IN PERCENT, 10,000 ITERATIONS)

Statistics \ n		DGP 1.1 100	300	500
$\widetilde{\mathcal{T}}^p_{1,n}(\mathbb{S}_1)$	$p=0.1$	4.05	4.71	5.09
	0.3	5.02	4.95	4.61
	0.5	5.22	4.74	4.87
	0.7	5.34	4.90	5.31
	0.9	4.35	6.63	6.07
$\widetilde{\mathcal{T}}^p_{\infty,n}(\mathbb{S}_1)$	$p=0.1$	3.86	4.44	4.49
	0.3	3.54	4.65	4.59
	0.5	5.42	4.40	5.05
	0.7	7.00	5.03	5.08
	0.9	4.21	6.27	4.71
$\widetilde{\mathcal{T}}^s_{1,n}$	$s=$ -0.5	4.36	4.72	4.26
	-0.3	4.32	4.77	4.08
	-0.1	4.23	4.71	3.83
	0.1	4.19	4.64	3.68
	0.3	4.01	4.08	3.25
	0.5	3.85	3.63	2.55
$\widetilde{\mathcal{T}}^s_{\infty,n}$	$s=$ -0.5	4.90	6.07	5.95
	-0.3	4.92	6.11	5.86
	-0.1	5.15	6.29	6.13
	0.1	4.89	6.22	5.94
	0.3	5.11	5.76	6.06
	0.5	5.05	5.18	5.57
$\widetilde{\mathcal{T}}_{1,n}(\mathbb{S}_1)$		3.71	4.21	3.60
$\widetilde{\mathcal{T}}_{\infty,n}(\mathbb{S}_1)$		4.91	6.40	6.65
$\widetilde{\mathcal{T}}_{1,n}(\mathbb{S}_2)$		4.06	4.65	4.12
$\widetilde{\mathcal{T}}_{\infty,n}(\mathbb{S}_2)$		4.77	5.94	5.74
R_n[a]		4.7		
ST_n[a]		5.1		
HW_n[a]		6.5		

[a]These results are those given in Hong and White (2005). Their number of replications is 1,000.

Table III: POWER SIMULATION AT 5% LEVEL (IN PERCENT, 3,000 ITERATIONS)

		DGP 1.2		DGP 1.3		DGP 1.4		DGP 1.5	
Statistics \ n		100	200	100	200	100	200	100	200
$\widetilde{\mathcal{T}}^{p}_{1,n}(\mathbb{S}_1)$	$p=$ 0.1	20.67	33.17	29.50	45.00	27.53	47.47	5.80	9.27
	0.3	33.27	59.37	8.43	10.57	8.77	12.07	9.77	15.40
	0.5	36.60	65.17	5.33	5.47	5.20	5.80	18.77	32.93
	0.7	34.13	60.30	9.33	11.37	5.37	6.10	76.17	97.33
	0.9	23.77	38.07	30.30	51.57	15.40	23.97	75.37	97.10
$\widetilde{\mathcal{T}}^{p}_{\infty,n}(\mathbb{S}_1)$	$p=$ 0.1	6.60	6.67	7.10	6.33	5.03	5.37	3.63	4.73
	0.3	10.57	23.97	3.60	5.40	3.70	5.40	5.70	11.00
	0.5	24.07	42.87	5.77	6.13	4.90	6.33	16.43	26.70
	0.7	29.53	41.67	9.80	6.87	6.93	4.20	72.97	92.77
	0.9	22.47	22.63	27.30	32.67	14.43	13.30	72.30	89.87
$\widetilde{\mathcal{T}}^{s}_{1,n}$	$s=$ -0.5	62.56	99.83	13.53	90.90	8.40	11.63	81.73	99.30
	-0.3	67.70	99.86	14.66	92.06	9.33	12.96	80.30	99.33
	-0.1	70.06	99.73	15.66	92.73	10.53	15.30	76.86	98.90
	0.1	71.30	99.70	16.46	93.40	11.86	18.10	69.66	97.73
	0.3	70.30	99.60	17.30	93.46	13.93	23.50	58.50	94.00
	0.5	67.36	99.46	19.93	93.33	18.63	32.80	44.90	82.73
$\widetilde{\mathcal{T}}^{s}_{\infty,n}$	$s=$ -0.5	43.80	79.50	7.83	13.46	5.70	6.10	74.46	98.73
	-0.3	49.66	83.96	8.50	13.66	6.60	6.76	71.40	98.03
	-0.1	55.63	87.80	9.83	15.33	8.03	9.16	66.43	96.53
	0.1	59.70	89.06	11.20	17.66	10.10	14.43	55.06	91.33
	0.3	61.40	88.30	14.33	24.86	15.46	25.60	40.76	79.13
	0.5	58.40	84.83	22.36	37.23	23.60	46.33	22.43	50.30
$\widetilde{\mathcal{T}}_{1,n}(\mathbb{S}_1)$		59.73	90.93	11.96	23.13	9.40	13.73	14.60	27.33
$\widetilde{\mathcal{T}}_{\infty,n}(\mathbb{S}_1)$		27.30	61.56	5.83	10.70	4.70	6.46	18.13	34.33
$\widetilde{\mathcal{T}}_{1,n}(\mathbb{S}_2)$		97.03	99.86	16.36	26.36	75.60	86.06	82.30	92.40
$\widetilde{\mathcal{T}}_{\infty,n}(\mathbb{S}_2)$		44.03	80.50	7.20	13.60	5.56	7.26	13.76	18.56
R_n[a]		13.8	25.4	26.4	52.2	15.0	7.2	59.8	75.4
ST_n[a]		12.4	22.0	61.2	90.0	27.8	52.0	81.6	98.4
HW_n[a]		14.0	27.0	37.6	67.6	20.6	35.2	69.6	95.6

[a]These results are those given in Hong and White (2005). Their number of replications is 500.

Table IV: POWER SIMULATION AT 5% LEVEL (IN PERCENT, 3,000 ITERATIONS)

		DGP 1.6		DGP 1.7		DGP 1.8		DGP 1.9	
Statistics $\setminus$ n		100	200	100	200	100	200	100	200
$\widetilde{\mathcal{T}}^{p}_{1,n}(\mathbb{S}_1)$	$p = 0.1$	1.30	6.83	1.00	5.33	2.77	3.73	27.83	47.70
	0.3	9.93	16.33	18.90	33.67	17.73	32.57	39.47	64.23
	0.5	8.33	10.87	9.27	11.37	31.77	59.80	43.13	68.00
	0.7	30.87	57.17	18.07	28.73	33.60	58.43	39.97	62.13
	0.9	26.43	41.27	24.00	39.03	22.37	34.63	26.30	37.50
$\widetilde{\mathcal{T}}^{p}_{\infty,n}(\mathbb{S}_1)$	$p = 0.1$	3.73	5.73	4.33	4.30	3.77	3.67	9.50	14.53
	0.3	6.00	10.47	8.93	16.23	5.83	12.80	22.23	44.33
	0.5	8.47	9.90	8.10	9.83	20.47	36.93	34.60	59.23
	0.7	33.63	47.20	16.67	18.17	30.03	39.93	37.83	54.90
	0.9	25.80	29.60	22.53	22.70	20.63	19.97	24.83	29.80
$\widetilde{\mathcal{T}}^{s}_{1,n}$	$s =$ -0.5	25.13	60.93	21.96	51.46	50.13	83.03	56.36	80.73
	-0.3	25.13	60.10	24.13	55.93	54.26	86.16	57.43	81.10
	-0.1	22.40	56.20	25.16	58.93	55.93	87.53	57.80	81.46
	0.1	20.13	52.10	25.10	59.96	56.20	87.26	58.46	81.96
	0.3	16.90	43.73	23.10	56.46	53.33	84.90	59.13	82.13
	0.5	12.36	32.30	17.53	47.60	48.93	80.23	59.50	82.20
$\widetilde{\mathcal{T}}^{s}_{\infty,n}$	$s =$ -0.5	23.13	54.23	20.80	43.36	40.46	75.83	55.70	80.06
	-0.3	22.03	50.46	21.50	46.80	45.56	79.86	57.20	80.83
	-0.1	20.50	45.50	23.03	49.93	49.60	82.36	57.83	81.53
	0.1	17.43	38.26	21.96	48.36	51.20	82.73	58.56	82.00
	0.3	14.16	30.10	20.50	45.70	49.70	80.60	59.33	82.40
	0.5	9.83	21.40	14.90	38.26	43.86	71.43	59.53	82.56
$\widetilde{\mathcal{T}}_{1,n}(\mathbb{S}_1)$		25.90	59.36	19.43	46.10	47.96	81.50	56.33	79.96
$\widetilde{\mathcal{T}}_{\infty,n}(\mathbb{S}_1)$		22.36	50.56	14.50	29.10	26.36	55.80	50.26	75.60
$\widetilde{\mathcal{T}}_{1,n}(\mathbb{S}_2)$		22.80	56.40	23.63	57.23	54.90	86.40	58.70	81.40
$\widetilde{\mathcal{T}}_{\infty,n}(\mathbb{S}_2)$		22.80	53.70	19.66	41.50	39.63	76.36	54.60	79.73
R_n[a]		31.8	65.2	24.6	80.8	14.2	34.6	60.2	84.0
ST_n[a]		34.8	72.8	25.8	86.8	13.4	23.8	55.8	79.8
HW_n[a]		34.0	74.0	25.6	85.4	17.0	26.2	60.8	84.6

[a]These results are those given in Hong and White (2005). Their number of replications is 500.

Table V: LEVEL SIMULATION AT 5% LEVEL (IN PERCENT, 10,000 ITERATIONS)

		DGP 2.1			DGP 2.2		
Statistics $\setminus$ n		100	300	500	100	300	500
$\widetilde{\mathcal{T}}^p_{1,n}(\mathbb{S}_1)$	$p = 0.1$	4.12	4.88	5.02	4.25	5.01	5.02
	0.3	5.12	5.22	5.07	5.20	5.12	5.23
	0.5	4.76	5.51	5.33	5.02	5.04	5.38
	0.7	5.81	5.63	5.19	5.14	4.51	5.32
	0.9	3.71	6.63	5.92	4.35	6.87	6.02
$\widetilde{\mathcal{T}}^p_{\infty,n}(\mathbb{S}_1)$	$p = 0.1$	3.86	6.38	6.04	4.33	6.44	6.32
	0.3	5.56	4.56	6.10	5.68	4.82	6.38
	0.5	5.17	5.34	5.60	5.60	5.06	5.77
	0.7	7.44	5.70	4.88	6.89	4.54	5.15
	0.9	3.58	6.25	7.43	4.14	6.47	7.49
$\widetilde{\mathcal{T}}^s_{1,n}$	$s =$ -0.5	4.37	4.66	4.32	4.18	4.71	4.21
	-0.3	4.27	4.82	4.40	4.29	4.75	4.13
	-0.1	3.98	4.57	4.06	4.18	4.47	3.97
	0.1	3.90	4.46	3.78	4.02	4.29	3.67
	0.3	3.46	3.95	3.30	3.86	3.96	3.04
	0.5	3.34	3.31	2.72	3.50	3.52	2.57
$\widetilde{\mathcal{T}}^s_{\infty,n}$	$s =$ -0.5	4.42	5.96	6.13	4.63	6.16	6.12
	-0.3	4.49	5.73	6.12	4.64	6.07	6.06
	-0.1	4.64	5.87	6.27	4.62	5.94	6.19
	0.1	4.41	5.66	6.28	4.53	5.65	6.04
	0.3	4.47	5.53	5.84	4.79	5.27	5.77
	0.5	4.78	4.95	5.21	4.78	4.95	5.22
$\widetilde{\mathcal{T}}_{1,n}(\mathbb{S}_1)$		3.69	4.34	3.50	3.62	3.90	3.46
$\widetilde{\mathcal{T}}_{\infty,n}(\mathbb{S}_1)$		5.22	6.65	6.50	5.00	5.98	6.97
$\widetilde{\mathcal{T}}_{1,n}(\mathbb{S}_2)$		3.94	4.33	3.98	3.71	4.41	4.09
$\widetilde{\mathcal{T}}_{\infty,n}(\mathbb{S}_2)$		4.87	5.53	6.02	4.89	5.62	6.25
M_n		2.89	4.43	4.75	3.98	5.33	5.62
RE_n		7.89	4.14	3.11	71.90	70.70	31.01
RR_n		9.28	4.30	2.74	64.39	60.78	66.15
SupW_n		4.77	4.41	4.57	2.93	0.94	1.31
AvgW_n		5.81	5.29	5.10	2.60	1.69	2.22
ExpW_n		5.34	5.35	5.04	1.78	2.35	3.13
$RECUSUM_n$		1.65	3.07	3.68	3.41	3.86	3.64
$O(N)LSCUSUM_n$		2.68	4.20	4.01	28.91	30.22	55.50

Table VI: POWER SIMULATION AT 5% LEVEL (IN PERCENT, 3,000 ITERATIONS)

		DGP 2.3		DGP 2.4		DGP 2.5		DGP 2.6		DGP 2.7	
Statistics $\setminus$ n		100	200	100	200	100	200	100	200	100	200
$\widetilde{\mathcal{T}}^{p}_{1,n}(\mathbb{S}_1)$	$p=$ 0.1	14.20	22.73	18.13	29.60	22.20	39.93	16.70	24.70	35.67	59.00
	0.3	19.06	31.86	27.47	48.43	20.53	37.60	7.80	10.37	22.87	39.23
	0.5	21.00	37.70	29.03	55.43	21.83	37.77	5.80	7.77	19.43	35.07
	0.7	19.23	31.03	29.03	51.70	22.33	38.20	7.67	10.17	27.37	48.37
	0.9	10.70	14.66	19.37	32.77	22.63	39.23	17.30	24.07	43.27	71.47
$\widetilde{\mathcal{T}}^{p}_{\infty,n}(\mathbb{S}_1)$	$p=$ 0.1	4.10	4.30	5.90	5.53	5.50	6.80	6.07	5.03	7.57	9.47
	0.3	4.43	8.70	7.70	21.80	6.87	16.43	4.77	8.40	7.20	18.10
	0.5	11.36	17.56	18.23	33.47	13.80	20.50	5.93	6.50	12.63	20.10
	0.7	15.56	16.53	25.27	44.40	18.83	32.30	7.30	9.83	22.47	41.37
	0.9	10.23	8.10	17.37	39.73	20.33	44.47	15.27	25.77	40.10	74.80
$\widetilde{\mathcal{T}}^{s}_{1,n}$	$s=$ -0.5	28.76	51.46	47.36	83.30	37.53	67.60	9.20	17.23	47.83	76.43
	-0.3	33.46	59.10	53.60	86.73	42.40	73.60	10.60	19.63	51.56	81.03
	-0.1	39.40	66.50	57.76	89.13	46.46	77.33	11.96	21.56	54.83	82.93
	0.1	46.20	73.23	60.73	90.73	48.80	80.13	13.36	23.63	57.36	84.66
	0.3	50.73	79.46	61.06	91.00	49.83	81.00	14.53	24.16	58.26	85.16
	0.5	58.43	85.13	60.23	90.00	50.03	80.10	15.66	26.00	59.20	84.03
$\widetilde{\mathcal{T}}^{s}_{\infty,n}$	$s=$ -0.5	22.60	43.50	33.90	68.30	22.96	49.46	6.10	10.56	31.66	59.56
	-0.3	28.70	53.10	41.16	75.53	28.20	56.63	7.06	12.56	36.50	64.93
	-0.1	37.13	62.40	47.10	80.76	33.66	63.53	8.93	14.90	40.60	69.56
	0.1	44.30	70.36	51.76	83.30	37.33	67.33	9.96	15.90	43.76	71.96
	0.3	51.06	77.03	55.50	84.86	41.23	70.16	11.73	18.43	47.63	72.53
	0.5	59.70	83.50	54.63	83.56	43.00	69.96	13.60	22.43	49.13	73.06
$\widetilde{\mathcal{T}}_{1,n}(\mathbb{S}_1)$		32.73	56.70	46.53	80.13	36.06	66.73	9.46	14.96	45.93	46.86
$\widetilde{\mathcal{T}}_{\infty,n}(\mathbb{S}_1)$		12.76	24.50	22.13	46.10	15.63	30.06	5.50	7.70	19.06	38.26
$\widetilde{\mathcal{T}}_{1,n}(\mathbb{S}_2)$		42.90	67.73	58.76	89.06	45.83	76.33	12.26	19.83	54.86	84.70
$\widetilde{\mathcal{T}}_{\infty,n}(\mathbb{S}_2)$		23.93	44.30	35.00	69.13	23.63	49.40	6.03	9.60	29.43	60.00
M_n		97.13	100.0	100.0	100.0	12.93	23.03	1.97	2.70	35.96	61.43
RE_n		99.16	100.0	100.0	100.0	100.0	100.0	94.96	100.0	79.50	89.13
RR_n		93.33	100.0	99.97	100.0	99.96	100.0	87.90	99.93	69.43	81.73
SupW_n		98.26	100.0	100.0	100.0	100.0	100.0	85.63	97.86	94.89	99.03
AvgW_n		98.26	99.96	100.0	100.0	100.0	100.0	73.76	92.60	94.87	99.02
ExpW_n		98.76	100.0	100.0	100.0	100.0	100.0	85.73	97.70	90.19	98.14
$RECUSUM_n$		87.13	99.40	6.77	9.92	9.30	11.43	14.73	19.10	15.60	20.20
$O(N)LSCUSUM_n$		99.16	100.0	19.67	24.29	22.43	26.60	15.50	22.03	83.03	92.40

Table VII: Power Simulation at 5% Level (in percent, 3,000 iterations)

		DGP 2.8		DGP 2.9		DGP 2.10		DGP 2.11		DGP 2.12	
Statistics $\setminus$ n		100	200	100	200	100	200	100	200	100	200
$\widetilde{\mathcal{T}}^p_{1,n}(\mathbb{S}_1)$	$p =$ 0.1	18.13	29.13	18.13	27.06	14.86	23.66	4.10	12.33	24.03	45.53
	0.3	27.47	49.87	25.50	43.26	20.06	37.43	37.83	63.23	20.53	34.53
	0.5	29.03	55.77	26.33	47.90	21.56	41.06	46.40	74.57	18.93	29.27
	0.7	29.03	50.97	25.46	45.06	22.70	39.90	35.87	63.67	15.20	24.10
	0.9	19.37	32.33	21.10	27.73	16.06	26.13	0.43	0.03	0.93	0.20
$\widetilde{\mathcal{T}}^p_{\infty,n}(\mathbb{S}_1)$	$p =$ 0.1	5.90	5.27	5.43	4.93	5.10	4.23	4.23	7.13	24.27	46.60
	0.3	7.70	20.83	7.86	15.83	6.43	14.50	31.27	43.87	23.07	30.87
	0.5	18.23	32.73	16.40	28.16	13.10	23.96	37.17	65.13	16.87	26.87
	0.7	25.27	44.43	22.43	28.73	19.76	25.56	38.13	56.77	17.60	20.17
	0.9	17.37	39.57	19.90	15.06	14.80	14.23	0.40	0.07	0.90	0.53
$\widetilde{\mathcal{T}}^s_{1,n}$	$s =$ -0.5	46.63	80.86	42.33	75.43	37.63	67.90	64.23	90.10	27.63	47.73
	-0.3	53.33	85.20	46.93	80.03	42.76	73.26	64.20	90.30	25.16	44.40
	-0.1	57.06	87.26	50.73	83.43	45.43	75.76	61.46	89.30	22.10	40.00
	0.1	59.73	88.63	53.53	85.30	47.36	77.46	55.23	86.33	17.76	33.00
	0.3	60.66	89.20	54.50	85.40	47.53	77.53	44.73	79.60	11.93	23.70
	0.5	60.13	88.20	53.70	83.86	46.26	75.40	31.13	64.43	7.06	13.96
$\widetilde{\mathcal{T}}^s_{\infty,n}$	$s =$ -0.5	34.30	66.83	29.60	62.03	26.00	52.93	55.86	84.30	22.80	36.33
	-0.3	40.60	73.86	34.66	69.00	31.70	60.06	55.60	84.50	21.10	35.70
	-0.1	46.56	80.56	40.30	74.23	37.00	65.90	54.46	83.53	19.46	33.33
	0.1	50.96	83.23	45.03	77.23	39.66	69.10	46.73	79.30	15.00	28.23
	0.3	53.66	84.20	47.76	78.23	42.50	70.40	37.33	70.43	10.70	21.10
	0.5	52.36	81.86	47.96	76.26	41.66	67.43	19.46	47.96	5.80	10.36
$\widetilde{\mathcal{T}}_{1,n}(\mathbb{S}_1)$		46.26	81.20	40.16	75.83	35.63	66.20	58.73	88.66	26.10	47.30
$\widetilde{\mathcal{T}}_{\infty,n}(\mathbb{S}_1)$		21.13	47.66	17.56	41.93	16.06	35.43	47.30	77.06	21.93	39.50
$\widetilde{\mathcal{T}}_{1,n}(\mathbb{S}_2)$		57.43	87.66	50.60	84.80	46.83	75.46	59.86	89.90	21.66	39.76
$\widetilde{\mathcal{T}}_{\infty,n}(\mathbb{S}_2)$		33.00	66.73	28.73	62.60	25.46	53.83	56.13	85.46	22.53	37.47
M_n		100.0	13.67	9.80	12.20	8.10	9.40	0.66	0.86	2.36	2.73
RE_n		100.0	100.0	78.06	60.00	36.23	3.12	7.77	4.91	80.70	89.16
RR_n		99.97	99.96	77.86	59.73	34.33	30.93	8.85	5.24	71.16	79.93
SupW_n		100.0	100.0	98.16	63.26	41.86	46.06	3.73	3.46	3.82	3.64
AvgW_n		100.0	100.0	92.80	50.16	34.16	32.63	4.35	3.63	3.34	3.15
ExpW_n		100.0	100.0	98.86	62.66	43.50	42.40	3.56	3.19	1.76	1.87
$RECUSUM_n$		6.77	9.20	7.03	9.50	5.46	8.50	0.23	0.39	1.26	3.10
$OLSCUSUM_n$		19.67	23.36	17.80	20.40	13.63	15.70	0.43	0.39	28.43	35.13