Econometric Theory, **29**, 2013, 419–446.
doi:10.1017/S0266466612000370

A SMOOTH TEST FOR THE EQUALITY OF DISTRIBUTIONS

ANIL K. BERA
University of Illinois at Urbana - Champaign

AUROBINDO GHOSH
Singapore Management University

ZHIJIE XIAO
Boston College

The two-sample version of the celebrated Pearson goodness-of-fit problem has been a topic of extensive research, and several tests like the Kolmogorov-Smirnov and Cramér-von Mises have been suggested. Although these tests perform fairly well as omnibus tests for comparing two probability density functions (PDFs), they may have poor power against specific departures such as in location, scale, skewness, and kurtosis. We propose a new test for the equality of two PDFs based on a modified version of the Neyman smooth test using empirical distribution functions minimizing size distortion in finite samples. The suggested test can detect the specific directions of departure from the null hypothesis. Specifically, it can identify deviations in the directions of mean, variance, skewness, or tail behavior. In a finite sample, the actual probability of type-I error depends on the relative sizes of the two samples. We propose two different approaches to deal with this problem and show that, under appropriate conditions, the proposed tests are asymptotically distributed as chi-squared. We also study the finite sample size and power properties of our proposed test. As an application of our procedure, we compare the age distributions of employees with small employers in New York and Pennsylvania with group insurance before and after the enactment of the "community rating" legislation in New York. It has been conventional wisdom that if community rating is enforced (where the group health insurance premium does not depend on age or any other physical characteristics of the insured), then the insurance market will collapse, since only older or less healthy patients would prefer group insurance. We find that there are significant changes in the age distribution in the population in New York owing mainly to a shift in location and scale.

1. INTRODUCTION

One of the old, celebrated problems in statistics is the two-sample version of Pearson (1900) goodness-of-fit problem (Lehmann, 1953; Darling, 1957). Suppose

The authors thank the co-editor, two referees, and Oliver Linton, Peter Phillips, Yanqing Fan, and seminar participants at the Tinbergen Institute, University of Amsterdam, University of Maryland, International Symposium on Econometrics of Specification Tests in 30 Years at Xiamen University, and other conferences for helpful comments. The usual disclaimers apply. Address correspondence to Zhijie Xiao, Dept. of Economics, Boston College, Chestnut Hill, MA 02467; e-mail: zhijie.xiao@bc.edu.

© Cambridge University Press 2012

https://doi.org/10.1017/S0266466612000370 Published online by Cambridge University Press

we have two samples $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ from two *unspecified* absolutely continuous distributions with cumulative distribution functions (CDFs), $F(x)$ and $G(x)$, respectively. The problem is to test the hypothesis $H_0 : F = G$. Most of the tests in the literature such as Kolmogorov-Smirnov, Cramér-von Mises, and Anderson-Darling are based on some distance measures between the two empirical distribution functions (EDFs), $F_n(x)$ and $G_m(x)$. Although widely used, the computation and implementation of these tests are not trivial, and being *omnibus* tests, they lack power in specific directions. Furthermore, they may not have good finite sample power properties (Janssen, 2000; Esconciano, 2009). A bootstrap version of the test of comparing two distributions of regression errors using an empirical process approach have been proposed by Mora and Neumeyer (2005).

We propose a test for H_0 using Neyman's (1937) smooth test principle. To motivate the proposed test, let us first consider the *one-sample* Pearson goodness-of-fit test of $H_0^{'} : F(x) = F_0(x)$, where $F_0(x)$ is a *specified* CDF with $f_0(x)$ as the corresponding probability density function (PDF). Define the probability integral transform (PIT)

$$Z_i = F_0(X_i) = \int_{-\infty}^{X_i} f_0(\omega)\, d\omega, \qquad i = 1, 2, \ldots, n. \tag{1}$$

If $H_0^{'}$ is true, then $Z_1, Z_2, \ldots, Z_n$ are independently and identically distributed (IID) as uniform random variates $U(0,1)$ irrespective of F_0. Test of $H_0^{'}$ is identical to the test of uniformity of Z in $(0,1)$ (Solomon and Stephens, 1985). Therefore, in some sense, all testing problems can be converted into testing only *one kind of hypothesis* (see Neyman, pp. 160–162; Bera and Ghosh, 2002, p. 178). Neyman considered the following *smooth* alternative to the uniform density,

$$h(z) = c(\theta) \exp\left[\sum_{j=1}^{k} \theta_j \pi_j(z)\right], \qquad 0 < z < 1, \tag{2}$$

where $c(\theta)$ is the constant of integration depending on $\theta_1, \theta_2, \ldots, \theta_k$, and $\pi_j(z)$ are orthogonal polynomials of order j satisfying

$$\int_0^1 \pi_i(z)\, \pi_j(z)\, dz = \delta_{ij} \quad \text{where } \delta_{ij} = \begin{cases} = 1, & \text{if } i = j \\ = 0, & \text{if } i \neq j. \end{cases} \tag{3}$$

Neyman called (2) a smooth alternative since with small θ's, $h(z)$ is close to, and has few intersections with, the density under the null hypothesis $U(0,1)$. Neyman (pp. 163–164) used $\pi_j(z)$ as the normalized Legendre polynomials $\pi_j(z) = a_{j0} + a_{j1} z + \cdots + a_{jj} z^j, a_{jj} \neq 0$, satisfying the orthogonality conditions in (3). Explicitly for $k = 1, 2, 3, 4$, these are: $\pi_1(z) = \sqrt{12}\left(z - \frac{1}{2}\right)$, $\pi_2(z) = \sqrt{5}\left(6\left(z - \frac{1}{2}\right)^2 - \frac{1}{2}\right)$, $\pi_3(z) = \sqrt{7}\left(20\left(z - \frac{1}{2}\right)^3 - 3\left(z - \frac{1}{2}\right)\right)$,

https://doi.org/10.1017/S0266466612000370 Published online by Cambridge University Press

$\pi_4(z) = 210\left(z-\frac{1}{2}\right)^4 - 45\left(z-\frac{1}{2}\right)^2 + \frac{9}{8}$. We can, therefore, test H_0' by testing $H_0'' : \theta_1 = \theta_2 = \cdots = \theta_k = 0$ in (2). Using the generalized Neyman-Pearson (N-P) lemma, Neyman derived the locally most powerful symmetric test for H_0'' against the alternative H_1 : At least one $\theta_j \neq 0$, for some j. Lehmann and Romano (2005, pp. 599–607) provide a lucid and up-to-date treatment of Neyman's smooth tests. Under H_0, asymptotically the test statistic

$$\Psi_k^2 = \sum_{j=1}^{k} u_j^2 \sim \chi_k^2, \quad \text{where } u_j = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \pi_j(Z_i), \qquad j = 1, \ldots, k. \tag{4}$$

Neyman suggested this test to rectify some of the drawbacks of the Pearson (1900) goodness-of-fit statistic (for composite hypotheses, see Thomas and Pierce, 1979). Bera and Ghosh showed this can also be derived as a Rao (1948) score test and presented a historical perspective.

Now turning to the problem of testing $H_0 : F = G$ in the two-sample case, let us start assuming that $F(.)$ is known. We construct a new random variable $Z = F(Y)$. The CDF of Z is given by

$$H(z) = \Pr(Z \leq z) = \Pr\left(Y \leq F^{-1}(z)\right) = G(Q(z)), \tag{5}$$

where $Q(z) = F^{-1}(z)$ is the quantile function of Z. Therefore, the PDF of Z can be written as (see Neyman, 1937, p. 161; Pearson, 1938, p. 138)

$$h(z) = \frac{d}{dz} H(z) = \frac{g(Q(z))}{f(Q(z))}, \qquad 0 < z < 1. \tag{6}$$

Although this is a ratio of two PDFs, $h(z)$ is a proper density function in the sense that $h(z) \geq 0$, $z \in (0, 1)$, and $\int_0^1 h(z)\,dz = 1$, if we assume that F and G are also strictly increasing functions (for further details, see Bera, Ghosh, and Xiao, 2010). Under $H_0 : F = G$, $h(z) = 1$; i.e., $Z \sim U(0, 1)$. Under the alternative hypothesis $H_1 : F \neq G$, $h(z)$ will differ from 1, providing a basis for the Neyman smooth test. Under the alternative, we take $h(z)$ as given in (2) and test $\theta_1 = \theta_2 = \cdots = \theta_k = 0$. Therefore, the test utilizes (6), which looks more like a "likelihood (density) ratio."

If $F(.)$ is known, following (4), the Neyman (1937) smooth statistic for testing $H_0 : F = G$ can be written as

$$\Psi_k^2 = \sum_{j=1}^{k} u_j^2, \quad \text{where } u_j = \frac{1}{\sqrt{m}} \sum_{i=1}^{m} \pi_j(Z_i), \tag{7}$$

with $Z_i = F(Y_i)$. Since $F(.)$ is unknown, to implement our test we need to replace Z_i by an estimate, $\hat{Z}_i, i = 1, 2, \ldots, m$. Once we replace Z_i by $\hat{Z}_i$ in Ψ_k^2,

https://doi.org/10.1017/S0266466612000370 Published online by Cambridge University Press

we need to reconsider its asymptotic distribution, and that is one of the central contributions of this paper.

The rest of the paper is organized as follows. Section 2 contains the main results of the paper. In Section 2.1 we first derive the asymptotic distribution of the proposed test statistic. Optimality criteria for the finite sample implementation of the test is discussed in Section 2.2. The general case with relaxed condition on the relative sample sizes is studied in Section 2.3. In Section 3 we apply the proposed test for comparing two age distributions in New York and Pennsylvania small group insurance markets. Finite sample behavior of the test and sample size selection are investigated in Section 4 through Monte Carlo experiments. Finally, the paper is concluded in Section 5 with a summary of our results and indications of future research.

2. THE MAIN RESULTS

2.1. Two-Sample Smooth Test

We consider two samples of n and m observations $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^m$ from *unspecified* absolutely continuous distributions with CDFs $F(x)$ and $G(x)$, and we test the hypothesis $H_0 : F = G$. Without loss of generality, we assume that $m \leq n$.

Without knowledge about the distribution function F, $Z_i = F(Y_i)$ is unknown, and thus Ψ_k^2 in (7) is not feasible in practice. We, therefore, consider using the EDF

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq x), \tag{8}$$

in place of F, to construct

$$\hat{Z}_i = F_n(Y_i) = \frac{1}{n}\sum_{l=1}^{n} I(X_l \leq Y_i), \qquad i = 1, \ldots, m, \tag{9}$$

where $I(.)$ is an indicator function. Substituting Z_i by $\hat{Z}_i$ in (7), we obtain the feasible version of the statistic,

$$\widehat{\Psi}_k^2 = \sum_{j=1}^{k} \hat{u}_j^2, \quad \text{where } \hat{u}_j = \frac{1}{\sqrt{m}}\sum_{i=1}^{m} \pi_j\left(\hat{Z}_i\right). \tag{10}$$

As noted earlier, the infeasible Ψ_k^2 converges to a χ^2-distribution with degree of freedom k and has all the local optimality properties of Rao's score test. We now show that under certain conditions, $\hat{\Psi}_k^2$ has the same limiting distribution. To achieve that goal, we need to demonstrate that the departure due to replacing $F(.)$ by the EDF $F_n(.)$ is asymptotically negligible. For convenience of asymptotic analysis, we make the following assumptions.

https://doi.org/10.1017/S0266466612000370 Published online by Cambridge University Press

Assumption 1. $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^m$ are *independent and identically distributed* with CDFs $F(x)$ and $G(x)$, respectively, and are independent of each other.

Assumption 2. $\mathbb{E}\left(\pi_j(Z_i)^2\right) < \infty$, for $j = 1, \ldots, k$.

We state the result in Theorem 1, proof of which is given in the Appendix.

THEOREM 1. *Under Assumptions 1 and 2 and the null hypothesis that* $F = G$, *if* $\frac{(\log\log n)m}{n} \to 0$ *as* $m, n \to \infty$, $\hat{\Psi}_k^2 \Rightarrow \chi_k^2$.

From Theorem 1 we can see that the size of the sample used in estimating the distribution function F should be larger in magnitude than the test sample size. If we estimate the CDF of X based on observations $\{X_i\}_{i=1}^n$, F_n converges to F at the rate $\sqrt{n}$ (pointwise), i.e., the estimation error in $\hat{Z}_i$ is of the order $n^{-1/2}$. Thus, as $\pi_j(.)'s$ can be shown to be first-order Lipschitz continuous, the accumulated estimation error in $\frac{1}{\sqrt{m}}\sum_{i=1}^m \pi_j\left(\hat{Z}_i\right)$ has the order $\sqrt{m}/\sqrt{n}$, which goes to zero if n increases to ∞ at a rate faster than m. Thus, to obtain consistent tests, we require that the sample size n used in estimating the empirical distribution should be larger than m, the test sample size. Incidentally, similar problem arises in various statistical procedures. For example, in simulation-based inference (e.g., Gourieroux and Monfort, 1996) the conditional moments are estimated based on simulations. If the number of simulations increases fast enough relative to the sample size, the resulting estimator has the same asymptotic distribution as the estimator based on known conditional moments.

2.2. Relative Magnitude of the Two Sample Sizes

Theorem 1 provides an upper bound for m given n, so that $\hat{\Psi}_k^2$ is asymptotically equivalent to Ψ_k^2. Under this condition, a wide range of sample sizes can be chosen, and all provide asymptotically equivalent tests, although the finite sample performance of these tests may differ substantially. A natural question is: What is the optimal rate of m relative to n that minimizes the order of size distortion?

THEOREM 2. *Under the null hypothesis,*

$$\hat{\Psi}_k^2 = \chi_k^2 + O_p\left(\frac{1}{\sqrt{m}}\right) + O_p\left(\sqrt{\frac{m}{n}}\right),$$

and thus the optimal relative magnitude of m and n that minimizes the size distortion is $m = O\left(n^{1/2}\right)$.

Heuristically, we can decompose $\frac{1}{\sqrt{m}}\sum_{i=1}^m \pi_j\left(\hat{Z}_i\right)$ into

$$\frac{1}{\sqrt{m}}\sum_{i=1}^m \pi_j(Z_i) + \frac{1}{\sqrt{m}}\sum_{i=1}^m \left[\pi_j\left(\hat{Z}_i\right) - \pi_j(Z_i)\right]. \tag{11}$$

https://doi.org/10.1017/S0266466612000370 Published online by Cambridge University Press

The first part of (11), $\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\pi_j(Z_i)$, converges to a standard normal variate. The larger the m, the faster it goes to the limiting normal distribution. The second part of (11), $\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\left[\pi_j\left(\hat{Z}_i\right)-\pi_j(Z_i)\right]$, which is the error term coming from estimating $F(.)$, converges to zero under the conditions in Theorem 1. The larger the n relative to m, the smaller the term. To optimize the sampling properties of the test, a trade-off has to be made to balance these two components, giving an optimal relative magnitude between m and n.

An exact analytical formula of the optimal relative sample sizes will be dependent on the exact formulation of the higher-order terms, and to the best of our knowledge, cannot be obtained. Theorem 2 gives the optimal relative magnitude between m and n, which substantially narrows the range of choices for the sample sizes. To minimize the distortion coming from estimating the distribution function, we prefer n to be large relative to m. On the other hand, to obtain fast convergence to χ^2 in the limit, we want a large m. Thus, a trade-off has to be made to minimize size distortion. We balance these two terms so that they are of the same order of magnitude, giving the optimal relative magnitude as $m = O(\sqrt{n})$. Monte Carlo experiment results reported in Section 4 indicate that a simple rule of thumb, such as $m = \sqrt{n}$ based on Theorem 2, provides satisfactory results.

Alternatively, we know that $\{X_l\}_{l=1}^{n}$ are from the same distribution with CDF F. We may divide the index set $\mathcal{N} = \{1, \ldots, n\}$ into two mutually exclusive and exhaustive sets $\mathcal{N}_1$ and $\mathcal{N}_2$ with cardinalities n_1 and n_2, with $n_1 + n_2 = n$, and define the training set

$$\mathcal{Z}_1 = \{(X_j), j \in \mathcal{N}_1\}$$

and the testing set

$$\mathcal{Z}_2 = \{(X_j), j \in \mathcal{N}_2\}.$$

Then estimate $F(\cdot)$ using data $\mathcal{Z}_1$ and construct

$$F_{n_1}(X_i) = \frac{1}{n_1}\sum_{j \in \mathcal{N}_1} I\left(X_j \leq X_i\right), \quad \text{for } i \in \mathcal{N}_2.$$

Notice that $\mathcal{Z}_1$ and $\mathcal{Z}_2$ are from the same distribution with CDF F, $F(X_i)$ $(i \in \mathcal{N}_2)$ are uniformly distributed, and $F_{n_1}(X_i)$ provides an estimator for the uniform distribution. Hence, we can calculate the estimated $\hat{\Psi}_k^2$ based on the EDF $F_{n_1}(X_i)$, $i = 1, 2, \ldots, n_2$, that should have an asymptotic χ^2 distribution with k degrees of freedom. Our objective is to minimize some distance between the empirical distribution function of $\hat{\Psi}_k^2$ and the χ_k^2 CDF in finite samples; i.e., it is equivalent to the solution for either x or t,

$$\arg\min_{n_2} d_{n_2}^r\left(\hat{F}_{\hat{\Psi}_k^2}(x), F_{\chi_k^2}(x)\right) \equiv \arg\min_{n_2} d_{n_2}^r\left(F_{\chi_k^2}\left(\hat{F}_{\hat{\Psi}_k^2}^{-1}(t)\right), t\right).$$

https://doi.org/10.1017/S0266466612000370 Published online by Cambridge University Press

If we take the criterion function $d^r_{n_2}(.)$ to be the Anderson-Darling statistic over r replications, the optimal sample size for the testing sample n_2 is a solution to

$$\arg\min_{n_2} -r + \frac{1}{r}\sum_{i=1}^{r}(2i-1)\left[\log(p_{(i)}) + \log\left(1-p_{(r-i+1)}\right)\right], \tag{12}$$

$$\text{where} \quad p_{(i)} = F_{\chi^2_k}\left(\hat{F}^{-1}_{\hat{\Psi}^2_k}\left(\hat{\Psi}^2_{(i)}\right)\right) \quad \text{and} \quad \hat{\Psi}^2_{(1)} \le \hat{\Psi}^2_{(2)} \le \cdots \le \hat{\Psi}^2_{(r)}.$$

For each value of n_2, we can calculate the above criterion function. We choose n_2 that minimizes the above criterion. Finally, we choose the test sample size $m = (n_2/n_1) \times n$.

2.3. The General Case

The test $\hat{\Psi}^2_k$ that we proposed in the previous section is an asymptotic test and, by appropriately choosing the relative sample sizes (so that one sample is smaller than the other), has an asymptotic χ^2_k distribution under the null hypothesis. The test is asymptotically distribution-free (ADF) and thus remains asymptotically equivalent to the original one-sample version smooth test Ψ^2_k. In some applications, we may have two samples with similar sample sizes, i.e., the case where m and n are of the same magnitude, say $m = \lambda n$, where $\lambda \in (0, 1]$. In this case, the distribution of $\hat{\Psi}^2_k$ will be affected by the estimation error of $F(\cdot)$, and it will no longer be asymptotically χ^2_k distributed. We now discuss the appropriate modification needed to restore the original ADF (χ^2_k) property.

For convenience of asymptotic analysis, we slightly modify Assumption 2 as follows.

Assumption 2′. $\mathbb{E}\left(\pi_j(Z_i)^{2+\epsilon}\right) < \infty$, for some $\epsilon > 0$, $j = 1, \cdots, k$.

Since the basic ingredients in constructing $\hat{\Psi}^2_k$ are $\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\pi_j\left(\hat{Z}_i\right)$, $j = 1, \ldots, k$, we first derive the joint distribution of $\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\pi_j\left(\hat{Z}_i\right)$, $j = 1, \ldots, k$. Let

$$\Pi_k = \left(\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\pi_1\left(\hat{Z}_i\right), \ldots, \frac{1}{\sqrt{m}}\sum_{i=1}^{m}\pi_k\left(\hat{Z}_i\right)\right)'.$$

The limiting behavior of Π_k is summarized in the following theorem.

THEOREM 3. *Under Assumptions* 1 *and* 2′*, and* H_0*:* $F = G$*, if* $m = \lambda n$*,* Π_k *converges to a* k*-dimensional normal variate with covariance matrix*

$$\Omega_k = I_k + \lambda\Delta_k, \tag{13}$$

where I_k *is the* k*-dimensional identity matrix and* Δ_k *is a* $k \times k$ *matrix whose* (j, r)*-th element is*

https://doi.org/10.1017/S0266466612000370 Published online by Cambridge University Press

$$\delta_{jr} = \mathbb{E}\left[\dot{\pi}_j(Z_1)\,\dot{\pi}_r(Z_2)\{F(Y_1 \wedge Y_2) - F(Y_1)\,F(Y_2)\}\right],$$
$$j = 1, \ldots, k; \quad r = 1, \ldots, k,$$

with $\dot{\pi}_j(.)$ *denoting the first derivative of* $\pi_j(.)$ *and* $Y_i \wedge Y_j = \min\left(Y_i, Y_j\right)$.

The result in Theorem 3 shows that when the two sample sizes are of the same order of magnitude, the limiting distribution of Π_k is different from its infeasible counterpart when the true CDF is used. Consequently, the limiting distribution of $\hat{\Psi}_k^2$ is no longer distribution free, and thus can not be directly used as a test statistic in our inference problem. However, Theorem 3 indicates that the limiting variate of Π_k is still multivariate normal, and an asymptotic χ^2-test can be constructed after appropriate standardization.

It is also instructive to note the change in the asymptotic covariance of Π_k; from I_k (equations (7) and (13)) it changes to $I_k + \lambda\Delta_k$ when estimated CDF is used. However, given the result in Theorem 3, we can construct a valid test by replacing Δ_k by its consistent estimator. Letting

$$\widehat{\delta}_{jr} = \frac{1}{nm^2}\sum_{l=1}^{n}\sum_{i=1}^{m}\sum_{t=1}^{m}\left[\dot{\pi}_j\left(\widehat{Z}_i\right)\dot{\pi}_r\left(\widehat{Z}_t\right)\left[I\left(X_l \le Y_i \wedge Y_t\right) - \widehat{Z}_i\,\widehat{Z}_t\right]\right]$$

and

$$\widehat{\Delta}_k = \begin{bmatrix} \widehat{\delta}_{11} & \cdots & \widehat{\delta}_{1k} \\ \cdots & \ddots & \cdots \\ \widehat{\delta}_{k1} & \cdots & \widehat{\delta}_{kk} \end{bmatrix},$$

we may estimate the covariance matrix Ω_k by

$$\widehat{\Omega}_k = I_k + \frac{m}{n}\widehat{\Delta}_k.$$

The modified testing statistic

$$\widetilde{\Psi}_k^2 = \Pi_k'\widehat{\Omega}_k^{-1}\Pi_k \tag{14}$$

is asymptotically distributed as χ_k^2 and can be used when m and n are of the same order.

Under the sample size condition of Theorem 1, $\widehat{\Omega}_k \to I_k$, the modified testing statistic $\widetilde{\Psi}_k^2$ is asymptotically equivalent to $\hat{\Psi}_k^2$, and the limiting result still holds in that case. Thus we may treat $\widetilde{\Psi}_k^2$ as a generalized version of the smooth test in Section 2, which includes the previous test as a special case.

Remark 1. The optimal choice of k has been studied extensively in the literature of data-driven smooth tests (Bickel and Ritov, 1992; Ledwina, 1994; Kallenberg and Ledwina, 1995a, 1995b, 1997; Hart, 1997; Kallenberg, 2002). They suggested that the main advantage of the data-driven smooth test is that unlike

https://doi.org/10.1017/S0266466612000370 Published online by Cambridge University Press

the conventional smooth test it provides a consistent test against all alternatives and it has also been shown to be more efficient. It is optimal in terms of the *vanishing shortcoming* or the discrepancy between the maximal attainable power and its achieved power (Inglot, Kallenberg, and Ledwina, 1994, 1998, 2000; Inglot and Ledwina, 2006a). Data-driven methods reduce the subjectivity of the choice of k by using criteria like the Akaike information criterion (AIC) or the Bayesian information criterion (BIC). As the BIC penalty is relatively heavy, it is more suitable for lower-dimensional models, whereas the AIC-based penalty will suit high-dimensional departures (Inglot and Ledwina, 2006a). Generalizing the penalty, like a combination of both the AIC and BIC approaches or choosing the smoothness of the density itself, has also been more recently advocated (Kallenberg, 2002; Inglot and Ledwina, 2006b; Fromont and Laurent, 2006).

Neyman's (1937) smooth test in the one sample case for fixed k case has been shown to be *asymptotically maximin* (Lehmann and Romano, 2005, Thm. 14.4.1, p. 602). However, several studies have discussed cases where increasing the order of the test k slowly to ∞ would have better power for alternative densities having heavier tails (Kallenberg, Oosterhoff, and Schriever, 1985; Inglot, Jurlewicz, and Ledwina, 1990; Eubank and LaRiccia, 1992).

A nonparametric data-driven directional test that checks for tail alternatives was also proposed (Albers, Kallenberg, and Martini, 2001). The procedure for selecting the truncation point k in an adaptive version of the Neyman (1937) smooth test is similar to the choice of the number of classes in the Pearson χ^2 test and has been discussed in Kallenberg et al. (1985). Fan (1996) and Fan and Huang (2001) looked at a similar adaptive Neyman test with choice of k with wavelet thresholding. A combination of different score test statistics that have nontrivial power on all directions at $n^{-1/2}$ rate has been proposed by Bickel, Ritov, and Stoker (2006).

3. AN APPLICATION TO COMPARING TWO AGE DISTRIBUTIONS

We consider an insurance market where several companies compete to provide health insurance for clients. These clients can be grouped into different risk categories depending on their "proneness" or "propensity" for having bad health. However, the main problem insurance companies face is *adverse selection*, since they cannot see beforehand what type of client they are insuring, high risk or low risk (Akerlof, 1970).

Rothschild and Stiglitz (1976) considered an insurance market setup where there are two types of clients, high and low risk. They claim that a health insurance contract based purely on risk categories will ensure that the high-risk client chooses to pay a higher premium. Moreover, both high- and low-risk clients will buy full (or complete) insurance coverage. However, if it is not possible to write a health insurance contract based on the risk categories due to either legislation or other restrictions, then there could be two possible scenarios. First, healthier (low-risk) individuals will choose to buy less than complete coverage while the less healthy (high-risk) individuals will buy full insurance, so the insurance

https://doi.org/10.1017/S0266466612000370 Published online by Cambridge University Press

market will still function, although this will not be as efficient as risk-based contracts. The second scenario happens if further regulations restrict or prohibit the selling of less than full insurance. In this case, the healthy or low-risk individuals will stop buying coverage, which means that gradually the insured population will be made up of more high-risk individuals. Thus, the insurance company has to pay out more often. This will cause the premium to go up, so healthier individuals will drop their coverage even further and this cycle will continue. Finally, this will result in a total collapse of the insurance market. This scenario is referred to as the "adverse selection death spiral" (Buchmueller and DiNardo, 2002).

To find the evidence of any existence of the adverse selection death spiral, the state of New York, where legislation for enforcing "community rating" (premium fixed by community and not by risk category) was enacted in 1993, can be compared with Pennsylvania, where no such legislation was enacted. We would like to test for the difference between the age distributions of the adult civilian population between 18 and 64, before and after 1993. The data are from the 1987–1996 March Current Population Survey, covering questions on whether an individual has insurance coverage and, if so, whether the coverage is through a small employer, and other similar questions. New York and Pennsylvania were selected due to their similarities, both geographically and demographically.

Our objective is to use the smooth test to determine if there is a difference between the age distributions in a state before and after 1993. The population selected was individuals who were covered by group insurance policies sponsored by employers with 100 or fewer employees. This population was divided into two parts, one for the individuals before 1993 and the other for those after 1993. Figure 1 gives the estimated PDFs using kernel density estimator. We estimated the PDF $f(x)$ of the sample $x_1, x_2, \ldots, x_n$ using kernel density

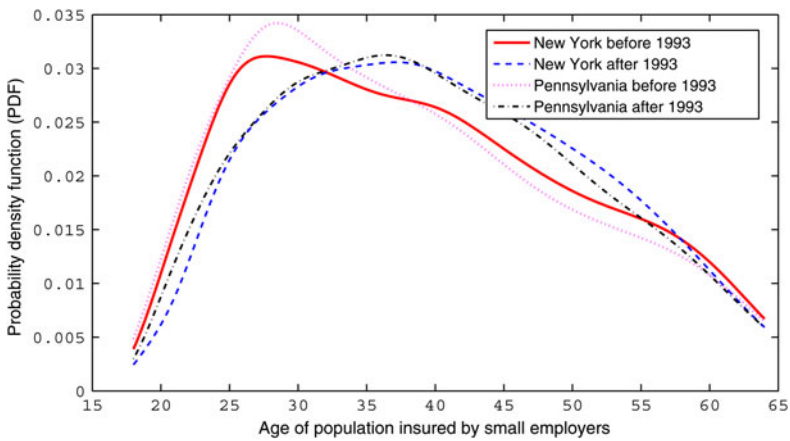


FIGURE 1. Density estimates for New York and Pennsylvania.

https://doi.org/10.1017/S0266466612000370 Published online by Cambridge University Press

TABLE 1. Goodness-of-fit statistics based on EDF

Test statistic	New York	Pennsylvania	Critical values (upper 0.1%)
KS	4.3492	4.4053	1.95
Kuiper	5.4809	5.3015	2.303
CvM	5.3503	5.1944	1.167
A-D	28.4875	25.2846	6.0

estimator $\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{x-x_i}{b}\right)$, where the bandwidth is *quadruple* of $b = 1.06\min(\hat{\sigma}_x, IQR/1.34)n^{-1/5}$, $\hat{\sigma}_x$ is the estimated standard deviation of X, IQR is the interquartile range for the sample, and $K(.)$ is the kernel suggested by Parzen (1962; also see Silverman, 1986, pp. 45–47).

Table 1 presents the values of some of the test statistics based on the EDF for the age distributions of groups insured by small employers before and after 1993 in New York and Pennsylvania along with the 0.1% critical values of all the modified statistics (Stephens, 1970; D'Agostino and Stephens, 1990). All standard tests clearly indicate that the two distributions are different. However, these test statistics are not informative about the nature of the deviation from the equality of two distributions.

Suppose now that the sample $y_1, y_2, \ldots, y_m$ comes from a population after 1993. The probability integral transforms (PITs) of each y_i are based on

$$\hat{z}_i = F_n(y_i), \qquad i = 1, 2, \ldots, m. \tag{15}$$

If the age distributions before and after 1993 are the same, then $\hat{z}_1, \hat{z}_2, \ldots, \hat{z}_m$ should be approximately distributed as $U(0, 1)$.

To elucidate more information from the data, we use the smooth test and its components based on the statistic $\hat{\Psi}_k^2$ in (10) with $k = 4$. Table 2 provides the smooth test results. For New York, we have $\hat{\Psi}_k^2 = 117.5011$, which is highly significant at any standard significance level. Therefore, we may conclude that the age distributions before and after 1993 are quite different. Now considering the four (asymptotically independent) components of $\hat{\Psi}_k^2$, each of which $\left(\hat{u}_j^2\right)$ is approximately distributed as χ_1^2, we note that the hypotheses of equality of two

TABLE 2. Smooth statistic and its four components

Source	$\hat{\Psi}_4^2$	$\hat{u}_1^2$	$\hat{u}_2^2$	$\hat{u}_3^2$	$\hat{u}_4^2$	$\tilde{\Psi}_4^2$
New York ($n = 4548, m = 2517$)	117.5011	69.0543	46.0559	0.819	1.572	83.3719
p-value $\left(\chi^2\right)$	0.0000	0.0000	0.0000	0.3655	0.2099	0.0000
Pennsylvania ($n = 3113, m = 1875$)	96.3788	64.9253	21.5569	8.7567	1.13	63.0314
p-value $\left(\chi^2\right)$	0.0000	0.0000	0.0000	0.0031	0.2878	0.0000

https://doi.org/10.1017/S0266466612000370 Published online by Cambridge University Press

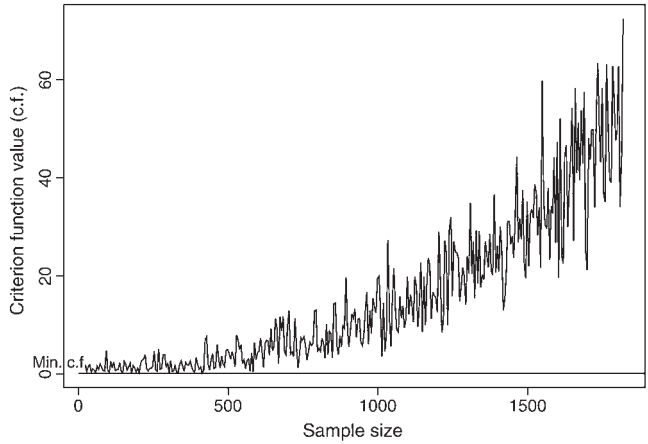


FIGURE 2. Criteria function for optimal sample size for closeness under H_0 ($n = 4548$; repl.$= 200$; $m_{opt} = 455$).

distributions are most possibly rejected due to the differences in the first- (location) and second- (scale) order moments. The contributions of $\hat{u}_3^2$ and $\hat{u}_4^2$ are indeed very small. As discussed in Section 2.2, there could be size distortion in uncritical use of $\hat{\Psi}_4^2$ with asymptotic χ_4^2 critical values. A way to rectify the size distortion of $\hat{\Psi}_4^2$ is to exploit the result in Theorem 2 and use a smaller test sample size m. Again, as we will discuss in Section 4, our simulation study shows that (see Figure 2) the optimal value of m is around 500. Therefore, we repeat the calculations of $\hat{\Psi}_4^2$ and its components with $n = 4,548$ but $m = 500$ (randomly drawn). The test statistics and the corresponding p-values are reported in Table 3. The qualitative result, i.e., the significance of $\hat{\Psi}_4^2$ and the two components $\hat{u}_1^2$ and $\hat{u}_2^2$, remain unchanged. Yet another way to to handle size distortion when m and n are of the same order of magnitude (without subsampling) is to use the modified statistic $\widetilde{\Psi}_4^2$ in (14). Its value for New York is 83.37, which is also highly significant. Therefore, we come to the same conclusion using various approaches. The results for the Pennsylvania data are similar (see Table 2). Thus, the population age distributions before and after 1993 in both states are indeed different irrespective of the community rating legislation, and the sources of the differences are mainly through location and scale.

TABLE 3. Neyman's smooth statistic for NY and components ($n = 4,548$, $m = 500$)

Source	$\hat{\Psi}_4^2$	$\hat{u}_1^2$	$\hat{u}_2^2$	$\hat{u}_3^2$	$\hat{u}_4^2$
Test statistic	27.6095	7.4413	13.7468	3.8773	2.544
p-value	0.0000	0.0064	0.0002	0.0489	0.1107

https://doi.org/10.1017/S0266466612000370 Published online by Cambridge University Press

4. MONTE CARLO RESULT

To study the finite sample properties of the smooth test, we generate the data under the null hypothesis of equality of distributions by subsampling from the population of group insured civilian adults before 1993 in New York, as used in our application in Section 3. In Figure 3 we plot the empirical size of the test as a function of m when the nominal level is 5%. The estimation sample size n is fixed at $5,000$, also drawn at random from the same population. We evaluate the size of the test for $m = 50$ to $3,500$ with an increment of 10 based on $2,000$ replications. There is a high relative variation in the empirical size; however, we observe an increasing trend of the test size with m, though it never crosses the 10% level. Furthermore, there are several values of m for which the actual sizes are close to the nominal size of 5%.

In order to fully explore the finite sample performance of the suggested tests, in addition to studying closeness of the actual and nominal sizes, we also evaluated the goodness of the component χ^2 approximations (χ^2_1 and χ^2_4, depending on the test statistics) of the distributions under the null hypothesis. Using the values of the components of the test statistics from $10,000$ replications with $n = 2,500$ and $m = 500$, we plot the estimated kernel densities and the corresponding theoretical χ^2 densities. In the panels of Figure 4, it is hard to distinguish between the theoretical χ^2_1 and the empirical densities of $\hat{u}^2_j$ $(j = 1, 2, 3, 4)$, the components of $\hat{\Psi}^2_4$. This is a worst-case scenario since here $m = 500$ is substantially high for $n = 2,500$. We also note that the estimated kernel density of $\hat{\Psi}^2_4$ is not too far from that of the theoretical χ^2_4 (not shown).

Although a good overall match of the empirical distribution of a test statistic to theoretical χ^2 is desirable, having the true size of the test close to the nominal

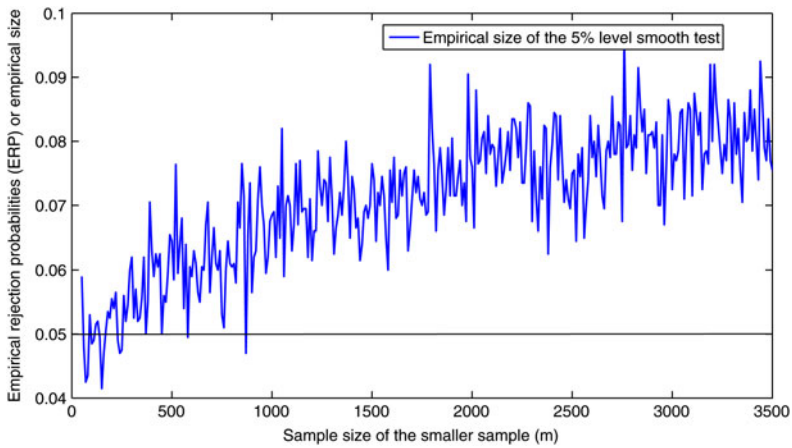


FIGURE 3. Test sample size and test size for $n = 5,000$ and increasing m $(r = 2,000)$.

https://doi.org/10.1017/S0266466612000370 Published online by Cambridge University Press

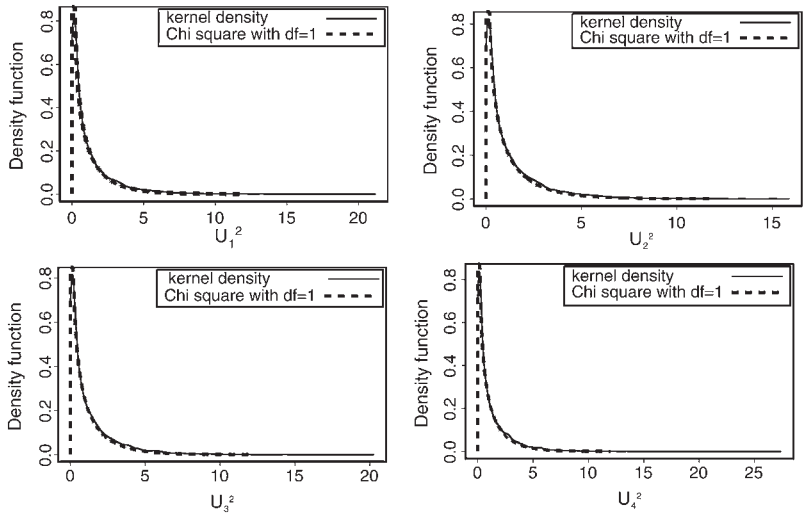


FIGURE 4. Density plots for individual U_i^2 under H_0 and theoretical χ_1^2.

level (say, 5%) is very important. In Table 4 we present the actual size when $n = 2,500$ and $m = 500$ and 50, with 10,000 replications. Theorem 2 suggests that for minimum size distortion m needs to be in the order of $\sqrt{n}$. From that point of view, the actual sizes when $m = 500$ are pretty good (only for $\hat{u}_4^2$, the size 0.083 is much higher than 5%); qualitatively, we could have inferred this by looking at the tail parts of the empirical distributions in the panels of Figure 4. For $m = 50$, the results are indeed very good; the actual sizes never cross the nominal 5% level.

Subsampling from the New York data used in the empirical application is convenient to study the finite sample size properties of the test. However, to study finite sample power-size trade-off in a systematic way, we need a different framework. We generate the data using the following mixture of Gaussian and log-normal densities:

$$X = B * X_1 + (1 - B) * X_2. \tag{16}$$

Observations under the null hypothesis are generated taking $B \sim Bernoulli(0.3)$, $\ln X_1 \sim N(-1, 4)$ and $X_2 \sim N(1.2, 1.21)$, and under the alternative $B \sim$

TABLE 4. Actual sizes for 5% nominal size tests

Source	$\hat{\Psi}_4^2$	$\hat{u}_1^2$	$\hat{u}_2^2$	$\hat{u}_3^2$	$\hat{u}_4^2$
Asymptotic null distribution	χ_4^2	χ_1^2	χ_1^2	χ_1^2	χ_1^2
Actual size ($n = 2500, m = 500$)	0.0577	0.0423	0.0529	0.0497	0.083
Actual size ($n = 2500, m = 50$)	0.0425	0.0473	0.0474	0.0461	0.0473

https://doi.org/10.1017/S0266466612000370 Published online by Cambridge University Press

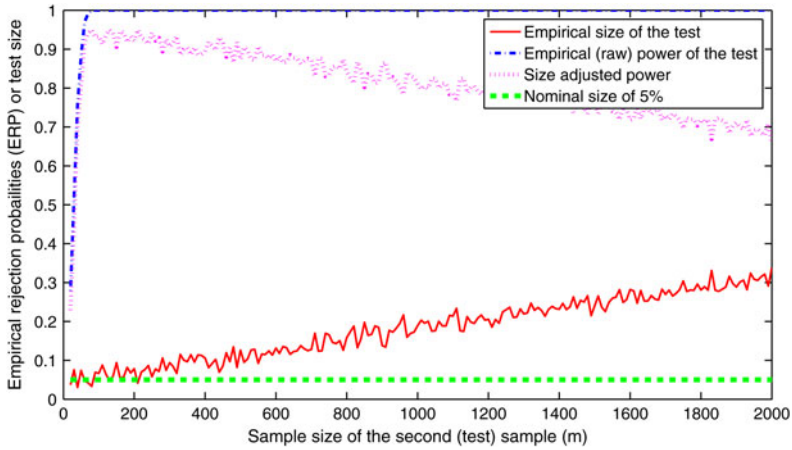


FIGURE 5. Size and power of the smooth test for simulated data ($n = 2,000$).

Bernoulli(0.5), $\ln X_1 \sim N(-0.1, 1)$ and $X_2 \sim N(1.75, 0.81)$. We take $n = 2,000$ and increase the value of m with increments of 10, up to $m = 2,000$ (so that m and n are of the same order of magnitude) and perform 2,000 replications. Figure 5 provides the size, (raw) power, and size-adjusted power of $\hat{\Psi}_4^2$. The test has very good power properties even at small sample size, say $m = 100$. The size-adjusted power, though decreasing with m, is also quite good. Of course, given these results our recommendation would be not to use $\hat{\Psi}_4^2$ mechanically when m and n are of similar order due to significant size distortion.

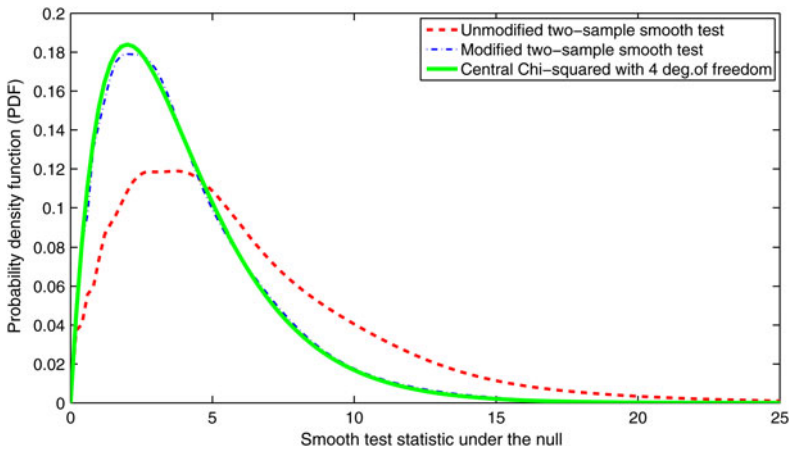


FIGURE 6. Density of the modified $\widehat{\Psi}_k^2$ under H_0 when $\gamma = 0.5$ ($n = 2,000, m = 1,000, r = 2,000$).

https://doi.org/10.1017/S0266466612000370 Published online by Cambridge University Press

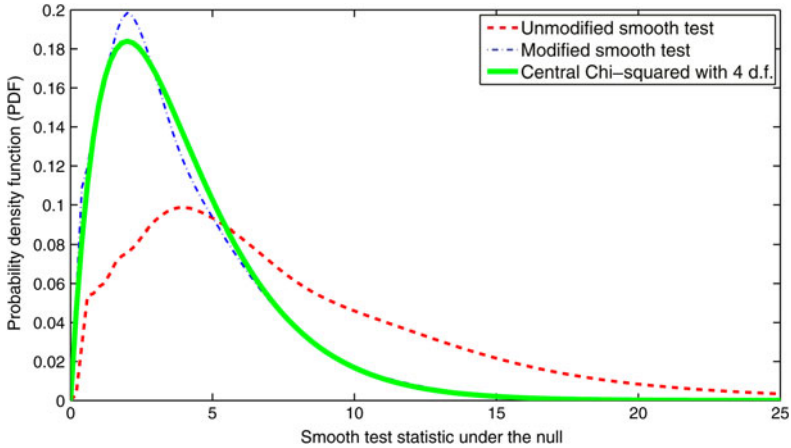


FIGURE 7. Density of the modified $\widehat{\Psi}_k^2$ under H_0 when $\gamma = 1$ ($n = 1,000, m = 1,000$, $r = 1,000$).

When $n = 2,000$ and $m = 1,000$ the empirical size of $\hat{\Psi}_4^2$ is close to 20%; however, for this case if we use our modified test $\widetilde{\Psi}_4^2$ the size reduces to close to 5%. To get an idea of the finite sample distribution of $\widetilde{\Psi}_4^2$ (relative to $\hat{\Psi}_4^2$ and χ_4^2) under the null hypothesis, we plot these three densities in Figure 6. It is clear that the kernel density function of $\widetilde{\Psi}_4^2$ almost overlaps with that of χ_4^2, while the density of $\hat{\Psi}_4^2$ has a much thicker tail requiring much larger critical values for valid inference. To see the impact of m on the modified test statistic $\widetilde{\Psi}_4^2$, we repeat the above experiment (of Figure 6) but with $n = m = 2,000$ and plot the

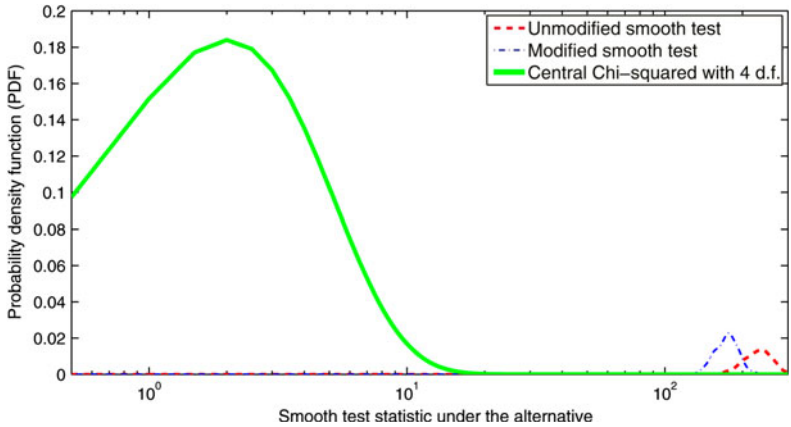


FIGURE 8. Density of the modified $\widehat{\Psi}_k^2$ under H_1 when $\gamma = 1$ ($n = 1,000, m = 1,000$, $r = 1,000$) (in log scale).

https://doi.org/10.1017/S0266466612000370 Published online by Cambridge University Press

three densities in Figure 7. Now, although the estimated kernel density of $\widetilde{\Psi}_4^2$ does not overlap with that of χ_4^2 completely, the deviation is not all that substantial. The departure of the empirical density of $\hat{\Psi}_4^2$ from χ_4^2 is more pronounced compared to that in Figure 6.

The empirical densities of $\hat{\Psi}_4^2$ and $\widetilde{\Psi}_4^2$ under the alternative hypothesis are plotted in Figure 8, for $n = m = 1,000$ (i.e., $\lambda = 1.0$) along with the χ_4^2 density. The distributions of the unmodified $\hat{\Psi}_4^2$ lie further to the right then those of the modified counterparts, reflecting the higher rejection rates for $\hat{\Psi}_4^2$ due to its much larger size.

5. SUMMARY AND FUTURE RESEARCH

We propose a smooth test for comparing the distributions in a two-sample setup. Unlike traditional *omnibus* goodness-of-fit tests such as the Kolmogorov-Smirnov or Cramér-von Mises procedures, the smooth test helps us identify the sources of departure from the null hypothesis of equality of two densities. Since the smooth test can also be derived from Rao's score principle, it enjoys some optimality properties (Bera and Bilias, 2001). We have also investigated the choices of the relative sizes of the estimation and test samples to minimize size distortion. Our Monte Carlo results reveal that the smooth test (with proper choice of the sample size) and its modified form have good finite sample properties in terms of size, power, and closeness to the theoretical χ^2 distributions.

There are several directions of future research that we would like to pursue.

Following Neyman (1937) we have kept the value of k fixed at 4. In some applications however (as discussed in Section 2.3), it may be desirable to choose k as an increasing function of the sample size using some model selection criterion. If k is too high, the effectiveness of the test in each direction could be diluted, as pointed out by Neyman (also see the discussion in Bera and Ghosh, 2002, p. 205). Janic-Wróblewska and Ledwina (2000) proposed a data-driven version of the smooth test, where k can take a maximum value $d(m) = o\left(\{m/\log m\}^{1/9}\right)$, which increases with m at a very slow rate. An adaptive version of the two-sample test method was shown to be efficient and locally most powerful under more realistic settings (Ducharme and Ledwina, 2003). It would be quite interesting to combine our methodology with that of Janic-Wróblewska and Ledwina.

As suggested by one of the referees, another very useful extension would be to consider nonorthogonal moment functions such as $\mathbb{E}\left(X^j - Y^j\right)$, $j = 1, 2, \ldots$ and follow the GMM-type formulation of Newey (1985). With this approach, though we lose local optimality property of Rao score principle (see, for example, Bera and Ghosh (2002) and references therein) and identification of the directions of departure from the null hypothesis, we could avoid the problem of choosing m and use the complete samples.

To obtain asymptotic chi-squared tests, we either impose restrictions on the relative magnitude of sample sizes or modify the original smooth test statistic.

https://doi.org/10.1017/S0266466612000370 Published online by Cambridge University Press

An alternative approach to deal with the preliminary estimation error is to use a bootstrap or subsampling method to construct a valid test. A bootstrap version of the smooth test may be constructed in the spirit of Mora and Neumeyer (2005; also see Horowitz, 2002).

We can extend our test similar to what Fan (1996) proposed adaptively using both Neyman's (1937) technique and a wavelet-based procedure for comparing global and local departures from the null hypothesis (also see Fan and Huang, 2001). Finally, the test can be modified to allow for possible dependence in the data, particularly in the context of time series or panel data (Bai, 2003).

REFERENCES

Akerlof, G. (1970) The market for "lemons": Qualitative uncertainty and the market mechanism. *Quarterly Journal of Economics* 89, 488–500.

Albers, W., W.C.M. Kallenberg, & F. Martini (2001) Data-driven rank tests for classes of tail alternatives. *Journal of the American Statistical Association* 96, 685–696.

Bai, J. (2003) Testing parametric conditional distributions of dynamic models. *Review of Economics and Statistics* 85, 531–549.

Bera, A.K. & Y. Bilias (2001) Rao's score, Neyrnan's C(α) and Silvey's LM tests: An essay on historical developments and some new results. *Journal of Statistical Planning and Inference* 97, 9–44.

Bera, A.K. & A. Ghosh (2002) Neyman's smooth test and its applications in econometrics. In A. Ullah, A. Wan, & A. Chaturvedi (eds.), *Handbook of Applied Econometrics and Statistical Inference*, pp. 177–230. Marcel Dekker.

Bera, A.K., A. Ghosh, & Z. Xiao (2010) Smooth Test for Equality of Distributions. Working paper presented at the International Symposium on Econometrics of Specification Tests in 30 Years, Xiamen University.

Bickel, P.J. & Y. Ritov (1992) Testing for goodness of fit: A new approach. In A.K.Md.E. Saleh (ed.), *Nonparametric Statistics and Related Topics*, pp. 51–57. North-Holland.

Bickel, P.J., Y. Ritov, & T. Stoker (2006) Tailor-made tests for goodness of fit to semiparametric hypotheses. *Annals of Statistics* 34, 721–741.

Buchmueller, T. & J. DiNardo (2002) Did community rating induce an adverse selection death spiral? Evidence from New York, Pennsylvania, and Connecticut. *American Economic Review* 92, 280–294.

D'Agostino, R.B. & M.A. Stephens (1986) *Goodness-of-Fit Techniques*. New York: Marcel Dekker.

Darling, D.A. (1957) The Kolmogorov-Smirnov, Cramér-von Mises tests. *Annals of Mathematical Statistics* 28, 823–838.

Ducharme, G.R. & T. Ledwina (2003) Efficient and adaptive nonparametric test for the two-sample problem. *Annals of Statistics* 31, 2036–2058.

Escanciano, J.C. (2009) On the lack of power of omnibus specification tests. *Econometric Theory* 25, 162–194.

Eubank, R.L. & V.N. LaRiccia (1992) Asymptotic comparison of Cramér-von Mises and nonparametric function estimation techniques for testing goodness-of-fit. *Annals of Statistics* 20, 1412–1425.

Fan, J. (1996) Test of significance based on wavelet thresholding and Neyman's truncation. *Journal of the American Statistical Association* 91, 674–688.

Fan, J. & L-S. Huang (2001) Goodness-of-fit tests for parametric regression models. *Journal of the American Statistical Association* 96, 640–652.

Fromont, M. & B. Laurent (2006) Adaptive goodness-of-fit tests in a density model. *Annals of Statistics* 34, 680–720.

https://doi.org/10.1017/S0266466612000370 Published online by Cambridge University Press

Gourieroux, C. & A. Monfort (1996) *Simulation-based Econometric Methods*. Oxford University Press.

Hart, J.D. (1997) *Nonparametric Smoothing and Lack of Fit Tests*. Springer-Verlag.

Horowitz, J.L. (2002) The bootstrap in econometrics. *Statistical Science* 18, Silver Anniversary of the Bootstrap, 211–218.

Inglot, T., T. Jurlewicz, & T. Ledwina (1990) On Neyman-type smooth tests of fit. *Statistics* 21, 549–568.

Inglot, T., W.C.M. Kallenberg, & T. Ledwina (1994) Power approximations to and power comparison of smooth goodness-of- fit tests. *Scandinavian Journal of Statistics* 21, 131–145.

Inglot, T., W.C.M. Kallenberg, & T. Ledwina (1998) Vanishing shortcoming of data driven Neyman's tests. In B. Szyszkowicz (ed.), *Asymptotic Methods in Probability and Statistics*: *A Volume to Honour Miklós Csörgo*, pp. 811–829. Elsevier.

Inglot, T., W.C.M. Kallenberg, & T. Ledwina (2000) Vanishing shortcoming and asymptotic relative efficiency. *Annals of Statistics* 28, 215–238 [Correction: ibidem 28 (2000), 1795].

Inglot, T. & T. Ledwina (2006a) Data-driven score tests for homoscedastic linear regression model: Asymptotic results. *Probability and Mathematical Statistics* 26, 41–61.

Inglot, T. & T. Ledwina (2006b) Towards data driven selection of a penalty function for data driven Neyman tests. *Linear Algebra and Its Applications* 417, 124–133.

Janic-Wróblewska, A. & T. Ledwina (2000) Data driven rank test for two-sample problem. *Scandinavian Journal of Statistics* 27, 281–297.

Janssen, A. (2000) Global power functions of goodness of fit tests. *Annals of Statistics* 28, 239–253.

Kallenberg, W.C.M. (2002) The penalty in data driven Neyman's tests. *Mathematical Methods of Statistics* 11, 323–340.

Kallenberg, W.C.M. & T. Ledwina (1995a) Consistency and Monte Carlo simulation of a data driven version of smooth goodness-of-fit tests. *Annals of Statistics* 23, 1594–1608.

Kallenberg, W.C.M. & T. Ledwina (1995b) On data driven Neyman's tests. *Probability and Mathematical Statistics* 15, 409–426.

Kallenberg, W.C.M. & T. Ledwina (1997) Data-driven smooth tests when the hypothesis is composite. *Journal of the American Statistical Association* 92, 1094–1104.

Kallenberg, W.C.M., J. Oosterhoff, & B.F. Schriever (1985) The number of classes in chi-squared goodness-of-fit tests. *Journal of the American Statistical Association* 80, 959–968.

Kendall, M.G. & A. Stuart (1973) *The Advanced Theory of Statistics*, 3rd ed., vol. 2. Hafner.

Ledwina, T. (1994) Data-driven version of Neyman's smooth test of fit. *Journal of the American Statistical Association* 89, 1000–1005.

Lehmann, E.L. (1953) The power of rank tests. *Annals of Mathematical Statistics* 24, 23–43.

Lehmann, E.L. & J.P. Romano (2005) *Testing Statistical Hypothesis*, 3d ed. Springer.

McLeish, D.L. (1974) Dependent central limit theorems and invariance principles. *Annals of Probability* 2, 620–628.

Mora, J. & N. Neumeyer (2005) The Two-Sample Problem with Regression Errors: An Empirical Process Approach. Working paper, Universidad de Alicante.

Newey, W.K. (1985) Generalized method of moments specification testing. *Journal of Econometrics* 29, 229–256.

Neyman, J. (1937) 'Smooth test' for goodness of fit. *Skandinaviske Aktuarietidskrift* 20, 150–199.

Parzen, E. (1962) On estimation of a probability density function and mode. *Annals of Mathematical Statistics* 33, 1065–1076.

Pearson, K. (1900) On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine, 5th series* 50, 157–175.

Rao, C.R. (1948) Large sample tests of statistical hypothesis concerning several parameters with applications to problems of estimation. *Proceedings of the Cambridge Philosophical Society* 44, 50–57.

Rothenberg, T.J. (1984) Approximating the distribution of econometric estimators and test statistics. In Z. Griliches and M.D. Intriligator (eds.), *Handbook of Econometrics*, vol. 2. North-Holland.

https://doi.org/10.1017/S0266466612000370 Published online by Cambridge University Press

Rothschild, M. & J. Stiglitz (1976) Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. *Quarterly Journal of Economics* 90, 630–649.
Shorack, G.R. & J. Wellner (1986) *Empirical Processes with Applications to Statistics*. Wiley.
Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis.* Chapman & Hall.
Solomon, H. & M.A. Stephens (1985) Neyman's test for uniformity. In S. Kotz and N.L. Johnson (eds.), *Encylcopedia of Statistical Sciences*, vol. 6, pp. 232–235. Wiley.
Stephens, M.A. (1970) Use of Kolmogorov-Smirnov, Cramer-von Mises and related statistics without extensive tables. *Journal of the Royal Statistical Society Series B* 32, 115–122.
Stock, J.H., J.H. Wright, & M. Yogo (2002) A survey of weak instruments and weak identification in generalized method of moments. *The Journal of Business and Economic Statistics* 20, 518–529.
Thomas, D.R., & D.A. Pierce (1979) Neyman's smooth goodness-of-fit test when the hypothesis is composite. *Journal of the American Statistical Association* 74, 441–445.

APPENDIX: Proofs

We first give two lemmas that will be used in our proof of Theorem 1.

LEMMA 1. *The normalized Legendre polynomials are given by*

$$\pi_j(z) = \frac{\sqrt{2j+1}}{j!}\frac{d^j}{dz^j}\left(z^2 - z\right)^j, \qquad 0 \le z \le 1, \qquad j = 1, 2, 3 \ldots. \tag{A.1}$$

Proof. The standard Legendre polynomials (see, for instance, Kendall and Stuart, 1973, p. 460) are

$$P_j(x) = \frac{1}{2^j j!}\frac{d^j}{dx^j}\left(x^2 - 1\right)^j, \qquad -1 \le x \le 1, \qquad j = 1, 2, 3 \ldots, \tag{A.2}$$

where

$$\int_{-1}^{1} P_j(x) P_l(x)\, dx = \begin{cases} 0, & \text{if } j \ne l, \\ \frac{2}{2j+1}, & \text{if } j = l. \end{cases} \tag{A.3}$$

Now, changing the variable to $z = \frac{x+1}{2}$, or $x = 2z - 1$,

$$\begin{aligned} \tilde{P}_j(z) &= \frac{1}{2^j j!} \times \frac{1}{2^j}\frac{d^j}{dz^j}\left\{2z-1)^2 - 1\right\}^j \\ &= \frac{1}{2^{2j} j!}\frac{d^j}{dz^j}\left\{4z^2 - 4z\right\}^j \\ &= \frac{1}{j!}\frac{d^j}{dz^j}\left\{z^2 - z\right\}^j. \end{aligned}$$

Since $\int_0^1 \left(\tilde{P}_j(z)\right)^2 dz = \int_{-1}^1 2^{-1}\left(P_j(x)\right)^2 dx = (2j+1)^{-1}$ (see equation (A.3)), normalizing we have

$$\pi_j(z) = \frac{\sqrt{2j+1}}{j!}\frac{d^j}{dz^j}\left(z^2 - z\right)^j, \qquad 0 \le z \le 1, \qquad j = 1, 2, \ldots. \tag{A.4}$$

■

https://doi.org/10.1017/S0266466612000370 Published online by Cambridge University Press

LEMMA 2. *If $\pi_j(.)$ is the normalized Legendre polynomial of degree j defined on* $[0,1]$, *then the first-order Lipschitz condition holds; that is,*

$$\left|\pi_j\left(\hat{z}\right)-\pi_j\left(z\right)\right| \leq M\left|\hat{z}-z\right|, \tag{A.5}$$

where M is a positive constant, and z and $\hat{z}$ are any two points between 0 *and* 1.

Proof. Using the mean value theorem on $\pi_j(z)$ in (A.1), we have

$$\begin{aligned}\pi_j\left(\hat{z}\right)-\pi_j\left(z\right) &= \frac{\sqrt{2j+1}}{j!}\left(\hat{z}-z\right)\frac{d}{dz}\pi_j\left(z^*\right)\\ &= \frac{\sqrt{2j+1}}{j!}\left(\hat{z}-z\right)\frac{d^{j+1}}{dz^{j+1}}\left(z^2-z\right)^j\Big|_{z=z^*}, \end{aligned}\tag{A.6}$$

where z^* is such that $\left|z^*-z\right| < \left|\hat{z}-z\right|$.

Expanding the polynomial $\left(z^2-z\right)^j$, and differentiating with respect to z up to order $j+1$,

$$\begin{aligned}\frac{d}{dz}\pi_j\left(z\right) &= \frac{\sqrt{2j+1}}{j!}\frac{d^{j+1}}{dz^{j+1}}\left(z^2-z\right)^j\\ &= \frac{d^{j+1}}{dz^{j+1}}\sum_{l=0}^{j}\left(z^2\right)^l\left(z\right)^{j-l}\left(-1\right)^{j-l}\\ &= \frac{\sqrt{2j+1}}{j!}\sum_{l=1}^{j}\frac{(l+j)!}{(l-1)!}\frac{j!}{l!\,(l-j)!}\left(-1\right)^{j-l}z^{l-1}\\ &\leq \frac{\sqrt{2j+1}}{j!}\sum_{l=1}^{j}l\frac{(l+j)!}{(l)!}\frac{j!}{l!\,(l-j)!}\left|\left(-1\right)^{j-l}\right|\left|z^{l-1}\right|\\ &\leq \frac{j\,(j+1)}{2}\sqrt{2j+1}\sum_{l=1}^{j}\frac{(l+j)!}{(l)!l!\,(l-j)!}\\ &= \frac{j\,(j+1)}{2}\sqrt{2j+1}\,{}_2F_1\left(1-j,2+j;2;-1\right), \end{aligned}\tag{A.7}$$

where the finite hypergeometric function ${}_2F_1(.)$ is defined by

$${}_2F_1\left(a,b;c;x\right) = 1+\frac{ab}{c}x+\frac{a\,(a+1)\,b\,(b+1)}{c\,(c+1)\,2!}x^2+\cdots. \tag{A.8}$$

Hence, using (A.3) and (A.7),

$$\begin{aligned}\left|\pi_l\left(\hat{z}\right)-\pi_l\left(z\right)\right| &= \left|\hat{z}-z\right|\left|\frac{d}{dz}\pi_j\left(z\right)\Big|_{z=z^*}\right|\\ &\leq \left|\hat{z}-z\right|\frac{j\,(j+1)}{2}\sqrt{2j+1}\left|{}_2F_1\left(1-j,2+j;2;-1\right)\right|\\ &= M_j\left|\hat{z}-z\right|, \end{aligned}\tag{A.9}$$

where M_j is a finite positive number for any finite positive interger j. ■

https://doi.org/10.1017/S0266466612000370 Published online by Cambridge University Press

Proof of Theorem 1. We need to show that $\widehat{\Psi}_k^2 - \Psi_k^2 = o_p(1)$. Notice that

$$\begin{aligned}\widehat{\Psi}_k^2 &= \sum_{j=1}^{k}\left[\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\pi_j(\widehat{Z}_i)\right]^2 \\ &= \sum_{j=1}^{k}\frac{1}{m}\left[\sum_{i=1}^{m}\pi_j(Z_i)+\sum_{i=1}^{m}\left[\pi_j(\widehat{Z}_i)-\pi_j(Z_i)\right]\right]^2 \\ &= \Psi_k^2 + R_{1,m,n} + R_{2,m,n},\end{aligned}$$

where

$$\Psi_k^2 = \sum_{j=1}^{k}\frac{1}{m}\left[\sum_{i=1}^{m}\pi_j(Z_i)\right]^2, \qquad R_{1,m,n} = \sum_{j=1}^{k}\frac{1}{m}\left[\sum_{i=1}^{m}\left[\pi_j(\widehat{Z}_i)-\pi_j(Z_i)\right]\right]^2,$$

$$R_{2,m,n} = 2\sum_{j=1}^{k}\frac{1}{m}\left[\sum_{i=1}^{m}\pi_j(Z_i)\right]\left[\sum_{i=1}^{m}\left[\pi_j(\widehat{Z}_i)-\pi_j(Z_i)\right]\right].$$

We show that, as $m, n \to \infty$, and $(\log\log n)m/n \to 0$, $R_{1,m,n} = o_p(1)$, $R_{2,m,n} = o_p(1)$. We first consider $R_{1,m,n}$,

$$R_{1,m,n} = \sum_{j=1}^{k}\frac{1}{m}\left[\sum_{i=1}^{m}\left[\pi_j(\widehat{Z}_i)-\pi_j(Z_i)\right]\right]^2 \le \sum_{j=1}^{k}\frac{1}{m}\left[\sum_{i=1}^{m}\left|\pi_j(\widehat{Z}_i)-\pi_j(Z_i)\right|\right]^2.$$

By Lemma 2, $\left|\pi_j(u)-\pi_j(v)\right| \le M\left|u-v\right|$, for some finite number $M > 0$. Thus

$$\begin{aligned}R_{1,m,n} &\le \sum_{j=1}^{k}\frac{1}{m}\left[\sum_{i=1}^{m}\left|\pi_j(\widehat{Z}_i)-\pi_j(Z_i)\right|\right]^2 \le M^2\sum_{j=1}^{k}\frac{1}{m}\left[\sum_{i=1}^{m}\left|\widehat{Z}_i - Z_i\right|\right]^2 \\ &\le M^2\sum_{j=1}^{k}\frac{1}{m}\left[m\max_{1\le i\le m}\left|\widehat{Z}_i - Z_i\right|\right]^2.\end{aligned}$$

Recalling $\hat{Z}_i = F_n(Y_i)$ and $Z_i = F(Y_i)$, and noting that $\sup_u |F_n(u) - F(u)|$ is maximally of order $\sqrt{(\log\log n)/n}$ as $n \to \infty$ (see, e.g., Shorack and Wellner, 1986), we have

$$\max_{1\le i\le m}\left|\widehat{Z}_i - Z_i\right| = O_p(\sqrt{(\log\log n)/n}),$$

and

$$R_{1,m,n} \le mM^2\sum_{j=1}^{k}\left[\max_{1\le i\le m}\left|\widehat{Z}_i - Z_i\right|\right]^2 = O_p((\log\log n)m/n),$$

which converges to 0 if $(\log\log n)m/n \to 0$ as $m, n \to \infty$.

Similarly,

$$\begin{aligned}\left|R_{2,m,n}\right| &= 2\sum_{j=1}^{k}\frac{1}{m}\left|\sum_{i=1}^{m}\pi_j(Z_i)\right|\left|\sum_{i=1}^{m}\left[\pi_j(\widehat{Z}_i)-\pi_j(Z_i)\right]\right| \\ &\le 2\sum_{j=1}^{k}\frac{1}{m}\left|\sum_{i=1}^{m}\pi_j(Z_i)\right|\left[\sum_{i=1}^{m}\left|\pi_j(\widehat{Z}_i)-\pi_j(Z_i)\right|\right].\end{aligned}$$

https://doi.org/10.1017/S0266466612000370 Published online by Cambridge University Press

Note that $\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\pi_j(Z_i)$ converges to a standard normal variable and

$$\sum_{i=1}^{m}\pi_j(Z_i) = O_p\left(m^{1/2}\right).$$

Again using Lemma 2,

$$\begin{aligned}\sum_{i=1}^{m}\left|\pi_j\left(\widehat{Z}_i\right)-\pi_j(Z_i)\right| &\le M\sum_{i=1}^{m}\left|\widehat{Z}_i - Z_i\right| \le Mm\max_{1\le i\le m}\left|\widehat{Z}_i - Z_i\right| \\ &= O_p(m\sqrt{(\log\log n)/n}).\end{aligned}$$

Thus,

$$\begin{aligned}\left|R_{2,m,n}\right| &\le 2\sum_{j=1}^{k}\frac{1}{m}\left|\sum_{i=1}^{m}\pi_j(Z_i)\right|\left[\sum_{i=1}^{m}\left|\pi_j\left(\widehat{Z}_i\right)-\pi_j(Z_i)\right|\right] \\ &= O_p\left(\frac{1}{m}\times m^{1/2}\times m\sqrt{(\log\log n)/n}\right) \\ &= O_p\left(\sqrt{(\log\log n)m/n}\right),\end{aligned}$$

converges to 0 when $(\log\log n)m/n \to 0$ as $m, n \to \infty$. ■

Proof of Theorem 2. From the proof of Theorem 1 we know that the asymptotic χ^2 test statistic $\widehat{\Psi}_k^2$ can be decomposed into $\Psi_k^2 + R_{1,m,n} + R_{2,m,n}$. We analyze each of these terms to show how the relative magnitudes of m and n affect the size of the test.

The leading term is

$$\Psi_k^2 = \sum_{j=1}^{k}\frac{1}{m}\left[\sum_{i=1}^{m}\pi_j(Z_i)\right]^2.$$

By construction, conditional on X, for each j, $V_{ji} = \pi_j(Z_i)$ $(i = 1, \ldots, m)$ are m independent and identically distributed random variables with mean zero and unit variance. Thus,

$$\frac{1}{\sqrt{m}}\sum_{i=1}^{m}V_{ji} \Rightarrow N(0,1) \equiv \xi_j,$$

where ξ_j $(j = 1, \ldots, k)$ are k independent standard normal variates. In addition, assuming that V_{ji} possesses moments up to the fourth order, by standard result of Edgeworth expansion (see, e.g., Rothenberg, 1984, and references therein), the probability density function of $\frac{1}{\sqrt{m}}\sum_{i=1}^{m}V_{ji}$ has the expansion

$$f(x) \approx \varphi(x)\left[1 + \frac{k_3 H_3(x)}{6\sqrt{m}} + \frac{3k_4 H_4(x) + k_3^2 H_6(x)}{72m}\right],$$

where $\varphi(.)$ is the density of standard normal, k_r is the rth cumulant, and H_r is the Hermite polynomial of degree r defined as $H_r(x) = (-1)^r\varphi^{(r)}(x)/\varphi(x)$. Thus, $m^{-1/2}\sum_{i=1}^{m}V_{ji}$ can

https://doi.org/10.1017/S0266466612000370 Published online by Cambridge University Press

be expanded into a leading term of standard normal variable ξ_j plus a second-order term, say $\frac{1}{\sqrt{m}}A_j$, of $O_p(m^{-1/2})$, and a third term, $\frac{1}{m}B_j$, of order $O_p(m^{-1})$, i.e.,

$$\frac{1}{\sqrt{m}}\sum_{i=1}^{m} V_{ji} \approx \xi_j + \frac{1}{\sqrt{m}}A_j + \frac{1}{m}B_j.$$

Therefore, Ψ_k^2 can be expanded as

$$\sum_{j=1}^{k}\xi_j^2 + \frac{1}{\sqrt{m}}A + \frac{1}{m}B + o_p\left(\frac{1}{m}\right),$$

where the leading term $\sum_{j=1}^{k}\xi_j^2$ is the χ^2 random variable with k degrees of freedom and the second term $\frac{1}{\sqrt{m}}A$ is of order $O_p(m^{-1/2})$. To obtain a good approximation of the χ^2 distribution, a large m is preferred.

Now we turn to the estimation of the distribution function. We have

$$\begin{aligned} R_{1,m,n} &= \sum_{j=1}^{k}\frac{1}{m}\left[\sum_{i=1}^{m}\left[\pi_j\left(\widehat{Z}_i\right) - \pi_j(Z_i)\right]\right]^2 \\ &= \frac{m}{n}\sum_{j=1}^{k}\left[\frac{1}{m}\sum_{i=1}^{m}\sqrt{n}\left[\pi_j\left(\widehat{Z}_i\right) - \pi_j(Z_i)\right]\right]^2 = \frac{m}{n}C. \end{aligned}$$

Notice that $U_{ji} = \sqrt{n}\left[\pi_j\left(\widehat{Z}_i\right) - \pi_j(Z_i)\right] = O_p(1)$, $\frac{1}{m}\sum_{i=1}^{m}U_{ji} = O_p(1)$, and

$$C = \sum_{j=1}^{k}\left[\frac{1}{m}\sum_{i=1}^{m}\sqrt{n}\left[\pi_j\left(\widehat{Z}_i\right) - \pi_j(Z_i)\right]\right]^2 = O_p(1),$$

and thus $R_{1,m,n} = O_p\left(\frac{m}{n}\right)$. Similarly,

$$\begin{aligned} R_{2,m,n} &= 2\sum_{j=1}^{k}\frac{1}{m}\left[\sum_{i=1}^{m}\pi_j(Z_i)\right]\left[\sum_{i=1}^{m}\left[\pi_j\left(\widehat{Z}_i\right) - \pi_j(Z_i)\right]\right] \\ &= 2\sqrt{\frac{m}{n}}\sum_{j=1}^{k}\left[\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\pi_j(Z_i)\right]\left[\frac{1}{m}\sum_{i=1}^{m}\sqrt{n}\left[\pi_j\left(\widehat{Z}_i\right) - \pi_j(Z_i)\right]\right] \\ &= O_p\left(\sqrt{\frac{m}{n}}\right), \end{aligned}$$

since

$$D = 2\sum_{j=1}^{k}\left[\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\pi_j(Z_i)\right]\left[\frac{1}{m}\sum_{i=1}^{m}\sqrt{n}\left[\pi_j\left(\widehat{Z}_i\right) - \pi_j(Z_i)\right]\right] = O_p(1).$$

Therefore,

$$\begin{aligned} \widehat{\Psi}_k^2 &= \Psi_k^2 + R_{1,m,n} + R_{2,m,n} \\ &= \sum_{j=1}^{k}\xi_j^2 + \frac{1}{\sqrt{m}}A + \sqrt{\frac{m}{n}}D + o_p\left(\frac{1}{\sqrt{m}} + \sqrt{\frac{m}{n}}\right), \end{aligned}$$

https://doi.org/10.1017/S0266466612000370 Published online by Cambridge University Press

where $\frac{1}{\sqrt{m}}A$, $\sqrt{\frac{m}{n}}D$, etc., are higher order terms that bring size distortion in finite sample, but are $o_p(1)$. In particular, the leading terms are of order $O_p\left(\frac{1}{\sqrt{m}}\right)$ and $O_p\left(\sqrt{\frac{m}{n}}\right)$, respectively. ■

Proof of Theorem 3. We first analyze the limiting behavior of $\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\pi_j\left(\hat{Z}_i\right)$, $j = 1,\ldots,k$. Notice that

$$\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\pi_j\left(\hat{Z}_i\right) = \frac{1}{\sqrt{m}}\sum_{i=1}^{m}\pi_j\left(Z_i\right) + \frac{1}{\sqrt{m}}\sum_{i=1}^{m}\dot{\pi}_j\left(Z_i\right)\left[\hat{Z}_i - Z_i\right] + \underbrace{\mathcal{R}_{m,n}}_{\sqrt{m}/n},$$

where the remainder term $\mathcal{R}_{m,n} = O_p(\sqrt{m}/n) = o_p(1)$, and the leading term $\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\pi_j\left(Z_i\right)$ follows a simple CLT, thus we focus on the behavior of $\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\dot{\pi}_j\left(Z_i\right)\left[\hat{Z}_i - Z_i\right]$. Notice that when $m = \lambda n = m_n$, $\lambda > 0$, we may rewrite

$$\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\dot{\pi}_j\left(Z_i\right)\left[\hat{Z}_i - Z_i\right] = \sum_{l=1}^{n}\frac{1}{n}\left(\frac{1}{\sqrt{m_n}}\sum_{i=1}^{m_n}\dot{\pi}_j\left(Z_i\right)\left(I\left(X_l \le Y_i\right) - F\left(Y_i\right)\right)\right) = \sum_{l=1}^{n}\xi_{nl},$$

where $\xi_{nl} = \frac{1}{n}\left(\frac{1}{\sqrt{m_n}}\sum_{i=1}^{m_n}\dot{\pi}_j\left(Z_i\right)\left(I\left(X_l \le Y_i\right) - F\left(Y_i\right)\right)\right)$, $l = 1,\ldots,n$, are uncorrelated over l. Let $\mathcal{F}_l = \sigma\left\{X_\ell, \ell \le l; Y_i, i = 1,\ldots,m_n\right\}$, then ξ_{nl} is $\mathcal{F}_l$ measurable and $\mathbb{E}\left(\xi_{nl}|\mathcal{F}_{l-1}\right) = 0$. Thus, $\sum_{l=1}^{n}\xi_{nl}$ is a summation of uncorrelated arrays, and its limiting behavior can be analyzed using the central limit theorem method developed for triangular arrays.

Notice that

$$s_n^2 = \mathbb{E}\left(\sum_{l=1}^{n}\xi_{nl}\right)^2 = \sum_{l=1}^{n}\mathbb{E}\xi_{nl}^2 \to \lambda\delta_{jj}, \quad \text{as } n \to \infty,$$

where

$$\delta_{jj} = \mathbb{E}\left[\dot{\pi}_j\left(Z_i\right)\dot{\pi}_j\left(Z_s\right)\left\{F\left(Y_i \wedge Y_s\right) - F\left(Y_i\right)F\left(Y_s\right)\right\}\right].$$

We restandardize ξ_{nl} by s_n and verify that

$$\max_{1\le l\le n}\left|s_n^{-1}\xi_{nl}\right| \xrightarrow{P} 0.$$

Notice that

$$\Pr\left(\max_{1\le l\le n}\left|s_n^{-1}\xi_{nl}\right| > \epsilon\right) \le \frac{\sum_{1\le l\le n}\mathbb{E}\left|s_n^{-1}\xi_{nl}\right|^2 1\left(\left|s_n^{-1}\xi_{nl}\right| > \epsilon\right)}{\epsilon^2},$$

and by Hölder's inequality,

$$\mathbb{E}\left|s_n^{-1}\xi_{nl}\right|^2 1\left(\left|s_n^{-1}\xi_{nl}\right| > \epsilon\right) \le \left(\mathbb{E}\left|s_n^{-1}\xi_{nl}\right|^{2p}\right)^{1/p}\left(\mathbb{E}\left|1\left(\left|s_n^{-1}\xi_{nl}\right| > \epsilon\right)\right|^q\right)^{1/q},$$

https://doi.org/10.1017/S0266466612000370 Published online by Cambridge University Press

where $\frac{1}{p}+\frac{1}{q}=1$. We look at $\left(\mathbb{E}\left|s_n^{-1}\xi_{nl}\right|^{2p}\right)^{1/p}$ and $\left(\mathbb{E}\left|1\left(\left|s_n^{-1}\xi_{nl}\right|>\epsilon\right)\right|^{q}\right)^{1/q}$. Notice that

$$s_n^{-1}\xi_{nl}=\frac{1}{ns_n}\left(\frac{1}{\sqrt{m_n}}\sum_{i=1}^{m_n}\dot{\pi}_j\left(Z_i\right)\left(I\left(X_l\leq Y_i\right)-F\left(Y_i\right)\right)\right).$$

By Minkowski's inequality,

$$\begin{aligned}&\left[\mathbb{E}\left(\frac{1}{\sqrt{m_n}}\sum_{i=1}^{m_n}\dot{\pi}_j\left(Z_i\right)\left(I\left(X_l\leq Y_i\right)-F\left(Y_i\right)\right)\right)^{2p}\right]^{1/2p}\\&\quad\leq\sum_{i=1}^{m_n}\left[\mathbb{E}\left(\frac{1}{\sqrt{m_n}}\dot{\pi}_j\left(Z_i\right)\left(I\left(X_l\leq Y_i\right)-F\left(Y_i\right)\right)\right)^{2p}\right]^{1/2p}\\&\quad=m_n^{1/2}\left\|\dot{\pi}_j\left(Z_i\right)\left(I\left(X_l\leq Y_i\right)-F\left(Y_i\right)\right)\right\|_{2p}.\end{aligned}$$

Thus,

$$\begin{aligned}\left(\mathbb{E}\left|s_n^{-1}\xi_{nl}\right|^{2p}\right)^{1/p}&=\left(\mathbb{E}\left|\frac{1}{ns_n}\left(\frac{1}{\sqrt{m_n}}\sum_{i=1}^{m_n}\dot{\pi}_j\left(Z_i\right)\left(I\left(X_l\leq Y_i\right)-F\left(Y_i\right)\right)\right)\right|^{2p}\right)^{1/p}\\&\leq\frac{1}{\left(ns_n\right)^2}\left(m^{1/2}\left\|\dot{\pi}_j\left(Z_i\right)\left(I\left(X_l\leq Y_i\right)-F\left(Y_i\right)\right)\right\|_{2p}\right)^2\\&=\frac{\lambda}{ns_n^2}\left(\left\|\dot{\pi}_j\left(Z_i\right)\left(I\left(X_l\leq Y_i\right)-F\left(Y_i\right)\right)\right\|_{2p}\right)^2\\&\approx\frac{1}{n\delta_{jj}}\left\|\dot{\pi}_j\left(Z_i\right)\left(I\left(X_l\leq Y_i\right)-F\left(Y_i\right)\right)\right\|_{2p}^2\end{aligned}$$

and, by Minkowski's inequality again,

$$\begin{aligned}\mathbb{E}\left|1\left(\left|s_n^{-1}\xi_{nl}\right|>\epsilon\right)\right|^q&=\mathbb{E}\left|1\left(\left|s_n^{-1}\xi_{nl}\right|>\epsilon\right)\right|\\&=\Pr\left(\left|\frac{1}{ns_n}\left(\frac{1}{\sqrt{m_n}}\sum_{i=1}^{m_n}\dot{\pi}_j\left(Z_i\right)\left(I\left(X_l\leq Y_i\right)-F\left(Y_i\right)\right)\right)\right|>\epsilon\right)\\&\leq\frac{1}{\epsilon^b}\mathbb{E}\left|\frac{1}{ns_n}\left(\frac{1}{\sqrt{m_n}}\sum_{i=1}^{m_n}\dot{\pi}_j\left(Z_i\right)\left(I\left(X_l\leq Y_i\right)-F\left(Y_i\right)\right)\right)\right|^b\\&\leq\frac{1}{\epsilon^b}\frac{1}{n^b s_n^b m_n^{b/2}}\left(\sum_{i=1}^{m_n}\left(\mathbb{E}\left|\dot{\pi}_j\left(Z_i\right)\left(I\left(X_l\leq Y_i\right)-F\left(Y_i\right)\right)\right|^b\right)^{1/b}\right)^b\\&=\frac{\lambda^{b/2}}{n^{b/2}s_n^b\epsilon^b}\left(\mathbb{E}\left\|\dot{\pi}_j\left(Z_1\right)\left(I\left(X_l\leq Y_1\right)-F\left(Y_1\right)\right)\right\|_b\right)^b\\&\approx\frac{1}{n^{b/2}\delta_{jj}^{b/2}\epsilon^b}\left(\mathbb{E}\left\|\dot{\pi}_j\left(Z_i\right)\left(I\left(X_l\leq Y_i\right)-F\left(Y_i\right)\right)\right\|_b\right)^b.\end{aligned}$$

https://doi.org/10.1017/S0266466612000370 Published online by Cambridge University Press

Thus,

$$\begin{aligned}
&\mathbb{E}\left|s_n^{-1}\xi_{nl}\right|^2 1\left(\left|s_n^{-1}\xi_{nl}\right| > \epsilon\right) \\
&\quad \le \left(\mathbb{E}\left|s_n^{-1}\xi_{nl}\right|^{2p}\right)^{1/p} \left(\mathbb{E}\left|1\left(\left|s_n^{-1}\xi_{nl}\right| > \epsilon\right)\right|^q\right)^{1/q} \\
&\quad \le \frac{\lambda}{ns_n^2}\left(\left\|\dot{\pi}_j(Z_i)\left(I(X_l \le Y_i) - F(Y_i)\right)\right\|_{2p}\right)^2 \\
&\qquad \times \frac{\lambda^{b/2q}}{n^{b/2q}s_n^{b/q}\epsilon^{b/q}}\left(\mathbb{E}\left\|\dot{\pi}_j(Z_i)\left(I(X_l \le Y_i) - F(Y_i)\right)\right\|_b\right)^{b/q} \\
&\quad = \frac{\lambda^{1+b/2q}}{n^{1+b/2q}s_n^{2+b/q}\epsilon^{b/q}}\left\|\dot{\pi}_j(Z_i)\left(I(X_l \le Y_i) - F(Y_i)\right)\right\|_{2p}^2 \\
&\qquad \times \left(\mathbb{E}\left\|\dot{\pi}_j(Z_i)\left(I(X_l \le Y_i) - F(Y_i)\right)\right\|_b\right)^{b/q}.
\end{aligned}$$

Under Assumptions 1 and 2′, by appropriately chosen p, q, and b,

$$\begin{aligned}
\Pr\left(\max_{1\le l\le n}\left|s_n^{-1}\xi_{nl}\right| > \epsilon\right) &\le \frac{\sum_{1\le l\le n}\mathbb{E}\left|s_n^{-1}\xi_{nl}\right|^2 1\left(\left|s_n^{-1}\xi_{nl}\right| > \epsilon\right)}{\epsilon^2} \\
&\le \frac{1}{\epsilon^{2+b/q}}\frac{\lambda^{1+b/2q}}{n^{b/2q}s_n^{2+b/q}}\left\|\dot{\pi}_j(Z_i)\left(I(X_l \le Y_i) - F(Y_i)\right)\right\|_{2p}^2 \\
&\qquad \times \left(\mathbb{E}\left\|\dot{\pi}_j(Z_i)\left(I(X_l \le Y_i) - F(Y_i)\right)\right\|_b\right)^{b/q} \\
&\to \infty, \quad \text{as } n \to \infty.
\end{aligned}$$

By analysis of McLeish (1974),

$$\sum_{l=1}^{n} s_n^{-1}\xi_{nl} \to N(0,1).$$

Thus,

$$\left(\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\dot{\pi}_j(Z_i)\left[\hat{Z}_i - Z_i\right]\right) \to N(0, \lambda\delta_{jj}).$$

Notice that $\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\pi_j(Z_i)$ and $\frac{1}{n}\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\sum_{l=1}^{n}\dot{\pi}_j(Z_i)\left(I(X_l \le Y_i) - Z_i\right)$ are uncorrelated, we have

$$\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\pi_j\left(\hat{Z}_i\right) \Rightarrow N(0, 1 + \lambda\delta_{jj}). \tag{A.10}$$

Thus, by result of (A.10) and application of the Cramer-Wold device,

$$\Pi_k = \left(\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\pi_1\left(\hat{Z}_i\right), \ldots, \frac{1}{\sqrt{m}}\sum_{i=1}^{m}\pi_k\left(\hat{Z}_i\right)\right)$$

https://doi.org/10.1017/S0266466612000370 Published online by Cambridge University Press

converges to a k-dimensional normal variate with covariance matrix $\Omega_k = I_k + \lambda \Delta_k$, where

$$\Delta_k = \begin{bmatrix} \delta_{11} & \cdots & \delta_{1k} \\ \cdots & \ddots & \cdots \\ \delta_{k1} & \cdots & \delta_{kk} \end{bmatrix},$$

since, when $j \neq r$,

$$\mathbb{E}\left[\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\pi_j(Z_i)\right]\left[\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\pi_r(Z_i)\right] = 0 \text{ because } \pi_j(Z_i) \text{ and } \pi_r(Z_i) \text{ are orthogonal ,}$$

$$\mathbb{E}\left[\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\pi_j(Z_i)\right]\left[\frac{1}{n}\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\sum_{l=1}^{n}\dot{\pi}_r(Z_i)\left(I(X_l \leq Y_i) - Z_i\right)\right] = 0,$$

$$\mathbb{E}\left[\frac{1}{n}\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\sum_{l=1}^{n}\dot{\pi}_j(Z_i)\left(I(X_l \leq Y_i) - Z_i\right)\right]\left[\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\pi_r(Z_i)\right] = 0,$$

and

$$\mathbb{E}\left[\frac{1}{n}\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\sum_{l=1}^{n}\dot{\pi}_j(Z_i)\left(I(X_l \leq Y_i) - Z_i\right)\right]$$
$$\left[\frac{1}{n}\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\sum_{l=1}^{n}\dot{\pi}_r(Z_i)\left(I(X_l \leq Y_i) - Z_i\right)\right] \to \lambda\delta_{jr}.$$

■

https://doi.org/10.1017/S0266466612000370 Published online by Cambridge University Press