Adaptive procedures for nested processes: application to equal employment

Wenjing Xu[*], Qing Pan, and Joseph L. Gastwirth

Typically, equal employment cases concern the fairness of an employer's hiring, promotion or layoff decisions. Sometimes both the hiring and promotion practices are under scrutiny. An unappreciated issue in the analysis of promotion data is that Fisher's exact test may have low power for detecting a legally meaningful disparity in promotion rates when the number of minorities previously hired was small due to unfair hiring. The two employment processes are nested. Thus, before promotion data is analyzed, one should first check whether the hiring was fair. The choice of an appropriate test for fairness in promotions should depend on the result of the test on hiring. Two adaptive procedures are presented in this paper. One uses the Breslow-Day test as the preliminary procedure for choosing between a test having a mixture χ^2 null distribution and a test assuming a common odds ratio between the success rates of minority and majority members in both hiring and promotion. The second adaptive procedure was motivated by a court's suggestion that when the fairness of both the hiring and promotion practices are under review, rather than applying Fisher's exact test to data of promotions made from those hired, the promotions should be viewed as a random sample from the applicants for the lower level position. The statistical properties of the test procedures are presented and the tests are applied to data from four actual cases. Because the choice of the test at the second stage depends on the results of the first stage analysis, in cases concerning promotion, courts should give plaintiffs access to data concerning hiring in the feeder positions. Thus, the courts should take a broad view of potentially relevant data at the discovery stage.

1. INTRODUCTION

In 1977, the United States Supreme Court accepted statistical hypothesis testing to assess whether minorities were fairly represented on juries in *Castaneda v. Partida*[1] and

received their fair share of hires in *Hazelwood*[2]. Since then, there have been numerous cases involving the fairness of hiring and/or promotion as well as equal pay. Various nested processes arise in related aspects of fair employment litigation. For example, there can be two stages in the hiring process. First, applicants must pass a written test and only those who pass continue on to an oral exam. Alternatively, an employer may check for improper workplace behavior of its employees by a haphazard process of selecting some to be investigated and then decide to terminate a subset of those investigated. Notice that unfairness can occur at the first stage in the choice of employees to be investigated, and/or the second stage when some employees are terminated. While a single plaintiff will allege that he/she was discriminated in either hiring or promotion, a class action may concern the fairness of both. Thus, different statistical methods will be preferred depending on the nature of the legal complaint.

Gastwirth and Greenhouse (1995) and Gastwirth (1997) pointed out that the available sample size for studying the promotion process depends on the number of those who were hired. They showed that anomalous results can occur when one analyzes the data for the two processes separately. In particular, it is possible for one employer in a labor market to hire and promote more minorities than a similar employer in the same labor market, but to have a statistically significant shortfall in promotions while the second employer, who hired and promoted fewer minorities, does not have a statistically significant shortfall in either hires or promotions. This paper proposes two adaptive procedures addressing such situation and shows that they generally have higher power than the commonly used tests, which focus on only one of the two related employment practices.

The statistical models appropriate for the types of data arising in these cases are presented in Section 2, where the sampling distributions of the test statistics are given. In Section 3 two adaptive tests are described; the first is appropriate when testing the fairness of both policies. It uses the Breslow-Day test (Breslow and Day, 1980, p. 142) as a preliminary check of the equality of the odds ratios of the hiring and promotion processes. When the data are consistent with a common odds ratio, a test based on an optimal estimate of that odds ratio is used. Otherwise, a mixture χ^2 test originally proposed by Chernoff (1954) is used. The second

---

*Corresponding author.

[1] *Castaneda v. Partida*, 430 U.S. 782, (1977).

[2] *Hazelwood School District et al. v. United States*, 433 U.S. 299, (1977).

Table 1. Data Structure under Study

Hiring Data			
	+	−	Total
Minority	X	$n_1 - X$	n_1
Majority	$m - X$	$n_2 - (m - X)$	n_2
Total	m	$n - m$	n

Promotion Data			
	+	−	Total
Minority	Y	$X - Y$	X
Majority	$k - Y$	$(m - X) - (k - Y)$	$m - X$
Total	k	$m - k$	m

procedure is designed for cases where the fairness of promotions is the primary concern but minorities might have been disadvantaged in the hiring process, which formalizes an approach adopted by the courts in *Robinson v. Union Carbide*[3].

Section 4 presents the results of simulation studies examining the powers of the various procedures in different situations. The findings imply that comparing the proportions of minorities and majorities chosen for promotion from an existing pool of employees by Fisher's exact test (Finkelstein and Levin, p.154–p.157, 2001), a commonly used method, has *low* power when minorities did not receive fair treatment in the hiring process and should not be used. The proposed adaptive procedures, however, have good power relative to other methods in this situation. This result implies that when courts allow discovery in promotion cases it is important that plaintiffs have access to data concerning the hiring or promotion to the feeder positions. The methods are illustrated on data from actual cases in Section 5.

2. THE MATHEMATICAL FORMULATION OF THE PROBLEM AND SAMPLING DISTRIBUTION OF THE STATISTICS

The data for analysis are summarized in Table 1, where n_i $(i = 1, 2)$ denotes the number of minority or majority applicants, respectively. Let X denote the number of minorities among the m hires. From the m hires, during the relevant time period, k are promoted, Y of whom are minority. The estimates of the log odds ratios of the hiring and promotion processes ($\hat{\theta}_h$ and $\hat{\theta}_p$) are

$$(1) \qquad \hat{\theta}_h = \log \frac{X/(n_1 - X)}{(m - X)/(n_2 - (m - X))},$$

$$(2) \qquad \hat{\theta}_p = \log \frac{Y/(X - Y)}{(k - Y)/((m - X) - (k - Y))}.$$

Under the null hypothesis $H_0 : \theta_h = \theta_p = 0$, $X \sim$ H(n, n_1, m) and $Y|X = x \sim$ H(m, x, k), where H(N, M, n) is the hypergeometric distribution with parameters N, M

[3]*Robinson v. Union Carbide*, 538 F. 2d 652 (5th Cir. 1976).

and n, that is, there are N subjects, of which M are special, and a sample of size n is drawn from all N.

Theorem 1. *Assume $X \sim H(x; n, n_1, m)$ and $Y|X = x \sim H(y; m, x, k)$. If n, n_1, m, $k \to \infty$ at the same rate, i.e., $\frac{n_1}{n} \to \gamma$, $\frac{m}{n} \to \lambda_1$, $\frac{k}{n} \to \lambda_2$ and $u_{x,n} := \sqrt{n}(x/n - \gamma\lambda_1) \to u_1$, $u_{y,n} := \sqrt{n}(y/n - \gamma\lambda_2) \to u_2$, then*

$$\sqrt{n}\left(\frac{X/n - \mu_1}{\sigma_1}, \frac{Y/n - \mu_2}{\sigma_2}\right)' \xrightarrow{\mathcal{L}} N\left((0,0)', \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right),$$

where $\mu_i = \gamma\lambda_i$, $\sigma_i = \sqrt{\gamma(1-\gamma)\lambda_i(1-\lambda_i)}$ $(i = 1, 2)$, and $\rho = \sqrt{\frac{\lambda_2(1-\lambda_1)}{\lambda_1(1-\lambda_2)}}$.

The proof utilizes the conditional distribution of promotions given hires and the joint distribution is obtained by removing the condition using the result of Fligner (1981), which extended results of Sethuraman (1961), Fligner and Hettmansperger (1979).

Proof. It is convenient to consider the random variables $U_{X,n} = \sqrt{n}(X/n - \gamma\lambda_1)$ and $U_{Y,n} = \sqrt{n}(Y/n - \gamma\lambda_2)$. Under the stated conditions, Feller (1968, p. 194) and Pinsky (2008) showed that

$$P(X = x) \sim \frac{\exp\left\{-\frac{(x - n\gamma\lambda_1)^2}{2n\gamma(1-\gamma)\lambda_1(1-\lambda_1)}\right\}}{\sqrt{2\pi n\gamma(1-\gamma)\lambda_1(1-\lambda_1)}},$$

which implies that

$$(3) \qquad P(U_{X,n} = u_{x,n}) \sim \frac{\exp\left\{-\frac{u_1^2}{2\gamma(1-\gamma)\lambda_1(1-\lambda_1)}\right\}}{\sqrt{2\pi\gamma(1-\gamma)\lambda_1(1-\lambda_1)}}.$$

Equivalently,

$$P(U_{X,n} = u_{x,n}) \sim f_{U_1}(u_1),$$

where $U_1 \sim N(0, \gamma(1-\gamma)\lambda_1(1-\lambda_1))$ with probability density function of $f_{U_1}(u_1)$. Using the same technique of Feller and Pinsky's, one can obtain

$$P(Y = y|X = x) \sim \frac{\exp\left\{-\frac{\left(y - \frac{\lambda_2}{\lambda_1}x\right)^2}{2n\lambda_2\gamma(1-\gamma)\left(1 - \frac{\lambda_2}{\lambda_1}\right)}\right\}}{\sqrt{2\pi\lambda_2 n\gamma(1-\gamma)\left(1 - \frac{\lambda_2}{\lambda_1}\right)}},$$

which implies that

$$(4) \qquad P(U_{Y,n} = u_{y,n}|U_{X,n} = u_{x,n})$$

$$\sim \frac{\exp\left\{-\frac{\left(u_2 - \frac{\lambda_2}{\lambda_1}u_1\right)^2}{2\gamma(1-\gamma)\lambda_2\left(1 - \frac{\lambda_2}{\lambda_1}\right)}\right\}}{\sqrt{2\pi\gamma(1-\gamma)\lambda_2\left(1 - \frac{\lambda_2}{\lambda_1}\right)}}.$$

Let U_2 be a random variable such that its conditional distribution given $U_1 = u_1$ is $N\left(\frac{\lambda_2}{\lambda_1}u_1, \gamma(1-\gamma)\lambda_2\left(1 - \frac{\lambda_2}{\lambda_1}\right)\right)$. Thus (4) implies

$$U_{Y,n}|U_{X,n} = u_{x,n} \xrightarrow{\mathcal{L}} U_2|U_1 = u_1 \sim N\left(\frac{\lambda_2}{\lambda_1}u_1, \gamma(1-\gamma)\lambda_2\left(1 - \frac{\lambda_2}{\lambda_1}\right)\right).$$

Since $U_{X,n} = \sqrt{n}(X/n - \gamma\lambda_1)$ are discrete random variables on a lattice of the form $L_x = \{u_{x,n}, x = 0, 1, \ldots, m\}$, it follows from *Theorem 1* of Fligner (1981) that

$$(U_{X,n}, U_{Y,n})' \overset{\mathcal{L}}{\to} N\left((0,0)', \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right),$$

where $\mu_i = \gamma\lambda_i$, $\sigma_i = \sqrt{\gamma(1-\gamma)\lambda_i(1-\lambda_i)}$ $(i = 1, 2)$, and $\rho = \sqrt{\frac{\lambda_2(1-\lambda_1)}{\lambda_1(1-\lambda_2)}}$. Thus, we have

$$\sqrt{n}\left(\frac{X/n - \mu_1}{\sigma_1}, \frac{Y/n - \mu_2}{\sigma_2}\right)' \to N\left((0,0)', \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right),$$

where $\mu_i = \gamma\lambda_i$, $\sigma_i = \sqrt{\gamma(1-\gamma)\lambda_i(1-\lambda_i)}$, $(i = 1, 2)$ and $\rho = \sqrt{\frac{\lambda_2(1-\lambda_1)}{\lambda_1(1-\lambda_2)}}$. Details are provided in Appendix A.1. □

In the notation of Theorem 1, $n_1 \sim n\gamma$, $m \sim n\lambda_1$ and $k \sim n\lambda_2$, expressions (1) and (2) are asymptotically equivalent to

$$\hat{\theta}_h \sim \log \frac{\frac{X}{n}/\left(\gamma - \frac{X}{n}\right)}{\left(\lambda_1 - \frac{X}{n}\right)/\left[(1 - \gamma - \lambda_1) + \frac{X}{n}\right]},$$

$$\hat{\theta}_p \sim \log \frac{\frac{Y}{n}/\left(\frac{X}{n} - \frac{Y}{n}\right)}{\left(\lambda_2 - \frac{Y}{n}\right)/\left[(\lambda_1 - \lambda_2) - \frac{X}{n} + \frac{Y}{n}\right]}.$$

Using the delta method (Appendix A.2), Theorem 1 implies

Corollary 1. $\sqrt{n}(\hat{\theta}_h, \hat{\theta}_p)' \overset{H_0}{\to} N\left((0,0)', \Omega_0\right)$, where

$$\Omega_0 = \begin{pmatrix} \frac{1}{\gamma(1-\gamma)\lambda_1(1-\lambda_1)} & 0 \\ 0 & \frac{1}{\gamma(1-\gamma)\lambda_2\left(1-\frac{\lambda_2}{\lambda_1}\right)} \end{pmatrix}.$$

Sometimes the available data on applicants are incomplete or do not accurately reflect the true qualified labor pool for the job. This can happen when the jobs were advertised in media primarily directed to a majority audience, when a company has a history or reputation for not giving minorities better paying positions or when a substantial fraction of applications are duplicates. Courts then use the demographic method, which determines the minority share (γ) of the qualified labor pool from Census data on the education and labor force participation of the population in the area (Rosenblum, 1977, 1978). Methods have been developed to estimate γ to account for commuting patterns (Dorseano, 1975, Gastwirth and Haber, 1976, Haber and Gastwirth, 1978) and the salary of the entry level position (Gastwirth, 1981). Then the number of minorities X among the m hires is assumed to have a binomial distribution with parameters m and γ, i.e., $X \sim Binomial(m, \gamma)$ and $Y|X = x \sim H(m, X, k)$. The analog of Theorem 1 is:

Theorem 2. *Assume that $X \sim Binomial(m, \gamma)$ and $Y|X = x \sim H(m, x, k)$, $m, k \to \infty$ at the same rate, i.e. $\frac{k}{m} \to \lambda$, then*

$$\sqrt{m}\left(\frac{X/m - \mu_1}{\sigma_1}, \frac{Y/m - \mu_2}{\sigma_2}\right)' \overset{\mathcal{L}}{\to} N\left((0,0)', \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right),$$

where $\mu_1 = \gamma$, $\mu_2 = \gamma\lambda$, $\sigma_1 = \sqrt{\gamma(1-\gamma)}$, $\sigma_2 = \sqrt{\lambda\gamma(1-\gamma)}$ and $\rho = \sqrt{\lambda}$.

Gastwirth and Greenhouse (1987) noted that the binomial model is appropriate when the number of hires is small relative to the external labor pool. They showed that the maximum likelihood estimate of the relative risk is $\frac{X/(m-X)}{\gamma/(1-\gamma)}$, which can be interpreted as the ratio of the odds of randomly selecting a member of the minority group from the m hires to the odds $\gamma/(1-\gamma)$ of selecting a minority member from the large external labor pool. Following Gastwirth and Greenhouse (1987), in the notation of Theorem 2, $\hat{\theta}_h = \log \frac{X(1-\gamma)}{(m-X)\gamma}$ and $\hat{\theta}_p \sim \log \frac{Y/(X-Y)}{(m\lambda-Y)/(m(1-\lambda)-X+Y)}$. Their joint asymptotic distribution is obtained by applying the delta method (Appendix A.3) to Theorem 2. Formally, this yields the analog of Corollary 1

Corollary 2. $\sqrt{m}(\hat{\theta}_h, \hat{\theta}_p)' \overset{H_0}{\to} N\left((0,0)', \Omega_0\right)$, where $\Omega_0 = \begin{pmatrix} \frac{1}{\gamma(1-\gamma)} & 0 \\ 0 & \frac{1}{\lambda(1-\lambda)\gamma(1-\gamma)} \end{pmatrix}$.

2.1 A mixture χ^2 test

Sometimes, the problem of interest is to test whether discriminatory practices affect both the hiring and promotion processes. Thus, one tests $H_0 : \theta_h = \theta_p = 0$ against H_a: $\theta_h < 0$ and $\theta_p < 0$.

Chernoff (1954) developed a procedure for testing hypotheses of this type, which was generalized by Self and Liang (1987). When Z has a multivariate normal distribution with mean θ, where $\theta \in \Theta$ is in a k-dimensional space, and known covariance matrix Ω, for testing $H_0 : \theta \in \Theta_0$ vs. $H_a : \theta \in \Theta_1$, Chernoff showed that the likelihood ratio test statistic is

$$T = -2\log \frac{\sup_{\theta \in \Theta_1} (2\pi)^{-k/2} |\Omega|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(z-\theta)'|\Omega|^{-1}(z-\theta)\right\}}{\sup_{\theta \in \Theta_0} (2\pi)^{-k/2} |\Omega|^{-1} \exp\left\{-\frac{1}{2}(z-\theta)'|\Omega|^{-\frac{1}{2}}(z-\theta)\right\}}$$

$$= \inf_{\theta \in \Theta_0} (z-\theta)'|\Omega|^{-\frac{1}{2}}(z-\theta) - \inf_{\theta \in \Theta_1} (z-\theta)'|\Omega|^{-\frac{1}{2}}(z-\theta).$$

Let $P\Lambda P'$ be the spectral decomposition of Ω^{-1}, then one obtains

$$(z-\theta)'\Omega^{-1}(z-\theta)$$
$$= (z-\theta)P\Lambda P'(z-\theta)$$
$$= \left[\Lambda^{\frac{1}{2}}P'(z-\theta)\right]'\left[\Lambda^{\frac{1}{2}}P'(z-\theta)\right]$$
$$= \left(\Lambda^{\frac{1}{2}}P'z - \Lambda^{\frac{1}{2}}P'\theta\right)'\left(\Lambda^{\frac{1}{2}}P'z - \Lambda^{\frac{1}{2}}P'\theta\right).$$

Let $\tilde{z} = \Lambda^{\frac{1}{2}}P'z$ and $\tilde{\theta} = \Lambda^{\frac{1}{2}}P'\theta$, then the likelihood ratio test statistic is

$$T = \inf_{\tilde{\theta} \in \tilde{\Theta}_0} \|\tilde{z} - \tilde{\theta}\|^2 - \inf_{\tilde{\theta} \in \tilde{\Theta}_1} \|\tilde{z} - \tilde{\theta}\|^2,$$

where $\tilde{\Theta}_i = \left\{\tilde{\theta} : \tilde{\theta} = \Lambda^{\frac{1}{2}}P'\theta \text{ for all } \theta \in \Theta_i\right\}$ $(i = 0, 1)$. For our test of interest, $\Theta_0 = (0, 0)$ and $\Theta_1 = (-\infty, 0) \times (-\infty, 0)$. Thus $\inf_{\tilde{\theta} \in \tilde{\Theta}_0} \|\tilde{z} - \tilde{\theta}\|^2 = \|\tilde{z}\|^2$. This implies that

$$\inf_{\hat{\theta}\in\hat{\Theta}_1} \|\tilde{z} - \tilde{\theta}\|^2$$

$$= \begin{cases} \sum_{i=1}^{2} \tilde{z}_i^2 & : \text{if } \tilde{z} = (\tilde{z}_1, \tilde{z}_2) \in (0, +\infty) \times (0, +\infty) \\ \tilde{z}_2^2 & : \text{if } \tilde{z} = (\tilde{z}_1, \tilde{z}_2) \in (-\infty, 0) \times (0, +\infty) \\ 0 & : \text{if } \tilde{z} = (\tilde{z}_1, \tilde{z}_2) \in (-\infty, 0) \times (-\infty, 0) \\ \tilde{z}_1^2 & : \text{if } \tilde{z} = (\tilde{z}_1, \tilde{z}_2) \in (0, +\infty) \times (-\infty, 0). \end{cases}$$

Therefore, the likelihood ratio test statistic is

$$(5) \quad T = \sum_{i=1}^{2} \tilde{z}_i^2 I(\tilde{z}_i < 0) \overset{H_0}{\sim} \rho\chi_2^2 + (1 - 2\rho)\chi_1^2 + \rho\chi_0^2,$$

where χ_j^2 $(j = 0, 1, 2)$ is a χ^2 distribution with j degrees of freedom. Self and Liang (1987) give the general formula for ρ:

$$\rho = \frac{cos^{-1}\left(\frac{I_{21}}{\sqrt{I_{11}I_{22}}}\right)}{2\pi},$$

where I_{ij}'s are the (i, j) entries of the information matrix $I(\theta) = \Omega^{-1}$. As the two log odds ratios $\hat{\theta}_h$ and $\hat{\theta}_p$ are independent under the null hypothesis, $I_{21} = 0$, implying $\rho = \frac{1}{4}$. Thus, the asymptotic distribution of the test statistic T in (5) under the null is $\frac{1}{4}\chi_2^2 + \frac{1}{2}\chi_1^2 + \frac{1}{4}\chi_0^2$. This result agrees with Example 3 in Chernoff (1954).

2.2 Common log odds ratio test

In his dissent in *Watson v. Fort Worth Bank (1988)*[4], the late Judge I. Goldberg observed that "If an employer discriminates against black applicants, a plausible inference is that he does not suddenly become a Frederick Douglass to his employees." He then noted that when the same decision-makers are involved in both hiring and promotion, statistical evidence concerning each of the employment practices reinforces the other. This implies that if an employer discriminates, this will be reflected in lower odds ratios of minorities to majorities in both processes, i.e. θ_h and θ_p are likely to be in the same direction and possibly close. When $\theta_h = \theta_p = \theta_c$, one can use both the hiring and promotion data to estimate θ_c. The common log odds ratio θ_c is estimated by a weighted average of $\hat{\theta}_h$ and $\hat{\theta}_p$, which is $\hat{\theta}_c = a\hat{\theta}_h + (1 - a)\hat{\theta}_p$, where a is chosen to minimize the variance of $\hat{\theta}_c$. Standard calculus yields $a = \frac{var(\hat{\theta}_p) - 2cov(\hat{\theta}_h, \hat{\theta}_p)}{var(\hat{\theta}_h) + var(\hat{\theta}_p) - 2cov(\hat{\theta}_h, \hat{\theta}_p)}$, and $var(\hat{\theta}_c) = \frac{var(\hat{\theta}_h)var(\hat{\theta}_p)}{var(\hat{\theta}_h) + var(\hat{\theta}_p) - 2cov(\hat{\theta}_h, \hat{\theta}_p)}$.

From Corollary 1 and Corollary 2, $\hat{\theta}_c$ is asymptotically normal. Thus for testing $H_0 : \theta_c = 0$ against $H_a : \theta_c < 0$, the standardized statistic $z = \frac{\hat{\theta}_c}{se(\hat{\theta}_c)} \overset{H_0}{\sim} N(0, 1)$ is appropriate.

---

[4] *Watson v. Fort Worth Bank & Trust*, 798 F.2d 791 (5th Cir., 1986). Ultimately the U.S. Supreme Court reversed the majority opinion, *Watson v. Fort Worth Bank & Trust*, 481 U.S. 977 (1988), conforming the appropriateness of Judge Goldberg's dissent opinion.

2.3 A comparison of the mixture χ^2 test and common log odds ratio test

A small simulation study was conducted to assess the performance of the mixture χ^2 and common log odds ratio tests. There are 800 applicants, of whom 400 are hired and 200 are subsequently promoted, and the majority forms three-fourths of the original applicants. The number of minorities hired and promoted are studied for different pairs of odds ratios θ_h and θ_p. Ten thousand replications are examined for each setting. Table 2 reports the power and size of the mixture χ^2 and common log odds ratio tests of the $H_0 : \theta_h = \theta_p = 0$ vs. $H_a : \theta_h < 0$ and $\theta_p < 0$, under different combinations of the true values of two odds ratios at the 0.05 significance level.

First, when the null hypothesis is true, the empirical significance level of both tests agrees with their nominal level (0.05). When the odds ratios of two stages are the same or close, the common log odds ratio test performs better than the mixture χ^2 test. When the odds ratios are noticeably different, the mixture χ^2 test has more power, especially in the upper right corner of Table 2, pertaining to the situation when minorities do well at the hiring stage but have few opportunities for promotion.

3. ADAPTIVE PROCEDURES

3.1 An adaptive procedure using Breslow-Day test as the preliminary check

The study in Section 2 suggests that an adaptive procedure (Hogg, 1974; Neuhäuser, 2001; O'Gorman, 2004; Miao and Gastwirth, 2009), which checks whether the data are consistent with the assumption of a common log odds ratio might be useful. When a procedure is used to select the statistical test for the main inference, it is preferable to use a significance level greater than 0.05 (Bancroft, 1964; Huber, 1972). Here, we use the Breslow-Day (BD) test at significance level .20 as the preliminary procedure to decide which test to use at the analysis stage. If the BD test fails to reject the hypothesis of a common odds ratio, the common log odds ratio test is applied to both the hiring and promotion data; otherwise the mixture χ^2 test is used.

Because the promotions are made from individuals who were hired, the numbers (X and Y) of minorities hired and promoted are not independent. However, they are conditionally independent. Furthermore, similar to the derivation of the asymptotic null distribution of the logrank test (Mantel, 1966; Fleming and Harrington, p. 112, 1991), the impact of this dependence is likely to be small. A simulation study in Appendix A.4 also shows that the actual size of BD test using the χ^2 distribution is close to the nominal value.

Table 2. Power of the Mixture χ^2 and Common Log Odds Ratio Tests

Test	Odds ratio at stage I	Odds Ratio at stage II							
		$\frac{5}{4}$	$\frac{6}{5}$	1	$\frac{5}{6}$	$\frac{4}{5}$	$\frac{3}{4}$	$\frac{2}{3}$	$\frac{1}{2}$
Mixture χ^2 test	$\frac{5}{4}$	0.002	0.002	0.018	0.104	0.142	0.212	0.390	0.837
Common log odds ratio test		0.001	0.001	0.003	0.014	0.014	0.021	0.039	0.158
Mixture χ^2 test	$\frac{6}{5}$	0.002	0.004	0.021	0.110	0.141	0.215	0.379	0.833
Common log odds ratio test		0.002	0.002	0.005	0.019	0.024	0.030	0.057	0.208
Mixture χ^2 test	1	0.023	0.023	**0.045**	0.151	0.192	0.270	0.442	0.861
Common log odds ratio test		0.013	0.017	**0.046**	0.114	0.137	0.178	0.260	0.547
Mixture χ^2 test	$\frac{5}{6}$	0.182	0.180	0.226	0.350	0.381	0.458	0.592	0.900
Common log odds ratio test		0.101	0.115	0.229	0.386	0.423	0.490	0.600	0.836
Mixture χ^2 test	$\frac{4}{5}$	0.258	0.266	0.295	0.417	0.451	0.520	0.648	0.917
Common log odds ratio test		0.150	0.169	0.285	0.470	0.513	0.566	0.668	0.886
Mixture χ^2 test	$\frac{3}{4}$	0.400	0.406	0.438	0.545	0.559	0.632	0.738	0.942
Common log odds ratio test		0.224	0.255	0.415	0.595	0.622	0.694	0.780	0.933
Mixture χ^2 test	$\frac{2}{3}$	0.682	0.673	0.704	0.776	0.788	0.818	0.881	0.976
Common log odds ratio test		0.440	0.468	0.639	0.797	0.818	0.852	0.915	0.979
Mixture χ^2 test	$\frac{1}{2}$	0.985	0.985	0.988	0.991	0.992	0.993	0.997	1.000
Common log odds ratio test		0.880	0.895	0.960	0.987	0.991	0.994	0.998	1.000

3.2 An adaptive procedure based on a court's suggestion

Several cases filed a few years after the passage of the Civil Rights Act, concerned the fairness of promotions at companies that only employed African-Americans in low paying positions. In *Robinson v. Union Carbide* (1976) the minority share of hires was consistent with their share of the local area's labor force; however, they were concentrated in menial jobs, i.e., they were segregated in the least favorable positions and were rarely promoted. The court decided to evaluate whether they were treated fairly in promotion by comparing their share of promotions to their share of the local area labor force and found the defendant had discriminated in promotions. The court essentially used an adaptive procedure to check the fairness of the employer's policy in assembling the candidate pool for promotions, i.e. those who were hired and subsequently eligible for promotion, to decide on the method for analyzing the promotion data. In most cases today applicant data are usually available and reliable, so an analog of the court's method is the following adaptive procedure. First, test the fairness of hiring using two-sided Fisher's exact test at the .20 level to determine whether the probability a minority applicant is hired is the same as that of a majority applicant. When the test indicates fair hiring or over-hiring, one applies two-sided Fisher's exact test to the promotion data from those hired. When the first stage test indicates possible unfairness in hiring, a one-sided Fisher's exact test is used to evaluate whether the promotions are consistent with a random sample from the applicant pool.

3.3 A comparison of the six testing procedures

As in Section 2.3, the statistical properties of six tests are evaluated in a simulation study. Again 10,000 datasets of 800 applicants, 400 of whom are hired and 200 subsequently promoted, where the majority forms three-fourths of the original applicants, are simulated. The same combinations of hiring and promotion odds ratios in Section 2.3 are used here. The powers in Table 3 are listed in the order of: the first adaptive procedure using the BD test as the preliminary check, mixture χ^2 test, common log odds ratio test, two-sided Fisher's exact test on promotion data, one-sided Fisher's exact test on promotion data, and the second adaptive procedure based on court's suggestion. The results indicate that all six procedures preserve the nominal .05 level of significance.

When the primary interest is examining whether promotions were made fairly when the hiring process was unfair, Fisher's exact test has low power (see the forth test in Table 3). For instance, when the first and second stage odds ratios are $\frac{3}{4}$ and $\frac{2}{3}$ respectively, the power of a two-sided Fisher's exact test is .34. In contrast, the powers of the first adaptive procedure, the mixture χ^2, the common log odds ratio tests and the second adaptive procedure are .77, .74, .78 and .68, respectively. The adaptive procedures perform well, because the discrimination at the hiring stage reduces the minority sample size in the analysis of the promotion data. When the hiring process is fair, the adaptive method inspired by the court's approach performs as well as the two-sided Fisher's exact test applied to promotions from hired employees.

Table 3. Power Comparison of the Six Procedures

Odds ratio at stage I	Odds ratio at stage II							
	$\frac{5}{4}$	$\frac{6}{5}$	1	$\frac{5}{6}$	$\frac{4}{5}$	$\frac{3}{4}$	$\frac{2}{3}$	$\frac{1}{2}$
$\frac{5}{4}$	0.001	0.001	0.006	0.044	0.061	0.097	0.208	0.682
	0.002	0.002	0.018	0.104	0.142	0.212	0.390	0.837
	0.001	0.001	0.003	0.014	0.014	0.021	0.039	0.158
	0.151	0.104	0.040	0.109	0.145	0.216	0.387	0.839
	0.003	0.006	0.040	0.177	0.229	0.322	0.517	0.904
	0.151	0.104	0.040	0.110	0.146	0.217	0.389	0.840
$\frac{6}{5}$	0.002	0.002	0.008	0.049	0.062	0.099	0.207	0.672
	0.002	0.004	0.021	0.110	0.141	0.215	0.379	0.833
	0.002	0.002	0.005	0.019	0.024	0.030	0.057	0.208
	0.131	0.108	0.037	0.106	0.141	0.215	0.397	0.839
	0.003	0.004	0.040	0.176	0.221	0.324	0.525	0.904
	0.130	0.108	0.038	0.110	0.144	0.217	0.399	0.840
1	0.014	0.017	**0.046**	0.120	0.148	0.204	0.325	0.740
	0.023	0.023	**0.045**	0.151	0.192	0.270	0.442	0.861
	0.013	0.017	**0.046**	0.114	0.137	0.178	0.260	0.547
	0.138	0.100	**0.037**	0.102	0.133	0.214	0.376	0.819
	0.004	0.006	**0.037**	0.168	0.213	0.304	0.504	0.891
	0.129	0.096	**0.051**	0.129	0.167	0.250	0.414	0.836
$\frac{5}{6}$	0.121	0.125	0.224	0.378	0.412	0.483	0.594	0.858
	0.182	0.180	0.226	0.350	0.381	0.458	0.592	0.900
	0.101	0.115	0.229	0.386	0.423	0.490	0.600	0.836
	0.133	0.099	0.037	0.095	0.129	0.198	0.362	0.791
	0.004	0.006	0.037	0.156	0.208	0.297	0.482	0.874
	0.110	0.095	0.121	0.262	0.304	0.380	0.547	0.876
$\frac{4}{5}$	0.175	0.194	0.283	0.460	0.501	0.552	0.660	0.892
	0.258	0.266	0.295	0.417	0.451	0.520	0.648	0.917
	0.150	0.169	0.285	0.470	0.513	0.566	0.668	0.886
	0.130	0.096	0.038	0.097	0.127	0.195	0.354	0.793
	0.004	0.005	0.037	0.159	0.202	0.293	0.483	0.874
	0.105	0.101	0.150	0.313	0.350	0.441	0.592	0.898
$\frac{3}{4}$	0.282	0.293	0.416	0.585	0.609	0.678	0.767	0.933
	0.400	0.406	0.438	0.545	0.559	0.632	0.738	0.942
	0.224	0.255	0.415	0.595	0.622	0.694	0.780	0.933
	0.128	0.097	0.034	0.094	0.129	0.185	0.341	0.781
	0.004	0.006	0.039	0.160	0.206	0.284	0.473	0.866
	0.102	0.118	0.214	0.397	0.440	0.528	0.675	0.921
$\frac{2}{3}$	0.540	0.549	0.653	0.787	0.810	0.843	0.906	0.976
	0.682	0.673	0.704	0.776	0.788	0.818	0.881	0.976
	0.440	0.468	0.639	0.797	0.818	0.852	0.915	0.979
	0.122	0.094	0.040	0.092	0.120	0.188	0.328	0.770
	0.004	0.006	0.039	0.157	0.195	0.286	0.452	0.860
	0.151	0.187	0.349	0.582	0.618	0.695	0.815	0.964
$\frac{1}{2}$	0.958	0.963	0.974	0.988	0.990	0.994	0.997	1.000
	0.985	0.985	0.988	0.991	0.992	0.993	0.997	1.000
	0.880	0.895	0.960	0.987	0.991	0.994	0.998	1.000
	0.112	0.089	0.039	0.085	0.114	0.164	0.291	0.719
	0.005	0.008	0.041	0.143	0.187	0.257	0.415	0.815
	0.502	0.547	0.747	0.893	0.914	0.947	0.974	0.998

Notes: for each combination of the two odds ratios at stages I and II, the six values correspond to the power of the following six tests: the first adaptive procedure using the Breslow-Day test as the preliminary check, mixture χ^2 test, common log odds ratio test, two-sided Fisher's exact test on promotion data assuming the promotions are made from the hires, one-sided Fisher's exact test on promotion data assuming the promotions are made from the hires, and the second adaptive procedure based on court's suggestion. $n_1 = 200, n_2 = 600, h = 400, p = 200$, 10,000 simulations in each setup.

Odds ratio at stage I	Odds ratio at stage II							
	$\frac{5}{4}$	$\frac{6}{5}$	1	$\frac{5}{6}$	$\frac{4}{5}$	$\frac{3}{4}$	$\frac{2}{3}$	$\frac{1}{2}$
$\frac{5}{4}$	0.003	0.005	0.011	0.031	0.041	0.060	0.116	0.334
	0.005	0.008	0.022	0.068	0.081	0.116	0.204	0.502
	0.003	0.003	0.007	0.015	0.020	0.023	0.043	0.107
	0.098	0.076	0.043	0.088	0.099	0.156	0.268	0.657
	0.004	0.009	0.038	0.125	0.144	0.210	0.341	0.738
	0.097	0.076	0.044	0.089	0.099	0.156	0.269	0.657
$\frac{6}{5}$	0.004	0.005	0.017	0.037	0.046	0.071	0.118	0.344
	0.007	0.007	0.025	0.073	0.087	0.125	0.205	0.509
	0.004	0.006	0.013	0.022	0.028	0.036	0.053	0.132
	0.089	0.075	0.042	0.090	0.110	0.154	0.260	0.640
	0.005	0.008	0.039	0.123	0.154	0.210	0.339	0.717
	0.089	0.075	0.043	0.091	0.111	0.156	0.262	0.641
1	0.020	0.028	**0.053**	0.093	0.112	0.136	0.190	0.424
	0.023	0.028	**0.052**	0.105	0.125	0.164	0.247	0.539
	0.020	0.028	**0.052**	0.086	0.103	0.119	0.158	0.312
	0.085	0.073	**0.042**	0.083	0.108	0.152	0.258	0.626
	0.005	0.009	**0.036**	0.120	0.150	0.210	0.335	0.711
	0.085	0.074	**0.051**	0.104	0.130	0.175	0.281	0.640
$\frac{5}{6}$	0.082	0.092	0.143	0.225	0.253	0.281	0.365	0.588
	0.096	0.101	0.134	0.203	0.226	0.258	0.354	0.627
	0.079	0.092	0.148	0.229	0.254	0.281	0.358	0.558
	0.089	0.074	0.041	0.078	0.105	0.144	0.256	0.612
	0.005	0.009	0.034	0.107	0.145	0.197	0.329	0.696
	0.104	0.100	0.123	0.202	0.229	0.274	0.394	0.699
$\frac{4}{5}$	0.104	0.115	0.176	0.267	0.293	0.337	0.413	0.629
	0.125	0.136	0.171	0.242	0.265	0.306	0.394	0.655
	0.099	0.110	0.178	0.273	0.298	0.344	0.412	0.608
	0.082	0.073	0.039	0.090	0.100	0.146	0.243	0.608
	0.005	0.010	0.037	0.125	0.141	0.200	0.319	0.687
	0.111	0.117	0.147	0.249	0.274	0.323	0.426	0.723
$\frac{3}{4}$	0.155	0.171	0.238	0.342	0.379	0.412	0.499	0.698
	0.193	0.198	0.232	0.316	0.340	0.373	0.461	0.698
	0.142	0.165	0.243	0.353	0.385	0.420	0.505	0.694
	0.082	0.069	0.043	0.087	0.101	0.139	0.238	0.605
	0.006	0.009	0.037	0.119	0.143	0.192	0.313	0.687
	0.145	0.155	0.205	0.326	0.352	0.402	0.503	0.781
$\frac{2}{3}$	0.279	0.302	0.384	0.497	0.521	0.566	0.640	0.814
	0.353	0.367	0.406	0.471	0.490	0.524	0.599	0.791
	0.254	0.280	0.388	0.504	0.531	0.577	0.653	0.818
	0.081	0.067	0.039	0.078	0.100	0.141	0.231	0.582
	0.008	0.009	0.037	0.111	0.138	0.193	0.303	0.665
	0.232	0.245	0.355	0.486	0.514	0.565	0.650	0.855
$\frac{1}{2}$	0.695	0.701	0.775	0.833	0.847	0.874	0.911	0.964
	0.797	0.797	0.826	0.851	0.860	0.878	0.899	0.952
	0.609	0.630	0.747	0.829	0.849	0.881	0.914	0.967
	0.080	0.068	0.044	0.073	0.095	0.129	0.220	0.558
	0.006	0.011	0.038	0.106	0.131	0.172	0.288	0.637
	0.618	0.650	0.745	0.846	0.864	0.882	0.923	0.979

Notes: for each combination of the two odds ratios at stages I and II, the six values correspond to the power of the following six tests: the first adaptive procedure using the Breslow-Day test as the preliminary check, mixture χ^2 test, common log odds ratio test, two-sided Fisher's exact test on promotion data assuming the promotions are made from the hires, one-sided Fisher's exact test on promotion data assuming the promotions are made from the hires, and the second adaptive procedure based on court's suggestion. $n_1 = 200, n_2 = 600, h = 400, p = 200$, 10,000 simulations in each setup.

Table 5. Pass and Promotion Data in Kirkland

Pass	+	−	Total
Minority	148	21	169
Majority	527	43	570
Total	675	64	739

Promotion	+	−	Total
Minority	21	127	148
Majority	204	323	527
Total	225	450	675

When one is concerned with detecting discriminatory practices that affect both hiring and promotion, the common log odds ratio test tends to have more power compared to the mixture χ^2 test when the two odds ratios are close. The first adaptive procedure, using the BD test as the preliminary check for the homogeneity of hiring and promotion odds, has power between those of the mixture χ^2 and common log odds ratio tests. Because the adaptive method does not incur a serious loss of power relative to either the mixture χ^2 or common log odds ratio tests in the settings where they have the highest power, the first adaptive method is recommended for general use.

In the real world some individuals who are hired leave fairly early and would not be considered as candidates for subsequent promotion. To assess the impact of this phenomenon on the statistical procedures, a simulation explored the situation where a random 30% of all individuals hired left employment before promotions were made. The empirical powers are presented in Table 4. As expected, random leaving reduced the power of the tests; however, the relative powers of the six statistical tests remained similar to the earlier results. Other simulations were carried for samples half the size of those reported here. While the powers of the tests were lower, the relative powers of the six tests considered remained the same. The results are not presented here but are available from the corresponding author.

4. APPLICATION TO DATA FROM ACTUAL CASES

In this section data from four cases, which involve a nested two-stage data structure, will be reanalyzed by the methods described previously. Many equal employment cases concern whether an exam has a disparate impact on a legally protected group. Often, only the pass rates are compared, however, not everyone who passes will ultimately be appointed to the job. When the positions are filled in rank-order, if the scores of minority group candidates are concentrated at the lower end of the passers, they will receive few actual positions. Three cases from New York, Akron and St. Louis illustrate the situation. Although the approaches

presented here differ from the one proposed in Miao and Gastwirth (2013), which compares the distributions of test scores of the passers from each group, similar conclusions are reached. The fourth case arose when an employer investigated some employees for misusing the e-mail system to forward material from a pornographic website. A higher fraction of minority employees were investigated and subsequently terminated by the employer.

4.1 Reanalysis of the data from Kirkland, et al. v. The New York State Civil Service Commission[5]

Edward Kirkland and other minority sergeants in the New York State Department of Correctional Services (DOCS) brought a class action in 1982, alleging that the exam for the position of Lieutenant carried out by DOCS and the New York Civil Service Commissions (CSC) resulted in an eligibility list that was discriminatory against blacks and Hispanics. The first stage analysis examines whether there is a statistically significant difference in the pass rates; while the second concerns the fairness of making appointments from the passers according to their rank on the scores. The data on both stages are given in Table 5. The two-sided Fisher's exact test applied to both the pass and promotion data yields p-values of .06 and 6.39×10^{-9}, respectively. This implies a highly significant difference in the promotion rates; however, the difference in the proportions of passing the exam just misses significance at the .05 level. If one's interest is in detecting unfairness in both processes, the first adaptive method chooses the mixture χ^2 test as the BD test is significant (p-value = .03 < .20). The mixture χ^2 test yields a p-value of $<10^{-6}$, indicating that the minorities were disadvantaged in both processes. When the focus is on the actual promotions, the second adaptive procedure is appropriate. The first stage test rejects the equality of the proportions of the two groups passing the exam since its p-value (.06) < .20, so a one-sided Fisher's exact test is used to assess whether the promotions were consistent with a random selection from the original pool of eligible candidates who took the exam. This yields a highly significant p-value of 7.48×10^{-10}, indicating a statistically significant disparity in the promotion process. Thus, both adaptive methods conclude that the protected group was significantly disadvantaged. Ultimately, the case was settled by an agreement that modified the rank-order selection procedure and assured that all passers would eventually be promoted.

4.2 Reanalysis of the data from United Black Firefighters Association v. Akron[6]

In 1988, the City of Akron developed a promotional examination for fire lieutenants. The promotional process has

[5] *Kirkland, et al. v. The New York State Civil Service Commission*, 711 F. 2d 1117 (2nd Cir., 1982).
[6] 1996 U.S. App. LEXIS 8011, 81 F. 3d 161 (6th Cir. 1996).

Table 6. Pass and Appointment Data from U.B.F v. Akron

Pass			
	+	−	Total
blacks	29	9	38
whites	89	10	99
Total	118	19	137

Appointment			
	+	−	Total
blacks	4	25	29
whites	41	48	89
Total	45	73	118

Table 7. Exam Pass and Promotion Data in Stewart v. Louis

Pass Exam				
	+	−	Total	Pass Rate
Minority	55	69	124	44.35%
Majority	119	64	183	65.03%
Total	174	133	307	56.68%

Promotion				
	+	−	Total	Pass Rate
Minority	4	51	55	7.27%
Majority	18	101	119	15.13%
Total	22	152	174	12.64%

four components: (1) written job knowledge test, (2) an "assessment center" portion consisting of the applicant's written analysis of a lieutenant's typical "in-basket" followed by an oral presentation; (3) service rating and (4) seniority. The last two factors were weighted less than the first two.

A total of 137 (38 blacks and 99 whites) took the exam in 1990. After the first two components, 118 candidates (29 blacks and 89 whites) remained. Then the service rating and seniority were incorporated and the top 45 candidates (4 blacks and 41 whites) were promoted. The data are presented in Table 6. When both stages are under scrutiny, applying Fisher's exact test to each stage separately yields two-sided p-values of .053 and .002. This indicates that blacks were statistically significant disadvantaged at the second stage, however, the difference between the pass rates on the exam is of "border-line" significance. Applying the first adaptive procedure, the BD test fails to reject the hypothesis of a common odds ratio (p-value = .38), so the common log odds ratio test is used to analyze the full dataset. It yields a p-value of 1.02×10^{-5} with a common odds ratio estimate (minority against majority) of .24. This implies a statistically significant and meaningful disparity in both the exam and appointment stages. When the second adaptive test is applied to the data, the p-value (.053) of Fisher's test of equality of passing rates is less than .20. Thus, the one-sided Fisher's exact test is used to compare the proportion of the original minority applicants who were actually appointed to the corresponding proportion of majority applicants. This yields a p-value of .0003, implying that blacks were statistically significantly disfavored in the appointment process. These conclusions differ from those of the legal decision, which said that the test did *not* have significant impact on the blacks. The court relied on a government guideline that there is no disparate impact since the ratio of the minority to majority pass rates (.85) is greater than the .80 or "four-fifths rule"[7]. Had the plaintiffs applied Fisher's exact test to the pass data, the two-sided p-value (.053) might have informed the court that the employment test in question deserved to be evaluated for job-relatedness.

4.3 Reanalysis of the data from Stewart, et al. v. City of St. Louis[8]

The city of St. Louis conducted promotion tests for Fire Captain in 2000, 2002 and 2004. Several minority fire fighters filed an EEOC charge of discrimination alleging that the promotion test for the City of St. Louis Fire Department for Fire Captain had an adverse impact on African-Americans. The data for the 2000 Captain exam is summarized in Table 7. At trial the defendant's expert stated that the promotion rates of those who passed should be compared in order to examine that the candidates of either race who were minimally qualified. He applied Fisher's exact test to the promotion data and concluded the difference in pass rate was not statistically significantly (p-value = .22). The plaintiffs' expert compared the African-American proportions (4/124 or 3.23%) of all applicants who were promoted to the majority proportion (18/183 or 9.84%). He found a statistically significant difference (p-value = .03) and noted that the ratio of the selection rates is 0.22, clearly less than four-fifths. The court decided that the plaintiffs' analysis was more relevant as the government guideline refers to selection procedures.

From the perspective that promotion depends on both stages, the first adaptive procedure is appropriate. The p-value (.97) of the BD test indicates that the common log odds ratio test should be used. It yields a p-value of 2.95×10^{-5} with an estimated common odds ratio of .43, indicating that there is a statistically significant difference in both the exam pass and promotion stages. Because the pass rate of minority applicants was statistically significant (p-value = .0004) lower than that of whites, the second adaptive procedure applies a one-sided Fisher's exact test to promotion data assuming the promotions are made from the original applicant pool. This yields a p-value of .02. Both adaptive procedures confirm the court's finding that the exam had a disparate impact. In the actual case, the city was able to validate the exam as job-related, so it prevailed.

---

[7]Uniform Guidelines, 29 C.F.R. 1607 (D), 2000.

[8]2007 U.S. Dist. LEXIS 38473, affirmed 532 F. 3d 939 (8th Cir., 2008).

Table 8. Investigation and Termination Data of Williams v. Wells Fargo

	Investigation				Termination		
	+	−	Total		+	−	Total
Minority	66	249	315	Minority	30	26	66
Majority	8	377	385	Majority	1	7	8
Total	74	626	700	Total	31	33	74

4.4 Reanalysis of the data from Williams v. Wells Fargo[9]

In November of 2005, as part of a routine quality assurance review of an employee's (Ms. B's) telephone and email activity, Wells Fargo noted that she had accessed an email and website called "orgasmic simulator". In the subsequent investigation, the company noticed that earlier Ms. B had received an email from another employee. Then the entire chain of senders and recipients of the "orgasmic simulator" email was found by looking in the "inbox", "deleted items" and "sent box" of the email accounts of employees. Ultimately, 74 Wells Fargo employees were investigated, of whom 66 were African-American. The investigation led to the termination of 31 employees, 30 of whom were African-American. The firing took place in three stages. After the first 16 employees were terminated in November 2005, a local newscast reported that most were minority. A senior executive saw the broadcast and suggested that a random sample of email records from all employees should be taken in order to ensure that the investigation was fair. That suggestion was *not* followed and two further sets of terminations were made on the basis of the original investigation: nine on December 2, 2005 and six on February 21, 2006. The plaintiffs challenged the fairness of both the initial review and the selection of employees who were terminated. At the time of the investigation, there were approximately 700 employees, 45% of whom were African-American. Of the employees investigated, 89% were African-American. The data are reported in Table 7. Applying the two-sided Fisher's exact test to the investigation and termination stages, separately, yields p-values of <.0001 and .054, respectively. This is very strong statistical evidence indicating that African-Americans were disadvantaged at the investigation stage. However, the p-value of the test applied to the termination data is .054, which is a border line significant result. Applying the first adaptive procedure to the investigation and termination data, the BD test fails to reject the equality of the two odds ratios (p-value = .71). Therefore, the common log odds ratio test is conducted. It yields a highly significant result (p-value = 2.08×10^{-27}) with an estimated common odds ratio of 10.61 (minority vs. majority), indicating that African-Americans had

[9] *Williams, et al. v. Wells Fargo Financial Acceptance Co.*, 564 F. Supp. 2nd 441 (E.D. PA, 2008).

ten times the odds of whites for being investigated and subsequently terminated. Because the minority group was disfavored in the first stage, the second adaptive procedure is also appropriate. A statistically significant result (p-value = 3.25×10^{-10}) was found, again providing highly significant evidence that minorities were terminated at a statistically higher rate than majority employees.

Consider a hypothetical scenario that Wells Fargo had terminated 2 white employees rather than just one along with 29 African-American employees. If the discriminatory investigation was ignored and the usual two-sided Fisher's exact test was applied to the termination data, a non-significant result (p-value = .26) would be found. In contrast when this data are jointly analyzed by the two adaptive methods, highly significant results (p-value < 10^{-10} and p-value = 6.72×10^{-9}, respectively) are found.

5. DISCUSSION & SUMMARY

In this paper, two adaptive procedures are developed for a commonly occurring nested two-stage problem in the legal setting.

One adaptive procedure is primarily focused on detecting possible unfairness in both. Following the work of Chernoff (1954), Self and Liang (1987), a mixture χ^2 test is obtained for our problem. The common log odds ratio test is motivated by the tendency for discrimination in one aspect of employer's policy to affect related policies. Then an adaptive method using the Breslow-Day test as the preliminary procedure to choose between the mixture χ^2 and common log odds ratio test is presented.

When the main focus is on the second stage, e.g. promotion, the commonly used Fisher's exact test may give a misleading result if the original hiring process was unfair. This occurs because a smaller than expected number of minority hires reduces the power of the Fisher's exact test to detect unfairness at the second stage (promotion). The second adaptive procedure is motivated by a court's suggestion. It checks the fairness of the hiring process before applying Fisher's exact test. When the hiring rates are not statistically significant (at level .20), a two-sided Fisher's exact test is applied to the promotion data. If there is statistical evidence suggesting possible unfairness at the hiring stage, a one-sided Fisher's exact test is used to check whether the promotions were consistent with a random sample of the original applicant pool.

A simulation study of the performance of the procedures shows that all of the six procedures preserve the desired .05 nominal level. The common log odds ratio test performs better than the mixture χ^2 test, when the true value of odds ratios of the two stages are close; while the mixture χ^2 test has higher power when the two odds ratios are substantially different or there is little evidence of unfairness

at the first stage. The power of the first adaptive procedure is always between the power of the mixture χ^2 and log odds ratio tests, and its power is closer to the more powerful of the two. Thus, the Breslow-Day adaptive test can be recommended, when both hiring and promotion are being examined, or when the promotions are being examined but the fairness of hiring is questionable. The simulation study found that Fisher's exact test has low power for detecting possible unfairness in promotion policy when there is a disparity in the hiring rates. It was found that the second adaptive method is as powerful as the standard Fisher's exact test when minorities received more than their expected number of hires. Data from the legal cases were reanalyzed with the adaptive procedures. Generally they provide stronger evidence of a disparity than Fisher's exact test. In *United Black Firefighters Association v. Akron* where the courts did not find that the promotion exam had a disparate impact, both adaptive procedures found a highly significant difference in the actual selection rates (p-value < .001).

The results demonstrate that the choice of a statistical test for a disparity in success rates at a second stage should depend on whether the employment practice at the previous stage was fair. This indicates that in order to properly study the fairness of an employer's promotion policy, one should check the fairness of the hiring process. An important aspect of the legal process is the discovery phase, in which each party can request relevant information from the other side. The finding that the power of the statistical test used to detect unfairness at the second (promotion, appointment) stage is noticeably reduced when minorities were disadvantaged at an earlier stage, e.g. hiring or appointments to the feeder positions, has an important implication for judges when they create the rules for discovery. In order to adequately assess the fairness of a second stage process, e.g. promotions, data on the employment practices from previous stages also need to be examined.

Further development of the methodology presented may be needed to cover specific real world situations. For example, in addition to evaluating existing employees for promotion, an employer may also consider external applicants. Then the log odds ratio of external candidates could be considered as a third stratum in the Breslow-Day test. Depending on the result, either a common log odds ratio test or a mixture χ^2 test could be used. Sometimes, several lower level jobs serve as "feeder" positions for a higher level one. The promotion log odds ratio (ϕ_p) can be estimated through a logistic regression incorporating relevant covariates including the position held. Then the estimated log odds ratio ($\hat{\phi}_p$) would be used with $\hat{\theta}_h$ in the first adaptive method described here. In summary, the approach presented here should provide a framework for studying these and other complications arising in real world applications.

APPENDIX A. APPENDIX SECTION

A.1 Detailed calculation for Theorem 1

The joint distribution of (U_1, U_2) can be calculated as follows

(6)
$$f_{U_1,U_2}(u_1,u_2) = f_{U_1}(u_1) \times f_{U_2|U_1}(u_2|u_1)$$

$$= \frac{\exp\left\{-\frac{u_1^2}{2\gamma(1-\gamma)\lambda_1(1-\lambda_1)}\right\}}{\sqrt{2\pi\gamma(1-\gamma)\lambda_1(1-\lambda_1)}} \times \frac{\exp\left\{-\frac{\left(u_2-\frac{\lambda_2}{\lambda_1}u_1\right)^2}{2\gamma(1-\gamma)\lambda_2\left(1-\frac{\lambda_2}{\lambda_1}\right)}\right\}}{\sqrt{2\pi\gamma(1-\gamma)\lambda_2\left(1-\frac{\lambda_2}{\lambda_1}\right)}}$$

$$= \frac{\exp\left\{-\frac{u_1^2}{2\gamma(1-\gamma)\lambda_1(1-\lambda_1)} - \frac{\left(u_2-\frac{\lambda_2}{\lambda_1}u_1\right)^2}{2\gamma(1-\gamma)\lambda_2(1-\lambda_2)\left(1-\frac{\lambda_2(1-\lambda_1)}{\lambda_1(1-\lambda_2)}\right)}\right\}}{2\pi\sqrt{\gamma(1-\gamma)\lambda_1(1-\lambda_1)}\sqrt{\gamma(1-\gamma)\lambda_2(1-\lambda_2)}\sqrt{1-\frac{\lambda_2(1-\lambda_1)}{\lambda_1(1-\lambda_2)}}}$$

$$= \frac{\exp\left\{-\frac{1}{2\left(1-\frac{\lambda_2(1-\lambda_1)}{\lambda_1(1-\lambda_2)}\right)}\left[\frac{u_1^2\left(1-\frac{\lambda_2(1-\lambda_1)}{\lambda_1(1-\lambda_2)}\right)}{\gamma(1-\gamma)\lambda_1(1-\lambda_1)} + \frac{\left(u_2-\frac{\lambda_2}{\lambda_1}u_2\right)^2}{\gamma(1-\gamma)\lambda_2(1-\lambda_2)}\right]\right\}}{2\pi\sqrt{\gamma(1-\gamma)\lambda_1(1-\lambda_1)}\sqrt{\gamma(1-\gamma)\lambda_2(1-\lambda_2)}\sqrt{1-\frac{\lambda_2(1-\lambda_1)}{\lambda_1(1-\lambda_2)}}}$$

$$= \frac{\exp\left\{-\frac{1}{2\left(1-\frac{\lambda_2(1-\lambda_1)}{\lambda_1(1-\lambda_2)}\right)}Q\right\}}{2\pi\sqrt{\gamma(1-\gamma)\lambda_1(1-\lambda_1)}\sqrt{\gamma(1-\gamma)\lambda_2(1-\lambda_2)}\sqrt{1-\frac{\lambda_2(1-\lambda_1)}{\lambda_1(1-\lambda_2)}}},$$

where

$$Q = \frac{u_1^2\left(1-\frac{\lambda_2(1-\lambda_1)}{\lambda_1(1-\lambda_2)}\right)}{\gamma(1-\gamma)\lambda_1(1-\lambda_1)} + \frac{\left(u_2-\frac{\lambda_2}{\lambda_1}u_1\right)^2}{\gamma(1-\gamma)\lambda_2(1-\lambda_2)}$$

$$= \frac{u_1^2}{\gamma(1-\gamma)\lambda_1(1-\lambda_1)} - \frac{2\sqrt{\frac{\lambda_2(1-\lambda_1)}{\lambda_1(1-\lambda_2)}}u_1u_2}{\gamma(1-\gamma)\sqrt{\lambda_1(1-\lambda_1)}\sqrt{\lambda_2(1-\lambda_2)}}$$

(7)
$$+ \frac{u_2^2}{\gamma(1-\gamma)\lambda_2(1-\lambda_2)}$$

Let $\sigma_i = \sqrt{\gamma(1-\gamma)\lambda_i(1-\lambda_i)}$ $(i = 1, 2)$ and $\rho = \sqrt{\frac{\lambda_2(1-\lambda_1)}{\lambda_1(1-\lambda_2)}}$, and replacing the terms in equation (7) with σ_i $(i = 1, 2)$ and ρ yields

(8)
$$Q = \frac{u_1^2}{\sigma_1^2} - \frac{2\rho u_1 u_2}{\sigma_1 \sigma_2} + \frac{u_2^2}{\sigma_2^2}.$$

Substituting the terms in (6) with σ_i $(i = 1, 2)$, ρ and Q in (6) with (8) yields

$$f_{U_1,U_2}(u_1,u_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$$
$$\times \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{u_1^2}{\sigma_1^2} - \frac{2\rho u_1 u_2}{\sigma_1 \sigma_2} + \frac{u_2^2}{\sigma_2^2}\right]\right\}.$$

Therefore

$$(U_{1n}, U_{2n})' \overset{\mathcal{L}}{\to} (U_1, U_2) = N\left((0,0)', \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}\right).$$

Thus, we have

$$\sqrt{n}\left(\frac{X/n - \mu_1}{\sigma_1}, \frac{Y/n - \mu_2}{\sigma_2}\right)' \to N\left((0,0)', \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right),$$

where $\mu_i = \gamma\lambda_i$, $\sigma_i = \sqrt{\gamma(1-\gamma)\lambda_i(1-\lambda_i)}$, $(i = 1, 2)$ and $\rho = \sqrt{\frac{\lambda_2(1-\lambda_1)}{\lambda_1(1-\lambda_2)}}$.

Similar calculation can be carried out for Theorem 2.

A.2 Proof of Corollary 1

Proof. From Theorem 1, one obtains

$$\sqrt{n}(X/n - \mu_1, Y/n - \mu_2)' \xrightarrow{\mathcal{L}} N\left((0,0)', \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}\right),$$

$$\hat{\theta}_h \sim \frac{X/(n\gamma - X)}{(n\lambda_1 - X)(n(1-\gamma-\lambda_1) + X)}$$

$$= \log \frac{\frac{X}{n}/\left(\gamma - \frac{X}{n}\right)}{\left(\lambda_1 - \frac{X}{n}\right)/\left[(1-\gamma-\lambda_1) + \frac{X}{n}\right]},$$

$$\hat{\theta}_p \sim \log \frac{Y/(X-Y)}{(n\lambda_2 - Y)/[n(\lambda_1-\lambda_2) - X + Y]}$$

$$= \log \frac{\frac{Y}{n}/\left(\frac{X}{n} - \frac{Y}{n}\right)}{\left(\lambda_2 - \frac{Y}{n}\right)/\left[(\lambda_1-\lambda_2) - \frac{X}{n} + \frac{Y}{n}\right]},$$

i.e., $(\hat{\theta}_h, \hat{\theta}_p) \sim \left(g_1\left(\frac{X}{n}, \frac{Y}{n}\right), g_2\left(\frac{X}{n}, \frac{Y}{n}\right)\right)$, where $g_1(t_1, t_2) = \log \frac{t_1/(\gamma - t_1)}{(\lambda_1 - t_1)/[(1-\gamma-\lambda_1)+t_1]}$ and $g_2(t_1, t_2) = \log \frac{t_2/(t_1 - t_2)}{(\lambda_2 - t_2)/[(\lambda_1 - \lambda_2) - t_1 + t_2]}$. Note that $\hat{\theta}_h|_{\frac{X}{n} = \mu_1} \sim 0$ and $\hat{\theta}_p|_{\frac{X}{n} = \mu_1, \frac{Y}{n} = \mu_2} \sim 0$, then applying the delta method yields

$$\sqrt{n}(\hat{\theta}_h, \hat{\theta}_p)' \xrightarrow{H_0} N\left((0,0)', \Gamma\begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}\Gamma'\right),$$

where

$$\Gamma = \begin{pmatrix} \frac{\partial\hat{\theta}_h}{\partial(X/n)} & \frac{\partial\hat{\theta}_h}{\partial(Y/n)} \\ \frac{\partial\hat{\theta}_p}{\partial(X/n)} & \frac{\partial\hat{\theta}_p}{\partial(Y/n)} \end{pmatrix}\Big|_{\frac{X}{n} = \mu_1, \frac{Y}{n} = \mu_2}$$

$$= \begin{pmatrix} \frac{1}{\gamma(1-\gamma)\lambda_1(1-\lambda_1)} & 0 \\ -\frac{1}{\gamma(1-\gamma)(\lambda_1-\lambda_2)} & \frac{\lambda_1}{\gamma(1-\gamma)\lambda_2(\lambda_1-\lambda_2)} \end{pmatrix}.$$

$$\Rightarrow \Gamma\begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}\Gamma'$$

$$= \Gamma\begin{pmatrix} \gamma(1-\gamma)\lambda_1(1-\lambda_1) & \gamma(1-\gamma)\lambda_2(1-\lambda_1) \\ \gamma(1-\gamma)\lambda_2(1-\lambda_1) & \gamma(1-\gamma)\lambda_2(1-\lambda_2) \end{pmatrix}\Gamma'$$

$$= \begin{pmatrix} \frac{1}{\gamma(1-\gamma)\lambda_1(1-\lambda_1)} & 0 \\ 0 & \frac{1}{\gamma(1-\gamma)\lambda_2\left(1-\frac{\lambda_2}{\lambda_1}\right)} \end{pmatrix}.$$

Therefore

$$\sqrt{n}(\hat{\theta}_h, \hat{\theta}_p)' \xrightarrow{H_0} N\left((0,0)', \begin{pmatrix} \frac{1}{\gamma(1-\gamma)\lambda_1(1-\lambda_1)} & 0 \\ 0 & \frac{1}{\gamma(1-\gamma)\lambda_2\left(1-\frac{\lambda_2}{\lambda_1}\right)} \end{pmatrix}\right).$$

$\square$

A.3 Proof of Corollary 2

Proof. From Theorem 2, one obtains

$$\sqrt{m}(X/m - \mu_1, Y/m - \mu_2)' \xrightarrow{\mathcal{L}} N\left((0,0)', \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}\right).$$

Also

$$\hat{\theta}_h = \log \frac{X(1-\gamma)}{(m-X)\gamma} = \log \frac{\frac{X}{m}(1-\gamma)}{\left(1 - \frac{X}{n}\right)\gamma},$$

$$\hat{\theta}_p \sim \log \frac{Y/(X-Y)}{(m\lambda - Y)/(m(1-\lambda) - X + Y)}$$

$$= \log \frac{\frac{Y}{m}/\left(\frac{X}{m} - \frac{Y}{m}\right)}{\left(\lambda - \frac{Y}{m}\right)/\left[(1-\lambda) - \frac{X}{m} + \frac{Y}{m}\right]},$$

i.e., $(\hat{\theta}_h, \hat{\theta}_p)$ is asymptotically equivalent to $\left(g_1\left(\frac{X}{m}, \frac{Y}{m}\right), g_2\left(\frac{X}{m}, \frac{Y}{m}\right)\right)$, where $g_1(t_1, t_2) = \log \frac{t_1(1-\gamma)}{(1-t_1)\gamma}$ and $g_2(t_1, t_2) = \log \frac{t_2/(t_1-t_2)}{(\lambda-t_2)/[(1-\lambda)-t_1+t_2]}$. Note that $\hat{\theta}_h|_{\frac{X}{m} = \mu_1} = 0$ and $\hat{\theta}_p|_{\frac{X}{m} = \mu_1, \frac{Y}{m} = \mu_2} \sim 0$, then applying the delta method, one obtains $\sqrt{m}(\hat{\theta}_h, \hat{\theta}_p)' \xrightarrow{H_0} N\left((0,0)', \Gamma\begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}\Gamma'\right)$, where

$$\Gamma = \begin{pmatrix} \frac{\partial\hat{\theta}_h}{\partial\frac{X}{m}} & \frac{\partial\hat{\theta}_h}{\partial\frac{Y}{m}} \\ \frac{\partial\hat{\theta}_p}{\partial\frac{X}{m}} & \frac{\partial\hat{\theta}_p}{\partial\frac{Y}{m}} \end{pmatrix}\Big|_{\frac{X}{m} = \mu_1, \frac{Y}{m} = \mu_2}$$

$$= \begin{pmatrix} \frac{1}{\gamma(1-\gamma)\lambda_1(1-\lambda_1)} & 0 \\ -\frac{1}{\gamma(1-\gamma)(\lambda_1-\lambda_2)} & \frac{\lambda_1}{\gamma(1-\gamma)\lambda_2(\lambda_1-\lambda_2)} \end{pmatrix}.$$

$$\Rightarrow \Gamma\begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}\Gamma' = \Gamma\begin{pmatrix} \gamma(1-\gamma) & \lambda\gamma(1-\gamma) \\ \lambda\gamma(1-\gamma) & \lambda\gamma(1-\gamma) \end{pmatrix}\Gamma'$$

$$= \begin{pmatrix} \frac{1}{\gamma(1-\gamma)} & 0 \\ 0 & \frac{1}{\lambda(1-\lambda)\gamma(1-\gamma)} \end{pmatrix}.$$

Therefore

$$\sqrt{m}(\hat{\theta}_h, \hat{\theta}_p)' \xrightarrow{H_0} N\left((0,0)', \begin{pmatrix} \frac{1}{\gamma(1-\gamma)} & 0 \\ 0 & \frac{1}{\lambda(1-\lambda)\gamma(1-\gamma)} \end{pmatrix}\right).$$

$\square$

A.4 Size of Breslow-Day test for the 5% and 20% test

Test	Odds Ratio							
	$\frac{5}{4}$	$\frac{6}{5}$	1	$\frac{5}{6}$	$\frac{4}{5}$	$\frac{3}{4}$	$\frac{2}{3}$	$\frac{1}{2}$
(setup 1)								
χ^2	.050	.048	.049	.049	.048	.049	.050	.045
χ^2 Emp	.049	.048	.049	.051	.047	.051	.046	.042
(setup 2)								
χ^2	.050	.047	.044	.049	.051	.047	.051	.051
χ^2 Emp	.049	.047	.044	0.049	.051	.047	.051	.050
(setup 3)								
χ^2	.206	.211	.214	.209	.205	.204	.199	.202
χ^2 Emp	.206	.204	.218	.216	.207	.201	.201	.202
(setup 4)								
χ^2	.216	.215	.219	.224	.220	.215	.207	.197
χ^2 Emp	.191	.215	.217	.220	.220	.189	.198	.189

χ^2 Emp: size of the BD test based on the empirical statistic values under the null;
setup 1: $N_1 = 600$, $N_2 = 200$, $h = 400$, $p = 200$, .05 leve test;
setup 2: $N_1 = 300$, $N_2 = 100$, $h = 200$, $p = 100$, .05 level test;
setup 1: $N_1 = 600$, $N_2 = 200$, $h = 400$, $p = 200$, .20 level test;
setup 2: $N_1 = 300$, $N_2 = 100$, $h = 200$, $p = 100$, .20 level test;

REFERENCES

BANCROFT, T. A. (1964). Analysis and inference for incompletely specified models involving the use of preliminary test(s) of significance. *Biometrics*, **20**, 427–442. MR0181066

BRESLOW, N. E. and DAY, N. E. (1980). *Statistical Methods in Cancer Research: Vol. 1 – The Analysis of Case-Control Studies.* Lyon, France, IARC Scientific Publications.

CHERNOFF, H. (1954). On the distribution of the likelihood ratio. *Annals of Mathematical Statistics*, **25**, 573–578. MR0065087

DORSANEO, W. V. (1975). Statistical evidence in employment discrimination litigation: Selection of the available population, problems and proposals. *Southwestern Law Journal*, **29**, 859–875.

FELLER, W. (1968). *Intoduction to Probability Theory and Its Applications.* John Wiley and Sons, Third Edition. MR0228020

FINKELSTEIN, M. O. and LEVIN, B. (2001). *Statistics for Lawyers.* Springer, Second Edition.

FLEMING, T. R. and HARRINGTON, D. P. (1991). *Counting Processes and Survival Analysis.* John Wiley and Sons, New York. MR1100924

FLIGNER, M. A. and HETTMANSPERGER, T. P. (1979). On the use of conditional asymptotic normality. *J. Roy. Statistical Society Ser. B*, **41**, 178–183. MR0547242

FLIGNER, M. A. (1981). A note on limit theorems for joint distributions with applications to linear signed rank statistics. *J. Roy. Statistical Society Ser. B*, **43**, 61–64. MR0610378

GASTWIRTH, J. L. (1981). Demographic mix of the available labor force. *Monthly Labor Review*, **104**, 50–57.

GASTWIRTH, J. L. (1997). Statistical evidence in discrimination cases. *Journal of the Royal Statistical Society Ser. A*, **160**, 289–303.

GASTWIRTH, J. L. and HABER, S. (1976). Defining the labor market for equal employment standards. *Monthly Labor Review*, **99**, 32–36.

GASTWIRTH, J. L. and GREENHOUSE, S. W. (1987). Estimating a common relative risk: Application in equal employment. *Journal of the American Statistical Association*, **82**, 38–45. MR0883335

GASTWIRTH, J. L. and GREENHOUSE, S. W. (1995). Biostatistical concepts and methods in the legal setting. *Statistics in Medicine*, **14**, 1641–1653.

HABER, S. and GASTWIRTH, J. L. (1978). Specifying the labor market for individual firms. *Monthly Labor Review*, **101**, 26–30.

HOGG, R. (1974). Adaptive robust procedures: A partial review and some suggestions for future applications and theory. *Journal of the American Statistical Association*, **69348**, 909–923. MR0461779

HUBER, P. J. (1972). The 1972 wald lecture, robust statistics: A review. *Annals of Mathematical Statistics*, **43**, 1041–67. MR0314180

MANTEL, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemotherapy Reports*, **50**(3), 163–170.

MIAO, W. and GASTWIRTH, J. L. (2009). A new two stage adaptive nonparametric test for paired differences. *Statistics and Its Inference*, **2**, 213–221. MR2516072

MIAO, W. and GASTWIRTH, J. L. (2013). Properties of statistical tests appropriate for the analysis of data in disparate impact cases (to appear in *Law, Probability and Risk*).

NEUHAUSER, M. (2001). An adaptive location-scale test. *Biometrical J.*, **43**, 809–819. MR1869487

O'GORMAN, T. W. (2004). *Applied Adaptive Statistical Methods: Tests of Significance and Confidence Intervals.* Cambridge University Press, Cambridge, UK. MR2023355

PINSKY, M. A. (2008). http://www.dartmouth.edu/~chance/teaching_aids/books_articles/probability_book/pinsky-hypergeometric.pdf.

ROSENBLUM, M. (1977). The use of labor statistics and analysis in title VII cases: *Rois*, *Chicago* and beyond. *Industrial Relations Law Journal*, **1**, 685–710.

ROSENBLUM, M. (1978). The external measures of labor supply: Recent issues and trends. *Connecticut Law Review*, **10**, 892–919.

SELF, S. G. and LIANG, K. Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, **82**, 605–611. MR0898365

SETHURAMAN, J. (1961). Some limit theorems for joint distributions. *Sankhyā Ser. A*, **23**, 379–386. MR0139190

Wenjing Xu
Department of Statistics
George Washington University, Washington, D.C., 20052
USA
E-mail address: jeanvane@gwu.edu

Qing Pan
Department of Statistics
George Washington University, Washington, D.C., 20052
USA
E-mail address: qpan@gwu.edu

Joseph L. Gastwirth
Department of Statistics
George Washington University, Washington, D.C., 20052
USA
E-mail address: jlgast@gwu.edu