Yet another look at mutual fund tournaments

Alexei P. Goriaev, Theo E. Nijman, and Bas J.M. Werker*†
New Economic School, Tilburg University, Tilburg University

October 30, 2003

Abstract

We revisit the empirical evidence on the tournament hypothesis for the behavior of mutual fund managers provided by Busse (2001). First, we give analytical expressions for the biases arising in volatility estimates (based on both daily and monthly data) due to first-order autocorrelation effects in daily fund returns. These calculations show that tests of the tournament hypothesis based on monthly data are more robust to autocorrelation effects than tests based on daily data. Second, to address the impact of cross-correlated fund returns on these tests, we provide explicit conditions under which the tests proposed in the literature have appropriate size properties.

Keywords: Contingency table test, Strict factor structure, Temporal aggregation, Tournament hypothesis.

JEL Classification: G11

*Corresponding author: Bas J.M. Werker, Econometrics and Finance Group, CentER, Tilburg University, P.O. Box 90153, 5000 LE, Tilburg, The Netherlands, Werker@TilburgUniversity.nl.

†The authors wish to thank Jeff Busse and two anonymous referees for many useful comments.

1 Introduction

During the last two decades, the mutual fund industry experienced tremendous growth both in number of funds and amount of assets under management. It is not surprising that this industry attracts a lot of attention of the regulatory agencies that would like to ensure that fund managers select investment strategies that are optimal from the investors' point of view. The joint occurrence of two well-established facts in the mutual fund industry may lead to an agency conflict between mutual fund managers and mutual fund shareholders. First, managers' compensation is typically based on a percentage of the fund's net assets (see, e.g., Khorana, 1996). Second, the top-performing funds receive the bulk of new cash inflows, while bad performance does not lead to significant outflows (see, e.g., Sirri and Tufano, 1998). Together with the observation that at least some investors look at calendar year fund performance for their investment decisions (see, Chevalier and Ellison, 1997, p. 1183), these effects suggest that mutual fund managers participate in annual tournaments where they compete for the top rankings. This leads to the conjecture that funds performing badly during the first part of the year have an incentive to increase risk in the second part of the year in order to try to catch up with mid-year winners at the end of the year. This conjecture is called the *tournament hypothesis*.

A number of studies verifies the tournament hypothesis from an empirical point of view. Brown, Harlow, and Starks (1996) find evidence supporting the tournament hypothesis using a contingency table methodology applied to monthly data of 334 growth funds over the period 1976-1991. This technique compares volatility changes from the first to the second semester with mid-year performance. Koski and Pontiff (1998) use regression analysis and find a negative relation between interim performance and subsequent change in risk, in line with the tournament hypothesis. Koski and Pontiff (1998) use 798 domestic equity funds from 1992-1994. Finally, Chevalier and Ellison (1997), using 398 growth and growth&income funds from 1982-1992, obtain different regression results depending on whether fund risk is measured on the basis of fund portfolio holdings or monthly fund returns.

While previous studies, using monthly fund returns, have found strong evidence in favor of the tournament hypothesis over the period studied, Busse (2001) finds no such evidence using the contingency table methodology applied to daily data of 230 US domestic equity funds in 1985-1995 (new entrants after 1984 are not included). This appears surprising, since daily data provide, in principle, much more precise volatility estimates and hence tests based on daily data can be expected to be more powerful in detecting evidence in favor of the tournament hypothesis than tests based on monthly data. Busse (2001) offers two explanations for this apparent paradox. First, Busse (2001) argues that biases in monthly volatility estimates due to autocorrelation in daily fund returns may adversely affect the tests based on monthly data. Such autocorrelation could be caused by mutual fund managers loading on small stocks that exhibit first-order autocorrelation due to non-synchronous trading effects (compare, e.g., Boudoukh, Richardson, and Whitelaw, 1994). Second, Busse (2001) notes that standard statistical tests used so far in the literature rely on the contestable assumption that mutual fund returns are cross-sectionally independent.

In this paper, we analyze both the impact of autocorrelation and cross-correlation on tests of the tournament hypothesis from an analytical point of view. First, we calculate the biases arising in volatility estimates based on monthly or daily return data due to first-order autocorrelation in the returns. More precisely, we express the ratio of the volatility over the second part of the year with respect to that over the first part of the year (henceforth, "standard deviation ratio" or SDR) in terms of the autocorrelation of daily fund returns. We do this both for volatilities estimated using daily data (henceforth called the "daily SDR") and for those based on monthly data (henceforth "monthly SDR"). These results show that, in line with Busse's (2001) claim, monthly SDR's are indeed (in absolute terms) more sensitive to changes in the autocorrelation pattern than daily SDR's. However, the smaller absolute bias in daily SDR's has a *larger* effect on the distribution of the tests of the tournament hypothesis, since volatility estimates based on daily data are more precise compared to volatility estimates based on monthly data. Therefore, at the end of the day, tests of the tournament hypothesis based on monthly data are more robust to autocorrelation in daily fund returns than tests based on daily data[1]. Thus, if the autocorrelation effects are such that they adversely affect the tournament tests based on monthly data, they would certainly affect the tests based on daily data. Our calculations also rationalize some of the differences in the autocorrelation of daily returns that Busse (2001) reports for funds whose monthly volatility increases/decreases during the second half of the year.

While we only consider contingency table tests as in Brown, Harlow, and Starks (1996), Busse (2001) also reports empirical evidence concerning the tournament hypothesis using regression techniques similar to those of Koski and Pontiff (1999). The reader will easily convince herself that our arguments extend to those techniques, because our discussion concerning both the bias in volatility estimates and the required form of (in)dependence in mutual fund returns are independent of the actual test employed (be it a contingency table test or a regression based tests). Busse (2001) also presents tests of the tournament hypothesis based on ratios of the residual volatility of fund returns in an MA(1) model. The conclusions from these tests for the 1985-1995 sample period are the same as those based on the total return volatility (see Table 2 and Table 3 of Busse, 2001). These test results are not corrected for cross-correlation effects in fund returns. However, it is important to note that Busse's (2001) and our results only hold for studies that are based on return data only. Studies using actual mutual fund holdings (like, e.g., Chevalier and Ellison, 1997) do not necessarily suffer from the biases discussed in the present paper. The extent to which mutual fund holdings are correlated in cross-sections of mutual funds and the effect on tests of the tournament hypothesis, still has to be established.

A second contribution of the present paper is the derivation of explicit conditions for the validity of the tournament hypothesis tests. We show that these tests implicitly assume that fund returns follow a factor structure with *uncorrelated* idiosyncratic errors across funds. This is what Chamberlain and Rothchild (1983) have named a *strict factor structure*. If this hypothesis is not satisfied, size corrected p-values can be obtained using simulation or bootstrap techniques. When using these size corrected p-values, the evidence in favor of the tournament hypothesis based on monthly fund returns disappears. Busse (2001) also reports size corrected test results based on daily fund returns, finding again no evidence in favor of the tournament hypothesis.

[1]Stated formally, the non-centrality parameter in the χ^2-distribution arising from autocorrelation effects, which is given by the squared bias over the sampling variance, is smaller if one uses monthly data.

Clearly, the tests we discuss in this paper are based on fund returns only. Given the limited power that these tests are found to have concerning the tournament hypothesis, an obvious next step would be to check whether other sources of information may shed more light on the empirical relevance of the tournament hypothesis. Besides mutual fund holdings, flows into mutual funds might be interesting here as well. For instance, Berk and Green (2002) discuss a model that relates flows to performance in rational markets in an ingenious way. Based on such a model, it might be possible to extract information from flows in and out mutual funds, that can be used to support or refute the tournament hypothesis. This is outside the scope of the present paper.

The structure of the present paper is as follows. In Section 2, we present the analytical results concerning the bias arising in volatility estimates based on daily and monthly data and show that tests of the tournament hypothesis based on daily data are more severely affected by autocorrelation effects than tests based on monthly data. Using these calculations, we also rationalize part of the results reported in Busse (2001) about the relation between monthly volatility changes and changes in the autocorrelation pattern of fund returns. Section 3 discusses the conditions under which the contingency table tests have the appropriate size and gives an illustrative empirical example. Section 4 concludes and the appendix gathers some proofs.

2 Effect of autocorrelated returns on SDR's

Following Busse (2001), we consider first-order moving average (MA(1)) specifications for daily fund returns. More precisely, using the same notation as Busse (2001), we have, for fund p and day d,

$$r_{p1d} = \mu_{p1} + \theta_{p1}\varepsilon_{p1,d-1} + \varepsilon_{p1d}, \quad d = 1, \ldots, D, \tag{1}$$

$$r_{p2d} = \mu_{p2} + \theta_{p2}\varepsilon_{p2,d-1} + \varepsilon_{p2d}, \quad d = D+1, \ldots, D_y, \tag{2}$$

where $d = 1, \ldots, D$ refers to the first part of the year (subindex "1") and $d = D+1, \ldots, D_y$ refers to the second half of the year (subindex "2"). In Appendix A, we calculate the biases in the daily and monthly SDR's as a function of the autocorrelation coefficients θ_{p1} and θ_{p2}. These standard deviation ratios are crucial in the contingency table tests for the tournament hypothesis that we shortly describe now. The daily SDR for fund p is given by

$$SDR_p = \sqrt{\frac{\frac{1}{D_y-D-1}\sum_{d=D+1}^{D_y}\left(r_{p2d} - \bar{r}_{p2}\right)^2}{\frac{1}{D-1}\sum_{d=1}^{D}\left(r_{p1d} - \bar{r}_{p1}\right)^2}}, \tag{3}$$

where $\bar{r}_{p1}$ denotes the average return over the first part of the year and $\bar{r}_{p2}$ that over the second part of the year. Let $Med(\bar{r}_{p1})$ denote the median average fund return over the first part of the year and let $Med(SDR_p)$ denote the median SDR. The contingency table test statistic for the tournament hypothesis given in Brown, Harlow, and Starks (1996) can now be written as

$$Q = \left(\frac{\text{number of funds with } \bar{r}_{p1} < Med(\bar{r}_{p1}) \text{ and } SDR_p > Med(SDR_p)}{\text{total number of funds}} - 0.25\right)^2. \tag{4}$$

In the present section we assume that returns are independently distributed. The next section focusses on the effect of cross-sectional dependence. Under the null hypothesis that past returns and subsequent risk-taking are independent, and that returns are indeed distributed independently over funds, Q follows asymptotically a χ^2 distribution with one degree of freedom. The corresponding critical values are routinely used in many empirical studies. In Section 3 and Appendix B the distribution of the test statistic Q is derived under more realistic assumptions.

From the results in the appendix, we find in first order approximation and under the null of constant idiosyncratic volatility during the year (i.e., $\sigma_{\varepsilon p1} = \sigma_{\varepsilon p2}$),

$$\text{Monthly SDR bias} \approx \theta_{p2} - \theta_{p1}, \tag{5}$$

$$\text{Daily SDR bias} \approx \bar{\theta}_p(\theta_{p2} - \theta_{p1}), \tag{6}$$

where $\bar{\theta}_p$ denotes the average MA(1) coefficient for fund p over the year, i.e., $\bar{\theta}_p = (\theta_{p1} + \theta_{p2})/2$.

Busse (2001) claims that first-order autocorrelation in daily fund returns biases monthly volatility estimates. The effect of the autocorrelation on the monthly SDR is apparent from equation (5), and substantiates Busse's (2001) claim. If, due to external reasons, fund managers load more on less liquid stocks during the second part of the year, they may increase first-order autocorrelation in daily fund returns and thereby influence the monthly SDR. However, the daily SDR's are also affected by the changing autocorrelation, albeit to a lesser extent since $\bar{\theta}$ is (see Table 5 of Busse, 2001) at most 0.2. The relevant question is what the effect of changing autocorrelation is on the contingency table tests. This effect is measured by the non-centrality parameter in the χ^2_1-distribution caused by the daily autocorrelation. This non-centrality parameter is given by the squared bias over the estimation variance (see, e.g., Godfrey, 1991, p. 18). This quantity can be calculated using the relative efficiency of both SDR estimates as given in Busse (2001). Given the reported daily MA(1) coefficients in Busse (2001) of about 0.17[2], the squared monthly SDR bias is about $(1/0.17)^2 = 34.60$ times that of the squared daily SDR bias. The relative efficiency of daily SDR's with respect to monthly SDR's is given in Busse (2001) as about 47.3[3]. Comparing the squared biases to the relative efficiency of the volatility estimates, we see that the contingency table test based on monthly data is more robust to changes in the first-order autocorrelation of daily returns than tests based on daily data. The autocorrelation effect alone can thus not account for the difference in empirical evidence concerning the tournament hypothesis. If changes in the first-order autocorrelation of daily fund returns, possibly generated by changes in the fund's loading on small stocks, affect the contingency table tests based on monthly fund returns, they will also affect the tests based on daily fund returns and in the same direction. However, Table 2 in Busse (2001) shows that, for some specifications, the monthly tests seem to confirm the tournament hypothesis, while the daily tests do not. In all cases, the direction of the rejection of the monthly and daily tests is opposite. We take this as evidence that changes in autocorrelation over the year

[2] This number is calculated as the MA(1) coefficient that induces a first-order autocorrelation which equals the average autocorrelation reported in Panel B of Table 5 in Busse (2001).

[3] This number is reported by Busse at the end of Section III.B. It is calculated neglecting serial correlation in daily fund returns.

are not the driving factor of these empirical results. We argue that these results are more due to ignored cross-sectional dependence which is also mentioned in Busse (2001). This is the topic of further investigation in the next section.

Once more we want to emphasize that the analysis in this section refers to the marginal effect of autocorrelation, not taking into account a possible interaction with the cross-sectional dependence effect which is also mentioned in Busse (2001). In the next section, we study the effect of cross-sectional dependence in more detail by providing explicit conditions under which the tournament tests employed so far in the literature have the correct size. Again, for expository reasons, we focus on the contingence table test of Brown, Harlow, and Starks (1996), but the results readily extend to regression based tests.

3 Contingency table test and strict factor structure

In this section, we take a closer look at the conditions needed for the contingency table test statistic (4) to have a χ^2_1 limiting null distribution. Busse (2001) notes that cross-correlation in fund returns violates the independence assumption used in deriving the χ^2 tests for equal cell frequencies. In the appendix, we show that actual independence of fund returns is not necessary for the contingency table test statistic to have a χ^2 limiting null distribution. As long as fund returns are generated from a factor model in which the idiosyncratic returns are uncorrelated across funds, χ^2 critical values may be used to obtain a test with the required size. Such a factor structure is what Chamberlain and Rothchild (1983) have named a strict factor structure. It is important to note that the test statistic is still based on the raw returns, not on estimated idiosyncratic returns.

Busse (2001) adopts a bootstrap technique to simulate p-values that are robust to possible cross-correlation of fund returns. He reports that, both for daily and monthly fund returns and using these robust p-values, no evidence in favor of the tournament hypothesis is found in the sample under consideration. If the cross-correlation is taken into account, neither daily nor monthly fund returns point in the direction of strategic risk taking by mutual fund managers. To illustrate this point and to make the paper self-contained, we perform the contingency table test for strategic risk-taking on monthly data for 26 annual tournaments. Similar tables are given in Brown, Harlow, and Starks (1996) and Busse (2001). Our sample of US growth funds comes from CRSP Survivor-Bias Free Mutual Fund Database[4] (data as of the end of 2002). The sample includes all funds with ICDI "Aggressive Growth" and "Growth and Income" objectives from 1976 to 2001[5]. In line with Busse (2001), we split years in two equal periods of six months when calculating monthly SDR's.

Table 1 reports the results. Recall that the tournament hypothesis states that the low-high frequency exceeds 25%. Using χ^2_1-based p-values, we find apparently significant results in almost half of the 26 annual tournaments. A joint test based on these p-values would strongly support the tournament hypothesis. However, for about half of the years for which the results appear significant, the results are opposite to the tournament hypothesis,

[4]Source: CRSP, Center for Research in Security Prices. Graduate School of Business, The University of Chicago [2002]. Used with permission. All rights reserved. www.crsp.uchicago.edu.

[5]Our resuls remain qualitatively the same when we use alternative compositions of the growth fund category, e.g., as defined by Morningstar. Results are in a previous version of this paper and available upon request.

i.e., the low-high frequency is less than 25% and losing funds are found to (relatively) reduce risk-taking over the second part of the year. In order to accommodate possibly cross-correlated idiosyncratic monthly mutual fund returns, we determine the distribution of the tournament tests, under the null of no strategic risk-taking, using simulation. To explain the procedure in more detail, for each month, we simulate independently a vector of (correlated) fund returns from a multivariate normal distribution. The mean vector and variance matrix of this distribution are estimated from the observed monthly fund returns. In these simulated fund returns there is, by construction, no tournament effect. For each null simulation, we calculate the realization of the contingency table test statistic (4). This is replicated 10,000 times from which the simulated p-values are obtained.

This way of simulating robust p-values has a particularly nice invariance property. Suppose that actual fund returns are generated from some factor model and one would want to base the tournament hypothesis test on the variances of the idiosyncratic errors in this factor model. Then we could still obtain valid simulated p-values for our test statistic as above (i.e., without simulating from the factor model, but from a multivariate normal distribution). To see why this is true, suppose that a factor model had been estimated and that fund returns were simulated using the estimated factor loadings, the observed factor values, and the variance matrix of idiosyncratic fund returns. In the end, the simulated fund returns would then again be normally distributed with exactly the same mean vector and variance matrix as above. This follows immediately from the standard orthogonality of the regression decomposition. Our zero-factor simulation hence generates fund returns that are distributionally equal to those generated from any other factor model.

Our simulations assume normality of monthly fund returns. Clearly, this normality assumption is innocuous, if sufficient regularity conditions are satisfied for a central limit theorem to hold true. It is important that the simulation setup allows (idiosyncratic) fund returns to be correlated across funds. Busse (2001) uses bootstrapped critical values. The advantage of such an approach is that it does not rely on any normality assumptions, which may be relevant when using daily data. The disadvantage is that it is computationally somewhat more intensive. Moreover, as Busse (2001) reports, p-values based on a Monte Carlo approach assuming normality do not produce materially different results from p-values based on a bootstrap approach.

The last column of Table 1 reports the cross-correlation robust p-values. In line with Busse (2001), all evidence in favor of the tournament hypothesis disappears once the cross-correlation is accounted for. Thus, when using corrected p-values, monthly and daily fund returns lead to the same conclusion, where, given the increased efficiency, tests based on daily fund returns can be more powerful in detecting evidence in favor of the tournament hypothesis.

4 Conclusion

The present paper confirms the conclusion in Busse (2001) that, for US equity funds over the sample periods studied so far, there is little empirical evidence in favor of the tournament hypothesis for mutual fund managers. We study the reasons for the difference in empirical results of the tournament test based on daily and monthly returns. We take an analytical point of view and argue that the source of spurious evidence found in the past is not so much a neglected temporal correlation in returns, but more a neglected cross-correlation between idiosyncratic fund returns. As documented by Busse (2001), autocorrelation in daily fund returns indeed biases both monthly and daily SDR's, but we show that tests based on monthly SDR's are more robust to these effects than tests based on daily SDR's. Thus, spurious (due to possible autocorrelation effects) evidence in favor of the tournament hypothesis based on monthly returns, would, ceteris paribus, certainly show up in tournament tests based on daily returns for the same sample period.

On the other hand, neglecting cross-correlation in fund returns may lead (as is also noted in Busse, 2001) to spurious inference. We show that cross-correlated fund returns do not necessarily invalidate the tournaments tests used so far in the literature, as long as the idiosyncratic fund returns in some factor model are uncorrelated across funds. When idiosyncratic return cross-correlation is accounted for, the empirical evidence (based on returns and using the techniques applied so far) in favor of the tournament hypothesis in the 1976-2001 sample of US funds we study, disappears. Clearly, the biases studied in the present paper do not necessarily invalidate the empirical evidence concerning the tournament hypothesis based on actual portfolio holdings of mutual funds.

A Appendix: Biases in daily and monthly SDR's

We consider continuously compounded daily returns r_{pjd} as given in (1)-(2). If D_m denotes the number of days in a month, the monthly returns, for month m, are given as

$$r_{pjm} = \sum_{d=1}^{D_m} r_{pjd}, \quad j = 1, 2.$$

It is well-known that the stationary variance of daily fund returns during the first half of the year, for fund p, is $(1+\theta_{p1}^2)\sigma_{\varepsilon p1}^2$, where $\sigma_{\varepsilon p1}$ denotes the idiosyncratic volatility of fund p, i.e., $\sigma_{\varepsilon p1}^2 = \text{Var}\{\varepsilon_{p1}\}$. This immediately implies

$$\text{Daily SDR} = \sqrt{\frac{\text{Var}\{r_{p2d}\}}{\text{Var}\{r_{p1d}\}}} = \sqrt{\frac{1+\theta_{p2}^2}{1+\theta_{p1}^2}\frac{\sigma_{\varepsilon p2}}{\sigma_{\varepsilon p1}}} \approx \frac{1+\frac{1}{2}\theta_{p2}^2}{1+\frac{1}{2}\theta_{p1}^2}\frac{\sigma_{\varepsilon p2}}{\sigma_{\varepsilon p1}},$$

where the latter approximation is immediate from $\sqrt{1+x^2} \approx 1 + \frac{1}{2}x^2$, for small x. Note that Busse (2001) reports MA(1) coefficients θ in the interval from 0.0 to 0.2, so that the squared autocorrelation coefficient is at most 0.04.

It is somewhat more complicated to calculate the monthly SDR. From the autocorrelation function of an MA(1) process, we obtain

$$\begin{aligned}\text{Var}\{r_{p1m}\} &= D_m\text{Var}\{r_{p1d}\} + 2(D_m-1)\text{Cov}\{r_{p1d}, r_{p1,d-1}\} \\ &= D_m(1+\theta_{p1}^2)\sigma_{\varepsilon p1}^2 + 2(D_m-1)\theta_{p1}\sigma_{\varepsilon p1}^2, \\ &= D_m(1+\theta_{p1})^2\sigma_{\varepsilon p1}^2 - 2\theta_{p1}\sigma_{\varepsilon p1}^2,\end{aligned}$$

and a similar relation for the second half of the year. Therefore, the monthly SDR is given by

$$\text{Monthly SDR} = \sqrt{\frac{\text{Var}\{r_{p2m}\}}{\text{Var}\{r_{p1m}\}}} = \sqrt{\frac{D_m(1+\theta_{p2})^2 - 2\theta_{p2}}{D_m(1+\theta_{p1})^2 - 2\theta_{p1}}}\frac{\sigma_{\varepsilon p2}}{\sigma_{\varepsilon p1}} \approx \frac{1+\theta_{p2}}{1+\theta_{p1}}\frac{\sigma_{\varepsilon p2}}{\sigma_{\varepsilon p1}},$$

an approximation based on the fact that D_m is much larger than the MA(1) coefficient.

Under the null of constant idiosyncratic volatility during the year (i.e., $\sigma_{\varepsilon p1} = \sigma_{\varepsilon p2}$),

$$\begin{aligned} \text{Monthly SDR bias} &= \text{Monthly SDR} - 1 \approx \theta_{p2} - \theta_{p1}, \\ \text{Daily SDR bias} &= \text{Daily SDR} - 1 \approx \frac{1}{2}(\theta_{p2}^2 - \theta_{p1}^2) = \frac{1}{2}(\theta_{p1} + \theta_{p2})(\theta_{p2} - \theta_{p1}), \end{aligned}$$

since for small x and y we have $\frac{1+x}{1+y} \approx x - y$.

B Appendix: Limiting distribution of the contingency table test

In this appendix, we derive the limiting distribution of the contingency table test mentioned in the main text assuming that monthly mutual fund returns are generated from a strict factor model. Clearly, the same results would hold true for any other data frequency. We assume for the moment that, for fund p in month m,

$$r_{pm} = \alpha_p + \beta_p^T F_m + \varepsilon_{pm},$$

where F_m denotes the vector of factors and where the idiosyncratic errors ε_{pm} are independently $N(0, \sigma_p^2)$ distributed[6]. Define the sample average and the sample volatility of fund p's returns over some period j consisting of k months as $\bar{r}_{pj}$ and $\hat{\sigma}_{pj}$, respectively:

$$\begin{aligned} \bar{r}_{pj} &= \frac{1}{k}\sum_{m=1+k(j-1)}^{kj} r_{pm}, \\ \hat{\sigma}_{pj} &= \sqrt{\frac{1}{k}\sum_{m=1+k(j-1)}^{kj} (r_{pm} - \hat{\mu}_p^{(j)})^2}. \end{aligned}$$

Let $\mathcal{F}$ denote the information in the factors over the complete observational period, i.e. $\mathcal{F} = \sigma(F_1, F_2, \ldots)$. Now, conditionally on $\mathcal{F}$ and under the null hypothesis, the statistics $\bar{r}_{p1}$, $\bar{r}_{p2}$, $\hat{\sigma}_{p1}$, and $\hat{\sigma}_{p2}$ are independently distributed. The independence of the risk-adjustment-ratio $\hat{\sigma}_{p2}/\hat{\sigma}_{p1}$ and the first semester average return $\bar{r}_{p1}$, implies that (conditionally on $\mathcal{F}$ and under the null) the standard χ^2-test statistic Q for independence of risk-adjustment-ratios and first semester returns (4) follows, asymptotically, a χ_1^2 distribution. Formally, under the null hypothesis,

$$\mathcal{L}(Q|\mathcal{F}) \to \chi_1^2.$$

[6] Clearly, normality is, asymptotically, irrelevant for the main results in this appendix as long as variances exist since then one may resort to a central limit theorem argument.

For regression tests, the independence of the idiosyncratic errors ε_{pm} across funds, would guarantee the validity of the standard t-test by the same arguments.

In case the idiosyncratic errors ε_{pm} are correlated across funds, the arguments above no longer hold, even asymptotically. In that case, the number of unbounded eigenvalues of the variance of fund returns is generally infinite and limiting results can no longer be established analytically in general.

References

Berk, J.B., and R.C. Green, 2002. Mutual fund flows and performance in rational markets, working paper.

Boudoukh, J., M. Richardson, and R. Whitelaw, 1994. A tale of three schools: Insights on autocorrelations of short-horizon stock returns, Review of Financial Studies **7**, 539–573.

Brown, Keith C., W. V. Harlow, and Laura T. Starks, 1996. Of tournaments and temptations: An analysis of managerial incentives in the mutual fund industry, Journal of Finance **51**, 85–110.

Busse, Jeffrey A., 2001, Another look at mutual fund tournaments. Journal of Financial and Quantitative Analysis **36**, 53–73.

Chamberlain, Gary, and Michael Rothchild, 1983. Arbitrage, factor structure, and mean-variance analysis on large asset markets, Econometrica **51**, 1281–1304.

Chevalier, Judith, and Glenn Ellison, 1997. Risk taking by mutual funds as a response to incentives, Journal of Political Economy **105**, 1167–1200.

Godfrey, L.G., 1991. Misspecification tests in econometrics, Econometric Society Monographs.

Khorana, Ajay, 1996. Top management turnover: An empirical investigation of mutual fund managers, Journal of Financial Economics **40**, 403–427.

Koski, Jennifer Lynch, and Jeffrey Pontiff, 1999. How are derivatives used? Evidence from the mutual fund industry, Journal of Finance **54**, 791–816.

Sirri, Erik R., and Peter Tufano, 1998. Costly search and mutual fund flows, Journal of Finance **53**, 1589–1622.

Year	# funds	Low-High frequency	χ^2-statistic	p-value (χ^2_1)	p-value (simulated)
1976	64	21.88	1.00	0.3173	0.5051
1977	64	17.19	6.25	0.0124	0.1164
1978	67	21.64	1.21	0.2715	0.4877
1979	66	27.27	0.55	0.4602	0.6328
1980	67	28.36	1.21	0.2715	0.6169
1981	65	20.77	1.86	0.1724	0.5312
1982	72	19.44	3.56	0.0593	0.2916
1983	76	14.47	13.47	0.0002	0.0368
1984	86	26.74	0.42	0.5176	0.7588
1985	92	21.74	1.57	0.2109	0.4384
1986	109	25.69	0.08	0.7738	0.8500
1987	133	23.68	0.37	0.5439	0.7600
1988	162	28.40	2.99	0.0839	0.4985
1989	159	30.50	7.70	0.0055	0.2600
1990	212	30.66	10.87	0.0010	0.2214
1991	217	27.19	1.66	0.1971	0.6337
1992	491	22.61	4.50	0.0339	0.7519
1993	510	21.27	11.33	0.0008	0.5584
1994	715	25.73	0.62	0.4322	0.9090
1995	915	25.25	0.09	0.7661	0.9626
1996	1134	35.63	204.87	0.0000	0.2344
1997	1412	14.02	272.24	0.0000	0.2135
1998	1693	18.96	98.81	0.0000	0.4150
1999	1983	21.58	37.04	0.0000	0.6746
2000	2269	25.39	0.54	0.4625	0.9663
2001	2708	11.19	826.45	0.0000	0.0468

Table 1: Results of the contingency table approach for the 26 annual tournaments. The Low-High column gives the percentage of funds with both a total return over the first six months below median and a risk adjustment ratio (SDR) above median. The χ^2-statistic tests the null hypothesis that population percentages are equal to 25%. Column five presents the p-values of the χ^2-statistic based on the χ^2_1 distribution. The last column reports simulated corrected p-values for the χ^2-statistic. See main text for details.