

## How accurate are the power calculations relied on by the SEC in its regulatory deliberations?

EFSTATHIA BURA<sup>†</sup>

*Department of Statistics, The George Washington University, Washington, DC 20052, USA and Biometrics, Vertex Pharmaceuticals, Cambridge, MA 02139, USA*

AND

JOSEPH L. GASTWIRTH<sup>‡</sup>

*Department of Statistics, The George Washington University, Washington, DC 20052, USA*

In two related decisions in *Chamber of Commerce of the United States of America v. Securities and Exchange Commission (SEC)*, the District of Columbia Federal Court of Appeals ruled that the SEC had not fully complied with some provisions of the Administrative Procedures Act when it required that the boards of investment companies managing mutual funds have at least 75% of their membership and the Chairman be independent directors. In preparation for renewed rule making, the Office of Economic Analysis of the SEC prepared a *Power Study* to respond to an industry-sponsored report claiming that the returns of funds with independent boards and chairs are not superior to funds with boards dominated by management. The Power Study concluded that the available studies on the effectiveness of independent board members do not have sufficient statistical power to detect a meaningful difference in the returns of the two types of funds. This paper demonstrates that the method used by the SEC in their power calculation is not correct, unless a very restrictive condition that rarely occurs in practice holds. When the appropriate power formulas are used, the expected power of studies of the same size as the ones examined by the SEC is actually *lower* than the corresponding results of the SEC. Thus, the results in the paper actually strengthen the argument that the SEC is advocating. The relevance of both the SEC and the industry studies to the main issue in the case is also questioned in the discussion.

*Keywords:* comparing two groups; hypothesis tests; indicator variables; regression; statistical power.

### 1. Introduction

Early in 2004 the Securities and Exchange Commission (SEC) proposed new conditions that Mutual Funds needed to satisfy in order to engage in otherwise prohibited transactions, i.e. those that the fund's advisor would also benefit.<sup>1</sup> In particular, the fund would be required to have a Chairman and at least 75% of its Directors independent of its advisors and sponsors.<sup>2</sup> The Commission adopted this and other new conditions in light of recent abuses in the mutual fund industry and felt that more

<sup>†</sup> Email: eburagwu@eduw.edu

<sup>‡</sup> Email: jlgast@eduw.edu

<sup>1</sup> See Investment Company Governance, Final Rule, 69 Fed. Reg. 46,378 (2 August 2004).

<sup>2</sup> See Role of Independent Directors of Investment Companies, Final Rule, 66 Fed. Reg. 3734 (16 January 2001) for the criteria a director needs to satisfy in order to be considered independent.

independent directors would provide greater oversight of activities that involve inherent conflicts of interest between the funds and their managers in order to protect shareholders. The industry challenged the new proposal. In *Chamber of Commerce of the United States of America v. Securities and Exchange Commission*,<sup>3</sup> the Court of Appeals of the D.C. Circuit held that while the Commission had the legal authority to establish additional conditions on the funds, its deliberations did not fully comply with the Administrative Procedures Act. In particular, the Commission had not considered the potential costs of more independent directors versus those of a proposed alternative involving fuller disclosure by investment companies. Within a matter of days after this decision was issued, the SEC published a Response Release<sup>4</sup> stating that it decided that it was unnecessary to reopen the rule-making record for comment as the information in the existing record together with other publicly available information that could be relied on was a sufficient base to remedy the deficiencies noted by the court. The Chamber of Commerce again challenged this procedure in part because when an agency determines that additional fact gathering is needed, public notice and comments are typically required. In *Chamber of Commerce of the United States of America v. Securities and Exchange Commission II*,<sup>5</sup> the court vacated the rule and gave the SEC time to reopen the record. The decision noted that the SEC estimated the cost of requiring an independent chair on the increased compensation for the chair and the cost of extra staff, the latter of which relied on 'extra record' summaries of two surveys, one of which was not even available during the previous comment period.<sup>6</sup>

In the first case, the Chamber referred to a study (Bobroff & Mack, 2004) supported by Fidelity Investments that had not found a statistically significant difference in the performance of funds with a high proportion of independent directors compared with other mutual funds. The Commission discounted that study because it had not incorporated other potential predictors.<sup>7</sup> The study encompassed all retail-oriented fund complexes with \$10 billion or more in long-term assets, which account for 83% of industry long-term fund assets. The data included in the study indicated that the distributions of assets managed by the two types of funds are different. For example, the asset mean for the management-chaired funds was 74,180 (\$ millions) with a standard deviation of 126,519 (\$ millions), while the asset mean for the independent-chaired funds was 38,490 (\$ millions) with a standard deviation of 35,091.89 (\$ millions).<sup>8</sup> Somewhat surprisingly, the SEC did not suggest the possible economies of scale fund groups managing more assets might have or any other financial or economic variables that should have been included in the study of differential returns of the two types of funds.

In anticipation of the further proceedings mandated by the second appellate decision, the SEC staff carried out a literature review (Spatt, 2006a) and a power study (Spatt, 2006b). The purpose of the power study, also referred to as the Office of Economic Analysis (OEA)/SEC memorandum, was to demonstrate that statistical tests based on regression models that include relevant predictors would not have adequate power to detect a meaningful impact of independent directors and chairs on the profitability of funds. The power study of the SEC is described in Section 2. It will be seen that the

<sup>3</sup> 412 F.3d 133 (D.C. Cir. 2005).

<sup>4</sup> See Investment Company Governance, Release No. 26,985, 70 Fed. Reg. 39390-39398 (7 July 2005).

<sup>5</sup> 443 F.3d 890 (D.C. Cir. 2006).

<sup>6</sup> Ibid at 907.

<sup>7</sup> See n. 5 at 143 where the opinion notes that the study did not rule out 'other important differences [than independence of the chairman] that may have impacted performance results,' 69 Fed.Reg. at 46,383 n. 52 (quoting study), and because it did not use a reliable method of calculating fund expenses.

<sup>8</sup> See Exhibit 4 in Bobroff & Mack (2004).

approach the SEC adopted is statistically correct only when the predictors have the same distribution in both types of funds. To illustrate the flaw in the logic of the OEA/SEC memorandum, we consider the simple case of a regression model with a single predictor in Section 3 and demonstrate that when the distribution of the predictor in the funds with independent directors is different than the corresponding distribution in the other funds, the power results given in the OEA/SEC memorandum are not correct. The correct power formula is also given in that section and its mathematical derivation is given in Appendix A.

Section 3 presents simulation studies illustrating the impact of the distributions of the predictor variable in the two groups of funds. Interestingly, using the appropriate formula, the power of a regression analysis to detect a 3% difference in returns, deemed meaningful by the SEC, is *lower* than the figures given in their memorandum. Thus, a correct analysis would strengthen the argument presented by the SEC staff.

Section 4 summarizes the implications of our findings and raises an additional issue about the universe of investments examined by the studies submitted to the court. In contrast to the criteria used to focus on comparisons of ‘similarly situated’ employees or applicants in discrimination cases, the studies of returns of investment companies included all of their investments and were not focused on the smaller number of securities transactions that involve a potential conflict of interest.

## 2. The SEC power study

The SEC staff examined the performance of independent- versus management-chaired mutual funds by considering the asset-weighted average annual returns for every family in the Center for Research in Securities Prices (CRSP) mutual fund database.<sup>9</sup> Using this data set, they computed the cross-sectional standard deviation for 2002 to be approximately 12%. In the memo, the SEC staff stated the following hypothesis testing problem:

$H_0$ : *Mutual funds with an independent chair produce returns no higher than those with a management chair.*

*versus*

$H_1$ : *Mutual funds with an independent chair produce higher returns than those with a management chair.*

The OEA/SEC memorandum presents power curves as a function of sample size (up to 500).<sup>10</sup> The analysis assumed that one in five fund families is chaired by an independent director or trustee. The sample size on the  $x$ -axis is the total number of funds in the two groups with a 1:4 allocation between the independent- and management-chaired funds. The power of the two-sample  $t$ -test to detect differences ranging from 1% to 5% (100–500 basis points) at level 5% is computed. They indicated that the power of a test, applied to a sample of 448 funds in total, to detect a 3% difference between independent-chaired and management-chaired mutual funds is approximately 55%. Figure 1a,b gives the power curves of both the one-sided and the two-sided  $t$ -test assuming the groups have the same standard deviation, 12%, used by the SEC. Figure 1b replicates the SEC graph indicating that they reported the power curve for a two-sided test even though the stated alternative appears to be one-sided. The left vertical line is drawn at 284.73 and the right at 443.59, as in the OEA/SEC memorandum, to facilitate comparison. Thus, when a two-sided  $t$ -test is used, a sample

<sup>9</sup> This is the CRSP Survivor-Free U.S. Mutual Fund Database produced by the Center for Research in Securities Prices at the University of Chicago Graduate School of Business.

<sup>10</sup> Figure 1; p. 9 of the OEA/SEC Memorandum (Spatt, 2006b).

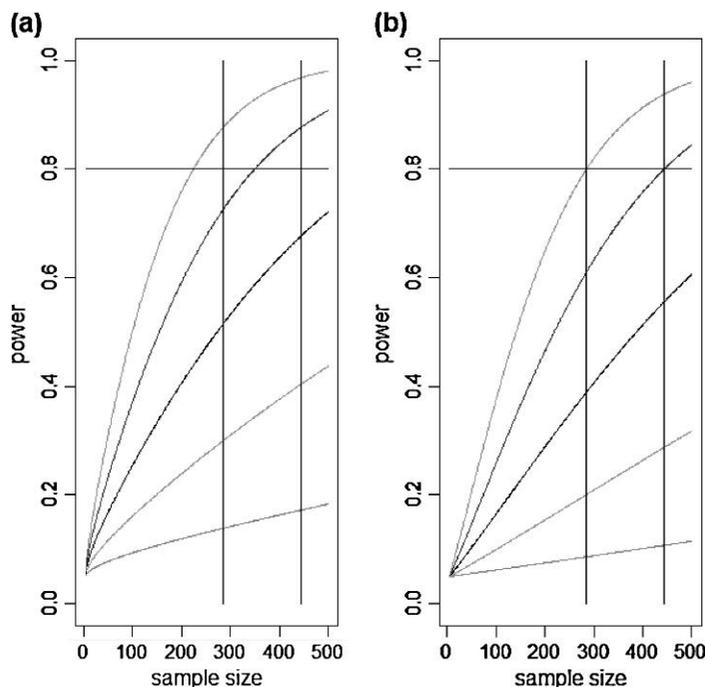


FIG. 1. Power curves for the two-sample  $t$ -test to detect differences of 1% (red), 2% (green), 3% (black), 4% (blue) and 5% (purple) at level 5% assuming the groups have the same standard deviation, 12%, used by the SEC. (a) Power curves for the one-sided two-sample  $t$ -test. (b) Power curves for the two-sided two-sample  $t$ -test, which replicate those of the OEA/SEC memorandum.<sup>14</sup>

size of about 285 is needed in order for the test to have 80% power to detect a 5% difference in mean returns. To detect a 4% difference with 80% power, a sample of size 444 is required.

Noting that a careful study of the difference in performance of the two groups of funds would control for other variables that affect returns, the SEC stated that a regression model incorporating the relevant predictors is appropriate. In order to calculate the power of a regression model, with a coefficient of determination  $R^2$ , the SEC multiplied the original variance of the returns in each group by  $(1 - R^2)$  and then used the power formula for the two-sample  $t$ -test assuming that the variance in returns in both groups equaled this smaller variance.<sup>11</sup> In particular, when the standard deviation of returns equals 12 and  $R^2 = 0.50$  (0.75) the resulting standard deviation of the error term in the regression becomes 8.5 (6).<sup>12</sup> Their power curves as a function of both sample size and magnitude of the difference in returns of the two groups<sup>13</sup> will be compared to our results in Section 3.

<sup>11</sup> The  $R^2$  statistic measures the proportion of the variance in the response (returns) that is explained by the predictors used in the regression.

<sup>12</sup> The total variation in returns is the sum of the variation due to the predictors plus that due to the error. Since the total standard deviation assumed by the SEC equals 12, the error variance changes with  $R^2$ . The numbers in the text are obtained from formula (7). For example,  $\sigma_e^2 = \text{Var}(y)(1 - R^2) = 12^2(1 - 0.50) = 72$  and the error standard deviation is  $\sqrt{72} = 8.485 \approx 8.5$ .

<sup>13</sup> See Figs 2 and 3 on p. 11 of the OEA/SEC memorandum.

<sup>14</sup> The power curves determined by the OEA/SEC are given in Fig. 1 on p. 9 of their memorandum (Spatt, 2006b).

*Comments:*

1. The OEA/SEC memorandum did not describe the particular explanatory variables nor did it present an explicit regression model relating a fund's returns and the predictors. It only used the overall explanatory ability of the model as measured by  $R^2$  to carry out their calculations based on the sample standard deviation (12%) of the funds' returns.
2. It is known that the joint distribution of the predictors affects the power of a regression and the sample size determination for future regression analysis (Kleinbaum *et al.*, 2008; Hsieh *et al.*, 1998).

**3. Power in linear regression: one predictor and one indicator variable**

Assume that a regression model is used to predict a response  $y$  from the information on a single independent variable  $x$  and membership in one of two groups. In the context of the SEC memorandum, the response ( $y$ ) is fund return and the two groups are the independent-chaired and management-chaired mutual funds. A reasonable independent variable might be the amount of assets ( $x$ ) a fund manages. Formally, the model is

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \beta_2 I_{ij} + \varepsilon_{ij}, \quad (1)$$

$$i = 1, 0, j = 1, \dots, n_i,$$

$I_{ij} = 1$  if the  $j$ th fund is independent ( $i = 1$ ), 0 otherwise,  $\varepsilon_{ij}$  iid  $N(0, \sigma_\varepsilon)$ ,  $x_{ij} \sim N(\mu_i, \sigma_i)$ .

The reduced model describes the situation where group membership does not affect  $y$  once the effect of  $x$  is accounted for and is modelled as

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \varepsilon_{ij}, \quad (2)$$

$$i = 1, 2, j = 1, \dots, n_i, \varepsilon_{ij} \text{ iid } N(0, \sigma_\varepsilon), x_{ij} \sim N(\mu_i, \sigma_i).$$

The reduced model (2) is obtained by setting  $\beta_2 = 0$  in the full model (1). The hypothesis stated in the OEA/SEC memorandum, that independent-chaired funds realize higher returns, is mathematically expressed as  $H_0: \beta_2 = 0$  versus  $H_1: \beta_2 > 0$ . The least squares estimate  $\hat{\beta}_2$  of  $\beta_2$  in the full model and its variance are given in Beals (1972, pp. 269–273):

$$\hat{\beta}_2 = \frac{\sum x_{ij}^2 \sum y_{ij} I_{ij} - \sum x_{ij} I_{ij} \sum y_{ij} x_{ij}}{\sum x_{ij}^2 \sum I_{ij}^2 - (\sum x_{ij} I_{ij})^2}, \quad (3)$$

$$\text{Var}(\hat{\beta}_2) = \frac{\sigma_\varepsilon^2 \sum x_{ij}^2}{\sum x_{ij}^2 \sum I_{ij}^2 - (\sum x_{ij} I_{ij})^2}. \quad (4)$$

When the errors are independent, normally distributed with a common variance,  $\hat{\beta}_2$  is normally distributed so that the distribution of the test statistic  $t = \hat{\beta}_2 / s(\hat{\beta}_2)$  has a Student's  $t$  distribution with  $n - 3$  degrees of freedom under the null, where  $s(\hat{\beta}_2) = \sqrt{\text{Var}(\hat{\beta}_2)}$ . The power for an  $\alpha$ -level

one-sided  $t$ -test of  $H_0: \beta_2 = 0$  to detect a true difference of  $\beta_2$  is given by (Kapadia *et al.*, 2005, pp. 529–531)

$$P\left(t(n-3) > t_\alpha(n-3) - \frac{\beta_2}{s(\hat{\beta}_2)}\right). \quad (5)$$

In the context of the regression model in (1),  $\beta_2$  quantifies the difference in performance as measured by the response  $y$  between the two groups. The probability in (5) assesses the likelihood of the test correctly deducing that the true  $\beta_2$  is not zero. To be meaningful in our context, a test needs sufficiently high power to identify a modest 2 or 3% difference in the returns of the two fund groups.

To be consistent with the OEA/SEC power study (Spatt, 2006b), we also assume that the variance of  $y$  is equal to 144 and consider the most realistic value of  $R^2$ , 0.50, used by the SEC. Given these conditions, we have to calculate the error and covariate variances in the two groups that are consistent with these values.

In (2), the indicator  $I$  follows a simple binomial or Bernoulli distribution with parameter  $\pi$ . When  $I = 1$ ,  $x$  has mean  $\mu_1$  and standard deviation  $\sigma_1$  and when  $I = 0$ ,  $x$  has mean  $\mu_0$  and standard deviation  $\sigma_0$ . In Appendix A, the formula for the variance of a fund's returns is derived:

$$\begin{aligned} \text{Var}(y) = & \beta_1^2(\sigma_1^2\pi + \sigma_0^2(1-\pi) + \pi(1-\pi)(\mu_1 - \mu_0)^2) + \beta_2^2\pi(1-\pi) \\ & + 2\beta_1\beta_2\pi(1-\pi)(\mu_1 - \mu_0) + \sigma_\varepsilon^2. \end{aligned} \quad (6)$$

In general,

$$R^2 = 1 - \frac{\sigma_\varepsilon^2}{\text{Var}(y)} \quad (7)$$

which yields

$$\sigma_\varepsilon^2 = \text{Var}(y)(1 - R^2).$$

When  $\sigma_1^2 = \sigma_0^2 = \sigma^2$  and the predictor has the same mean (average) in both groups, i.e.  $\mu_1 = \mu_0$ , formula (6) for  $\text{Var}(y)$  reduces to

$$\beta_1^2\sigma^2 + \beta_2^2\pi(1-\pi) + \sigma_\varepsilon^2.$$

In this *special* situation where  $\mu_1 = \mu_0$  and  $\sigma_1^2 = \sigma_0^2 = \sigma^2$ ,  $R^2$  is the fraction of the variance of  $y$  due to the effect of both the group and the predictor. The remaining portion,  $1 - R^2$ , of the variance of  $y$  is the error variance  $\sigma_\varepsilon^2$ . Consequently, the SEC approach is correct in this special case.

To examine the more general situation when  $\mu_1$  need not equal  $\mu_0$ , we follow the SEC and assume that the regression has  $R^2 = 0.50$  and  $\text{Var}(y) = 144$ . Thus,  $\sigma_\varepsilon^2 = \text{Var}(y)/2 = 144/2 = 72$ . Then, (6) implies that for any set of coefficients  $(\beta_1, \beta_2)$  the group variances must satisfy

$$\beta_1^2(\sigma_1^2\pi + \sigma_0^2(1-\pi) + \pi(1-\pi)(\mu_1 - \mu_0)^2) + \beta_2^2\pi(1-\pi) + 2\beta_1\beta_2\pi(1-\pi)(\mu_1 - \mu_0) = 72.$$

Thus,

$$\beta_1^2(\sigma_1^2\pi + \sigma_0^2(1-\pi) + \pi(1-\pi)(\mu_1 - \mu_0)^2) = 72 - \beta_2^2\pi(1-\pi) - 2\beta_1\beta_2\pi(1-\pi)(\mu_1 - \mu_0)$$

or

$$\sigma_1^2\pi + \sigma_0^2(1-\pi) = [72 - \beta_2^2\pi(1-\pi) - 2\beta_1\beta_2\pi(1-\pi)(\mu_1 - \mu_0) - \beta_1^2\pi(1-\pi)(\mu_1 - \mu_0)^2]/\beta_1^2.$$

When  $\sigma_1^2 = \sigma_0^2 = \sigma^2$ , then the common variance  $\sigma^2$  equals the right-hand side of the last equality. In our study, we set  $\beta_0 = 5$  and  $\beta_1 = 2$  and assume that the group variances are equal. The power for testing  $\beta_2 > 0$  is computed for sample sizes ranging from  $n = 4$  to  $n = 500$  and for  $\beta_2 = 1, 2, 3, 4, 5$ . For the funds with independent chairs, the predictor ( $x$ ) is assumed to follow a normal distribution with mean  $\mu_1$  and standard deviation  $\sigma$  ( $x \sim N(\mu_1, \sigma)$ ), while  $x \sim N(\mu_0, \sigma)$  in funds without independent chairs. As in the OEC/SEC power study, 20% ( $n/5$ ) of the funds have independent chairs and the significance level is set to  $\alpha = 0.05$ . We investigate how the power of the test is affected by the difference in the distributions of the predictor in the two groups by considering several scenarios that depend on the magnitude of the difference  $\mu_1 - \mu_0$  between the average values of the predictor ( $x$ ) in the two groups.

In the OEA/SEC memorandum, approximate power of 0.80 is achieved for a difference of 5% ( $\beta_2 = 5$ ) at  $n = 143.86$  and for a difference of 4% ( $\beta_2 = 4$ ) at  $n = 223.54$ . To approximate the power over a narrow window of sample sizes centered at 144 and 224, in Table 1 we report the average power along with its standard deviation over sample size ranging from 135 to 150 and in Table 2 for samples of size 215–230 for the four different  $x$  distributions we use to compute the power curves in Figs 2 and 3. In both tables and figures, the sample size is the total sample size of the two groups with the number of the independently chaired funds being one fifth of the total.

The first two lines in Tables 1 and 2 show that a slight difference between the means of the covariate ( $\mu_1 = 16, \mu_0 = 15$ ) in the groups causes a very small decrease in power. Thus, the

TABLE 1 *The average and standard deviation of the simulated power of the t-test when  $\beta_2 = 4, 5$  for  $n = 135, \dots, 150$ , for testing  $H_0: \beta_2 = 0$  versus  $H_1: \beta_2 > 0$  in the full model where  $x \sim N(\mu_1, \sigma)$  in the independent-chaired fund group and  $x \sim N(\mu_0, \sigma)$  in the other group*

	$\beta_2 = 4$		$\beta_2 = 5$	
	Average power	Standard deviation	Average power	Standard deviation
$\mu_1 = \mu_0 = 15$	0.749	0.054	0.881	0.030
$\mu_1 = 16, \mu_0 = 15$	0.724	0.035	0.862	0.031
$\mu_1 = 21, \mu_0 = 15$	0.522	0.053	0.633	0.050
$\mu_1 = 22, \mu_0 = 15$	0.403	0.051	0.464	0.037

Note. The error term is  $N(0, \sigma_\epsilon)$  with  $\sigma_\epsilon = \sqrt{72}$ .

TABLE 2 *The average and standard deviation of the simulated power of the t-test when  $\beta_2 = 4, 5$  for  $n = 215, \dots, 230$  for testing  $H_0: \beta_2 = 0$  versus  $H_1: \beta_2 > 0$  in the full model where  $x \sim N(\mu_1, \sigma)$  in the independent-chaired fund group and  $x \sim N(\mu_0, \sigma)$  in the other group*

	$\beta_2 = 4$		$\beta_2 = 5$	
	Average power	Standard deviation	Average power	Standard deviation
$\mu_1 = \mu_0 = 15$	0.880	0.028	0.963	0.012
$\mu_1 = 16, \mu_0 = 15$	0.878	0.030	0.967	0.014
$\mu_1 = 21, \mu_0 = 15$	0.678	0.064	0.802	0.030
$\mu_1 = 22, \mu_0 = 15$	0.560	0.036	0.626	0.050

Note. The error term is  $N(0, \sigma_\epsilon)$  with  $\sigma_\epsilon = \sqrt{72}$ .

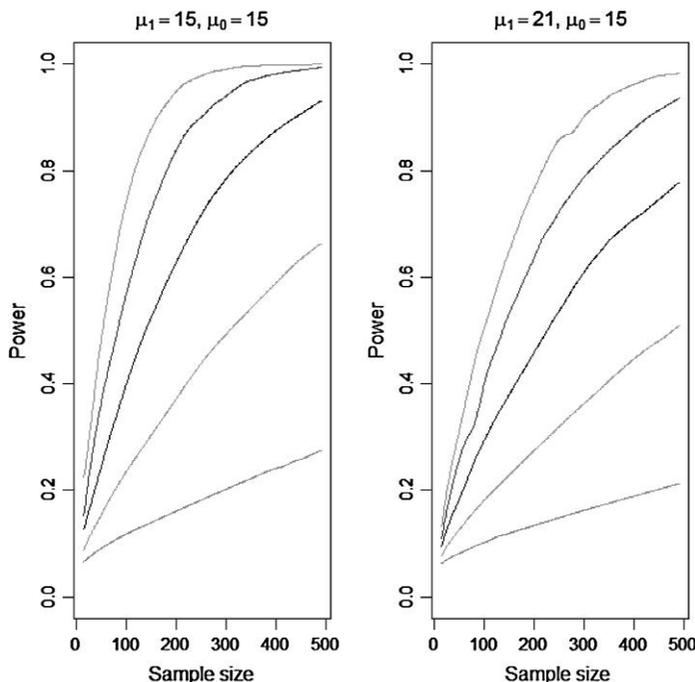


FIG. 2. Smoothed power curves for testing  $H_0: \beta_2 = 0$  versus  $H_1: \beta_2 > 0$ . Friedman's (1984) super smoother was used to smooth the simulated power curves. Left panel: the means and the standard deviations of the covariate are the same in the two groups. As the total variance of  $y = 144$ , the error term is  $N(0, \sigma_e)$  with  $\sigma_e = \sqrt{72}$ . Red corresponds to  $\beta_2 = 1$ , green to  $\beta_2 = 2$ , black to  $\beta_2 = 3$ , blue to  $\beta_2 = 4$  and purple to  $\beta_2 = 5$ . The common variance  $\sigma^2 = 4.24, 4.22, 4.20, 4.17, 4.12$  for  $\beta_2 = 1, 2, 3, 4, 5$ , respectively. Right panel: the covariate satisfies  $x \sim N(21, \sigma)$  in the independent-chaired fund group and  $x \sim N(15, \sigma)$  in the other group, while the group variances remain the same, i.e.  $\sigma_1^2 = \sigma_0^2 = \sigma^2$ . The error term again is  $N(0, \sigma_e)$  with  $\sigma_e = \sqrt{72}$ . Red corresponds to  $\beta_2 = 1$ , green to  $\beta_2 = 2$ , black to  $\beta_2 = 3$ , blue to  $\beta_2 = 4$  and purple to  $\beta_2 = 5$ . The common variance  $\sigma^2 = 3.35, 3.19, 3.0, 2.79, 2.54$  for  $\beta_2 = 1, 2, 3, 4, 5$ , respectively.

SEC approximation is adequate if the means are close. On the other hand, when there is a noticeable difference between the two groups' covariate means ( $\mu_1 = 21, \mu_0 = 15$ ), the power of the regression declines substantially. This can be seen in Fig. 2 where the power curves for detecting each value of  $\beta_2$  in the right panel are lower than the corresponding curves in the left panel for all sample sizes.

We show next that in the OEA/SEC memorandum a power analysis for the *two-sided* two-sample  $t$ -test is carried out, even though the hypothesis is stated in terms of a one-sided alternative (see pp. 6–7 in Spatt, 2006b). In Fig. 3, the power curves for testing  $H_0: \beta_2 = 0$  versus  $H_1: \beta_2 \neq 0$  when the predictor has the *same* distribution in the two groups are plotted. Table 3 reports the average and standard deviation of the power values for  $n = 135, \dots, 150$  and when  $n = 215, \dots, 230$  for testing  $H_0: \beta_2 = 0$  versus  $H_1: \beta_2 \neq 0$  when the true difference in average return for the two groups is  $\beta_2 = 1, 2, 3, 4, 5$ . These approximate the power for the sample sizes 144 and 224 obtained by the SEC.

As expected from the theory presented in Section 3, the power curves in Fig. 3 and the entries in Table 3 agree with the reported power analysis of the OEA/SEC memorandum. The dashed vertical line in Fig. 3 indicates the sample size 143.86 on the  $x$ -axis and the solid vertical line marks the

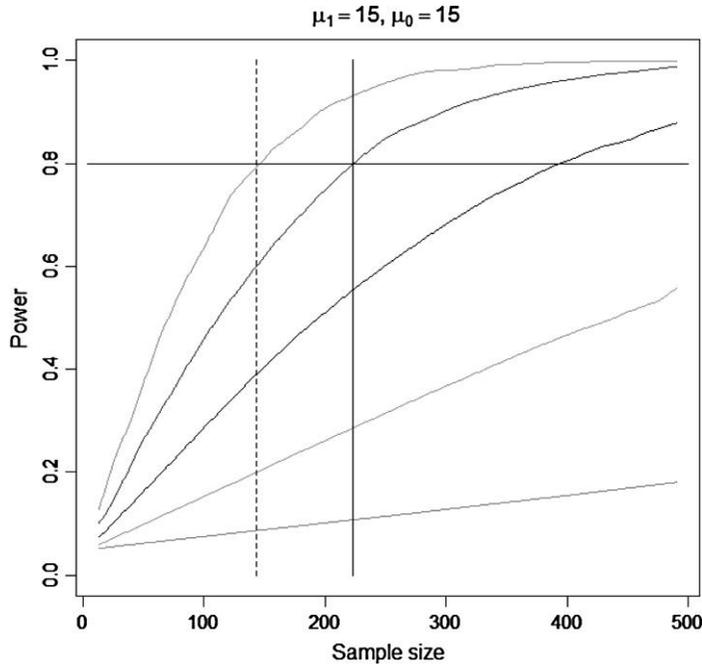


FIG. 3. Smoothed power curves as a function of sample size for testing  $H_0: \beta_2 = 0$  versus  $H_1: \beta_2 \neq 0$ . Friedman’s (1984) super smoother was used to smooth the simulated power curves. The horizontal solid line marks power 80%. The dashed vertical line marks sample size 143.86 in the  $x$ -axis and the solid vertical line marks sample size 223.54. The covariate satisfies  $x \sim N(15, \sigma)$  in the independently chaired fund group and  $x \sim N(15, \sigma)$  in the management-chaired group. The error term is  $N(0, \sigma_e)$ . Red corresponds to  $\beta_2 = 1$ , green to  $\beta_2 = 2$ , black to  $\beta_2 = 3$ , blue to  $\beta_2 = 4$  and purple to  $\beta_2 = 5$ .

TABLE 3 *The average and standard deviation of the simulated power of the t-test for testing  $H_0: \beta_2 = 0$  versus  $H_1: \beta_2 \neq 0$  in the full model where  $x \sim N(15, \sigma)$  in both the independently chaired and the management-chaired fund groups (identical distributions in the two groups)*

	Approximate power $n = 144$		Approximate power $n = 224$	
	Average power	Standard deviation	Average power	Standard deviation
$\beta_2 = 1$	0.087	0.005	0.106	0.005
$\beta_2 = 2$	0.195	0.014	0.290	0.019
$\beta_2 = 3$	0.394	0.035	0.527	0.032
$\beta_2 = 4$	0.605	0.066	0.807	0.044
$\beta_2 = 5$	0.807	0.041	0.935	0.021

Note. The error term is  $N(0, \sigma_e)$  with  $\sigma_e = \sqrt{72}$ .

sample size 223.54. These were the sample sizes identified in the OEA/SEC power study required to achieve 80% power.<sup>15</sup> This finding leads us to conclude that the SEC considered the two-sided test although the memorandum describes a one-sided alternative. Table 4 reports the same quantities as Table 3 but here the predictor distribution has mean 21 in the independent group and 15 in the other

<sup>15</sup> The sample sizes determined by the OEA/SEC are given in Fig. 2 on p. 11 of their memorandum (Spatt, 2006b).

TABLE 4 *The average and standard deviation of the simulated power of the t-test for testing  $H_0: \beta_2 = 0$  versus  $H_1: \beta_2 \neq 0$  in the full model where  $x \sim N(21, \sigma)$  in the independently chaired fund group and  $x \sim N(15, \sigma)$  in the management-chaired fund group*

	Approximate power $n = 135, \dots, 150$		Approximate power $n = 215, \dots, 230$	
	Average power	Standard deviation	Average power	Standard deviation
$\beta_2 = 1$	0.074	0.003	0.087	0.003
$\beta_2 = 2$	0.140	0.009	0.205	0.009
$\beta_2 = 3$	0.256	0.023	0.387	0.023
$\beta_2 = 4$	0.388	0.043	0.551	0.043
$\beta_2 = 5$	0.523	0.078	0.707	0.078

Note. The error term is  $N(0, \sigma_\epsilon)$  with  $\sigma_\epsilon = \sqrt{72}$ .

group. For every value of  $\beta_2$ , the power is noticeably smaller in Table 4 than in Table 3. Thus, the greater the difference between the averages of the predictors in the two groups, the lower the power of the test to detect a difference in returns.

#### 4. Discussion

This paper illustrates the effect an incorrect statistical calculation can have on the strength of the conclusions derived from it. The calculation made by the SEC is only correct in the unrealistic situation where the distribution of the relevant predictor variables is the same in both groups. As demonstrated in the power simulations in Section 3, even if the only difference between the two types of funds is in the means of the distributions of a single predictor, the calculation made by the SEC *overestimates* the statistical power of the test. While our study only considered the case of one predictor variable, the results should hold in the more general case where there are several predictors. Indeed, it is less likely that the joint distribution of several predictors will be identical in both groups of fund managers. In view of the large difference in the amounts (assets) managed by the two types of funds, this problem is likely to occur if the parties develop a full regression model in future proceedings. Other important results concerning the power of tests in multiple linear regressions are given by Andrews (1989) and Gatsonis & Sampson (1989).

Recent studies in the financial and business literature have been concerned with factors, e.g. the cost of acquiring information about a firm, that are related to the effectiveness of outside directors. The paper by Duchin *et al.* (2008) explores this topic and provides references to literature. The relevance of the studies available to the SEC focused on comparing the returns of funds with a high fraction of independent directors to those with the legally specified minimum number to the proposed regulation is logically questionable. The regulation concerns those funds whose sponsors will make trades where there is a potential conflict of interest, e.g. an investment bank is managing an initial public offering and also purchasing some of those shares for the funds it manages. One expects such trades to be a *relatively small* proportion of the trades made by a mutual fund. Hence, studies of the profitability of this special type of trade relative to other trades made by the fund or to similar ‘special’ trades made by funds with independent chairs and boards would be more focused on the problem the SEC regulation is addressing. This problem of choosing a statistically sound ‘control’ group arises in the choice of comparison groups (or individual comparators) in equal

employment cases where courts have expressed concern with studies that are overinclusive, e.g. include all employees rather than those qualified for the job or not subject to the same employment practice being scrutinized, or underinclusive, e.g. comparing a plaintiff to a subset of all comparable applicants or employees rather than the full set.<sup>16</sup>

## REFERENCES

- ANDREWS, D. W. K. (1989) Power in econometric applications. *Econometrica*, **57**, 1059–1090.
- BEALS, R. E. (1972) *Statistics for Economists: An Introduction*, Chicago, IL: Rand McNally and Co.
- BOBROFF, G. & MACK, T. (2004) Assessing the Significance of Mutual Fund Board Independent Chairs, submitted as part of Fidelity’s comment letter to the Commission for consideration of Investment Company Governance rules. <http://www.sec.gov/rules/proposed/s70304/fidelity031804.htm>.
- DUCHIN, R., MATSUSAKA, J. & OZBAS, O. (2008) When Are Outside Directors Effective? University of Southern California Law and Economics Working Paper Series. Working Paper 71. <http://law.bepress.com/usclwps/lewps/art71>.
- FEDERAL REGISTER (2001) Role of Independent Directors of Investment Companies, Final Rule, 66 Fed. Reg. 3734. US Government Printing Office, Washington, DC.
- FEDERAL REGISTER (2004) Investment Company Governance, Final Rule, 69 Fed. Reg. 46, 378. US Government Printing Office, Washington, DC.
- FRIEDMAN, J. H. (1984) A variable span scatterplot smoother. Laboratory for Computational Statistics, Technical Report 5, Stanford University, Department of Statistics.
- GATSONIS, C. & SAMPSON, A. R. (1989) Multiple correlation: exact power and sample size calculations. *Psychological Bulletin*, **106**, 516–524.
- HSIEH, F. Y., BLOCH, D. A. & LARSEN, M. D. (1998) A simple method of sample size calculation for linear and logistic regression. *Statistics in Medicine*, **17**, 1623–1634.
- KAPADIA, A. S., CHAN, W. & MOYE, L. (2005) *Mathematical Statistics with Applications*. Boca Raton, FL: Chapman and Hall/CRC.
- KLEINBAUM, D. G., KUPPER, L. L., NIZAM, A. & MULLER, K. (2008) *Applied Regression Analysis and Other Multivariate Methods* (4th ed.). North Scituate, MA: Duxburg Press.
- SPATT, C. S. (2006a) *Literature Review on Independent Mutual Fund Chairs and Directors*. OEA/SEC MEMORANDUM. [www.sec.gov/rules/proposed/s70304/oeamemo122906-litreview.pdf](http://www.sec.gov/rules/proposed/s70304/oeamemo122906-litreview.pdf).
- SPATT, C. S. (2006b) *Power Study as Related to Independent Mutual Fund Chairs*. OEA/SEC MEMORANDUM. <http://www.sec.gov/rules/proposed/s70304/oeamemo122906-powerstudy.pdf>.

## Appendix A

From model (1), we have

$$\text{Var}(y) = \beta_1^2 \text{Var}(x) + \beta_2^2 \text{Var}(I) + 2\beta_1\beta_2 \text{Cov}(x, I) + \sigma_\varepsilon^2. \quad (\text{A.1})$$

<sup>16</sup> For example, in *Blair v. Henry Filters, Inc.*, 101 FEP Cases 1345 (6th Cir. 2007) the statistical analyses of both the plaintiff (Ibid at n. 13) and the defendant (at n. 14) were faulty. In particular, the defendant submitted data on the ages of employees in all three of its companies rather than the one in which the plaintiff was employed. The opinion referred to *Bender v. Hecht’s Department Stores*, 455 F.3d 612 (6th Cir. 2006), where criteria for determining the relevant comparison population in RIF cases is described. This issue was also discussed by Judge Posner in *Crawford v. Indiana Harbor Belt Railroad Co.*, 461 F.3d 844, 846 (2006), where he notes that the comparators need not be ‘nearly identical’ to the plaintiff but should be ‘sufficiently comparable’ to suggest that the plaintiff ‘was singled out for worse treatment’.

We assume that the indicator  $I$  follows the Bernoulli distribution with parameter  $\pi$ . Also, we assume that when  $I = 1$ ,  $x$  has mean  $\mu_1$  and standard deviation  $\sigma_1$  and when  $I = 0$ ,  $x$  has mean  $\mu_0$  and standard deviation  $\sigma_0$ . Thus,  $\text{Var}(I) = \pi(1 - \pi)$ . Also,

$$\text{Cov}(x, I) = E(xI) - E(x)E(I) = \pi\mu_1 - \pi(\pi\mu_1 + (1 - \pi)\mu_0) = \pi(1 - \pi)(\mu_1 - \mu_0),$$

$$\text{Var}(x) = E(\text{Var}(x|I)) + \text{Var}(E(x|I)),$$

where

$$\text{Var}(x|I) = \begin{cases} \sigma_1^2 & \text{with prob } \pi, \\ \sigma_0^2 & \text{with prob } (1 - \pi). \end{cases}$$

$$\text{Also, } E(x|I) = \begin{cases} \mu_1 & \text{with prob } \pi, \\ \mu_0 & \text{with prob } (1 - \pi). \end{cases}$$

Let  $W = \text{Var}(x|I)$  and  $Z = E(x|I)$ . Then,  $E(\text{Var}(x|I)) = E(W) = \sigma_1^2\pi + \sigma_0^2(1 - \pi)$  and

$$\text{Var}(E(x|I)) = \text{Var}(Z) = E(Z^2) - (E(Z))^2$$

$$= \mu_0^2(1 - \pi) + \mu_1^2\pi - (\mu_0(1 - \pi) + \mu_1\pi)^2 = \pi(1 - \pi)(\mu_1 - \mu_0)^2.$$

$$\text{Thus, } \text{Var}(x) = \sigma_1^2\pi + \sigma_0^2(1 - \pi) + \pi(1 - \pi)(\mu_1 - \mu_0)^2.$$

In summary, substituting in (A.1), we obtain

$$\begin{aligned} \text{Var}(y) = & \beta_1^2(\sigma_1^2\pi + \sigma_0^2(1 - \pi) + \pi(1 - \pi)(\mu_1 - \mu_0)^2) + \beta_2^2\pi(1 - \pi) \\ & + 2\beta_1\beta_2\pi(1 - \pi)(\mu_1 - \mu_0) + \sigma_\varepsilon^2. \end{aligned}$$