# A screen for fraudulent return smoothing in the hedge fund industry

Nicolas P.B. Bollen† and Veronika Krepely

Owen Graduate School of Management
Vanderbilt University

January 2006

†Corresponding author. Email Nick.Bollen@owen.vanderbilt.edu. Phone (615) 343-5029. The authors thank an anonymous referee, George Aragon, Cliff Ball, Stephen Brown, Bill Christie, Greg Dyra, Yanqin Fan, Wayne Ferson, Joel Hasbrouck, David Hsieh, Ron Masulis, Steve Ross, John Van, and seminar participants at Louisiana State University, Vanderbilt University, and the UVA Conference on Probability, Financial Derivatives, and Asset Pricing for helpful comments and suggestions. The Dean's Fund for Research and the Financial Markets Research Center at the Owen Graduate School of Management, Vanderbilt University, provided generous research support.

# A screen for fraudulent return smoothing in the hedge fund industry

*Abstract*

This paper constructs a statistical screen for fraudulent return smoothing in the hedge fund industry. We show that if true returns are independently distributed, and a manager fully reports gains but delays reporting losses, then reported hedge fund returns will feature conditional serial correlation. Simulation evidence indicates that the power of the screen is restricted by the limited histories of some funds, but may still be sufficient to deter fraudulent return smoothing. Empirical evidence shows that the probability of observing conditional serial correlation is related to the volatility and magnitude of investor cash flows, consistent with managerial smoothing in response to the risk of capital flight.

**A screen for fraudulent return smoothing in the hedge fund industry**

## I. Introduction

The hedge fund industry has experienced a recent surge in popularity, with the number of funds and assets under management increasing at a much faster rate than in the mutual fund industry. The growth has generated a corresponding increase in aggregate managerial income. Incentive contracts are highly lucrative, usually including a guaranteed management fee between 1% and 2% of fund assets and a performance fee between 15% and 20% of fund profits. Moderately successful managers can quickly become extremely wealthy, which explains a flight of talent from mutual funds and investment banks' proprietary trading departments.[1] Critics in the popular press argue that demand for hedge funds has also contributed to increased instances of fraud, which are usually discovered ex post given the historically low transparency of the industry.[2]

As described by the SEC (2003), "safe harbor" exemptions in the 1933 Securities Act, the 1940 Investment Company Act, and the 1940 Investment Advisers Act allow hedge funds to avoid substantial disclosure and record-keeping requirements. The SEC (2005) suggests that some managers are abusing their autonomy, citing as evidence 51 hedge fund fraud enforcement cases in the past five years. In light of the tremendous growth of the hedge fund industry, the opacity with which hedge funds operate, and the resulting potential for fraud, the SEC recently adopted Rule 203(b)(3)-2 under the Investment Advisers Act to eliminate the private adviser exemption. As a result, most U.S. hedge fund managers will be required to register as investment advisers by 1 February 2006.

The specter of new regulation has received a mixed reception. Proponents state that the registration requirement will support the SEC's authority to examine hedge fund operating procedures, including internal valuation of fund assets. Opponents argue that hedge funds are already subject to the anti-fraud provisions of existing rules, that hedge funds exempt from registration are owned by sophisticated investors, and that the SEC will likely be unable to prevent fraudulent activity, even after requiring adviser registration. Random examinations are unlikely to be effective, given the thousands of existing hedge funds and the limited resources of the SEC. An initial quantitative filter may help regulators select funds that have an increased risk of fraud; yet, as noted by dissenting SEC commissioners Atkins and Glassman (2004), such a screen has not yet been developed.

In this paper, we construct one example of a statistical screen that regulators and sophisticated investors could use to select funds for further scrutiny. We then apply the screen to one of the commercially available databases of self-reported hedge fund returns to answer two questions. First, is the accuracy and duration of the available data sufficient to provide the statistical power necessary to correctly identify funds for in-depth examinations? Second, does the screen select funds with an increased risk of fraud?

We design the screen to identify funds which possess a time series pattern of reported returns that is consistent with artificial smoothing. Identifying artificially smoothed returns is important for at least three reasons. First, a hedge fund manager might smooth returns in order to lower the apparent risk of a fund, thereby increasing its risk-adjusted performance and making the fund more attractive to potential investors. Institutional investors would benefit from knowing which funds have suspicious time series patterns because they would then be able to execute more informed due diligence prior to the investment decision. Second, smoothed returns result in fund assets being either overvalued or undervalued. Research in the mutual fund market timing scandal, including studies by Boudoukh et al. (2002) and Zitzewitz (2003), show that inaccurate valuations result in wealth transfers among different investors as they subscribe to or redeem from funds. Similarly, in the context of hedge fund fraud, the SEC's litigation release for Marque Funds includes specific mention of the impact on investors when shareholder activity occurs with inaccurate valuations. Third, and perhaps most important, analysis of the SEC's litigation releases reveals that misrepresentation of returns is one of the most common types of fraud, and is the type of fraud we would expect a returns-based filter could detect. Artificially smoothed returns by themselves would not cause a fund to implode, but may be associated with more flagrant activity, such as hiding losses, and may be a precursor of more serious abuses of managerial discretion.

Any statistical filter requires a specific assumption about managerial behavior. One possibility is that some managers underreport both gains and losses in an effort to lower the measured volatility of hedge fund returns and improve risk-adjusted performance. This behavior implies that fund assets are sometimes overvalued and sometimes undervalued by the manager. We argue instead that the structure of hedge fund incentive contracts and the competitive nature of the hedge fund industry provide more of an incentive to underreport losses than gains, a pattern we call "conditional smoothing." We show analytically how this type of managerial behavior distorts various aspects of the fund's statistical footprint, generating conditional correlation with a fund's risk exposures and conditional serial correlation. The latter constitutes our statistical screen. While the reporting algorithm we study represents only one possible scheme, it is motivated by the compensation structure and competitive nature of the industry, and is consistent with anecdotal evidence.[3]

In a closely related paper, Getmansky et al. (2004) document substantial positive serial correlation in reported monthly hedge fund returns. The authors examine innocuous explanations for serial correlation, including time varying expected returns, time varying leverage, and marking illiquid assets to market using extrapolation. They also point out that serial correlation can be caused by purposeful managerial smoothing of contemporaneous and lagged asset returns. We strive to identify fraudulent smoothing by applying a more detailed model of managerial behavior. Under the assumptions of our model, the statistical flag of conditional serial correlation is only triggered when managers misreport fund returns.

Our empirical analysis has two parts. First, before testing actual hedge funds for conditional smoothing, we carefully examine the small sample properties of our statistical screen. From a practical perspective, this is an important step. A type I error, which in this context is falsely flagging a fund with no conditional smoothing, may cause regulators to spend limited resources auditing the wrong fund. A type II error, which here is failing to flag a fund in which conditional smoothing is present, may allow a fraudulent fund to avoid an audit. We assess the likelihood of type I and type II errors by using simulated returns calibrated to match the risk exposures of a large database of hedge funds. When returns are generated under the null hypothesis, we reject at the significance level. This result indicates a low probability of a false positive. When returns are generated under the alternative, and the underlying factors are revealed to the econometrician, we reject the null 80% of the time using 120-month histories. For 36-month histories, power drops to 33%. And when the econometrician must infer the underlying factors, power drops a modest amount at all history lengths. This result suggests that the screen will fail to identify a substantial fraction of the conditional smoothing cases it is designed to detect. However, the statistical power may be high enough to serve as a deterrent.

Second, we analyze the returns of actual hedge funds. Consistent with the results of Getmansky et al. (2004), funds that invest in the most liquid assets, such as CTAs and long-only funds, are generally not serially correlated. In contrast, almost 40% of the Event-Distressed and Global Emerging funds feature statistically significant positive unconditional serial correlation. When applying our screen, we find that only 5% of the funds feature *conditional* serial correlation, consistent with the low frequency of reported fraud cases. In cross-sectional analysis, the risk of capital flight, as measured by the magnitude and volatility of investor cash flows, is the most significant and robust predictor of observing conditional serial correlation.

Our results suggest that quantitative filters with reasonable statistical power can be developed. We discuss several important caveats. First, only a subset of the actual fraud cases pursued by the SEC is related to patterns of reporting that a returns-based screen might detect. Second, fraudulent return smoothing may become less likely if more funds delegate valuations and external reporting to a prime broker or auditor. Third, the presence of conditional serial correlation is only indicative of fraudulent return smoothing. We do show that conditional serial correlation is almost never found in samples of hedge fund indices and mutual funds, indicating that it is not a typical statistical property of asset returns. This evidence aside, statistical screens of the type developed in this paper should not be interpreted as tests for fraud, but rather as a means for selecting funds for more in-depth scrutiny.

The rest of the paper is organized as follows. Section II discusses fraudulent return smoothing in practice to motivate our statistical screen. Section III reviews related academic studies. Section IV develops an econometric model of conditional smoothing of hedge fund returns and a corresponding screen for fraud. Section V describes the data. Section VI presents our empirical analyses and interprets the results. Section VII offers concluding remarks.

## II. Fraudulent return smoothing in practice

This section examines actual cases of fraudulent reporting to motivate our assumption that some managers are more desirous of smoothing losses than gains.

### *A. National Australia Bank*

The fraud perpetrated by the currency options trading desk at National Australia Bank (NAB) provides a detailed example of the mechanics of fraudulent return smoothing.[4] Though NAB is not a hedge fund, its experience is useful because it illustrates some important features of hedge fund failures and is publicly documented, providing an in-depth look at the impact of incentives on reporting behavior.

In January 2004, NAB announced cumulative losses totaling AUD 360 million generated by the currency options trading desk. The traders had discovered a way of entering false transactions into NAB's computerized record-keeping system in order to conceal losses, thereby meeting profit targets to which the traders' compensation was tied. A detailed consulting report prepared by PricewaterhouseCoopers (PwC, 2004) chronicles the sequence of the currency desk's realized monthly profits and losses, and the profits and losses reported to NAB's record-keeping system. These are listed in Table 1. According to the PwC report, the currency desk began October 2002 with a cumulative overstatement of its portfolio of AUD 7,792,000. In October, the desk earned AUD 8,946,000 but reported only AUD 974,000 in order to eliminate the existing cumulative overstatement. The desk underreported income in three other months (March, July, and October 2003) in order to reduce existing cumulative overstatements. In November and December 2002, the desk fully reported its profits. In the remaining nine months in Table 1, the desk overreported profits. In five of these months, the desk reported positive profits when in fact the desk realized losses. In summary, the currency traders often underreported losses but only underreported profits when there was an existing overstatement of the desk's portfolio. This pattern of asymmetric misreporting motivates the conditional smoothing algorithm that we use to construct our screen.

*B. SEC enforcement actions*

To relate fraudulent return smoothing to fraud cases recently pursued by the SEC, we collect information on hedge fund fraud from SEC litigation releases. We find 53 enforcement actions and place them into six categories: "Misappropriation," "Misrepresent returns," "Misrepresent strategy," "Fraudulent offerings," "Ponzi schemes," and "Other." We place funds in the "Other" category if they do not fit any of the first five groups, including, for example, illegal short sale and market timing activities. Table 2 lists the number of fraud cases which fall into each category. Two lists are created. The first places each fund in one category that best captures the primary reason for the charge of fraud. When misappropriation is mentioned, it is always listed as the primary type of fraud. The second places each fund in all categories that are related to the fund's case file. Fraudulent return smoothing would fit in the category of misrepresenting returns, which is the primary type of fraud in at least 15 of the 53 cases and is present in at least 34 of them.

The litigation reports generally describe extreme behavior, such as outright stealing of fund assets or reporting wildly inflated asset values, which often cause the funds to collapse. The return smoothing we describe does not necessarily endanger investors to the same degree. However, three aspects of the fraud cases in Table 2 are salient. First, misrepresenting returns often occurs in conjunction with misappropriation and fraudulent offerings, indicating that screening for a potentially benign pattern in returns may identify other types of fraud. Second, several cases detail misrepresentation over extended periods of time, sometimes using forged reports from clearing brokers and auditors. This indicates that some managers would have the ability to engage in long-term fraudulent return smoothing, even with some level of external monitoring. Third, for the 34 cases related to misrepresenting returns, the SEC's litigation releases never mention undervaluation of assets.[5] This provides some justification for our assumption that fund managers have more of an incentive to smooth on the downside than the upside.

## III. Related literature

Prior literature studies the impact of managerial incentives on performance and risk-taking in the mutual fund and hedge fund industries. Brown et al. (1996) find that mutual funds with relatively poor returns over the first six months of a calendar year tend to feature increases in risk over the remainder of the year. They argue that in the presence of asset-based management fees, managers of poorly performing funds have an incentive to increase volatility. By doing so, they increase the probability of high returns in the remainder of the year, which would subsequently attract new investor capital. Chevalier and Ellison (1997) and Koski and Pontiff (1999) find similar results, though Busse (2001) argues that the results are not attributable to managerial activity but rather statistical properties of daily stock returns.

Since the incentive structure of a typical hedge fund rewards managers with large performance-based fees, we might expect an even greater set of interrelations between capital inflows, fund performance, and managerial incentives in the hedge fund industry. Agarwal et al. (2003) find that, as has been documented extensively in the mutual fund literature, hedge funds with relatively high returns in a given year experience larger subsequent capital inflows than other funds. Furthermore, Agarwal et al. construct an annual measure of managerial incentives that incorporates high-water marks and a record of prior capital inflows and show that it is positively related to fund performance in the following year.[6] Agarwal and Naik (2000) find that winners tend to repeat only at the quarterly frequency, however, and Brown et al. (1999) find no evidence of performance persistence at the annual frequency, suggesting that the relation between incentives and returns is short-lived. A question remains: by what mechanism do managers temporarily enhance performance?

Recent reports of fraud in the hedge fund industry suggest that hedge fund managers may be tempted to "manage" reported returns.[7] Two academic studies examine this issue. Asness et al. (2001) note that illiquid assets held by hedge funds can lead to changes in hedge fund values that are non-synchronous with changes in common benchmarks. If reported hedge fund values are stale, traditional estimates of volatility and correlation with benchmarks can be biased downwards, thereby improving the risk-adjusted performance of the funds. The authors regress hedge fund index returns on contemporaneous and lagged observations of the S&P 500 return, analogous to the techniques employed by Scholes and Williams (1977) and Dimson (1979), and find that the lagged factor can fully explain abnormal returns in the hedge fund indices.

In an effort to distinguish unintentional stale pricing from managerial smoothing, Asness et al. also separate observations of contemporaneous and lagged benchmark returns by sign. When the aggregate hedge fund index is regressed on the S&P 500, the coefficient on lagged negative returns is 0.79, whereas the coefficient on lagged positive returns is only 0.17, suggesting that more smoothing occurs when benchmark returns are low, consistent with discretionary managerial valuation. In addition, actual exposure to the S&P 500, as measured by the sum of coefficients on contemporaneous and lagged returns, is 1.38 during down markets and only 0.44 during up markets, suggesting that true factor exposures may change over time. This pattern of market timing is perverse, however, because it describes a manager that consistently changes exposure to risk in the wrong direction. Jagannathan and Korajczyk (1986) find that asymmetric coefficients such as these can occur when asset returns are more or less option-like than the chosen factors. Hedge fund return distributions feature option-like properties, as shown in Fung and Hsieh (1997), hence the Asness et al. results may be a consequence of their factor selection.

Getmansky et al. (2004) examine the econometric properties of reported hedge fund returns by specifying a linear factor model for asset returns and a moving average algorithm that transforms asset returns to reported returns. Getmansky et al. show that their model can generate the non-synchronous relation studied by Asness et al. (2001), and captures other time series properties of hedge fund returns. Furthermore, by focusing on the univariate distribution of hedge fund returns, the Getmansky et al. approach avoids the issue of factor selection. The authors argue that their algorithm can represent both valid attempts to mark-to-market fund assets, and deliberate efforts by managers to smooth reported returns and so improve reported risk-adjusted performance. We introduce in the next section a model of conditional smoothing to distinguish between these two activities.

## IV. An econometric model of conditional smoothing

In Subsection A, we review the Getmansky et al. (2004) framework. In Subsection B, we present our model of conditional smoothing and the corresponding screen for conditional serial correlation. In Subsection C, we discuss implementation issues.

### *A. Unconditional smoothing*

We denote by $R_t$ the return of a hedge fund's assets in period $t$ and assume $R_t$ satisfies the following linear single-factor model:

$$(1) \qquad R_t = \mu + \beta\Lambda_t + \varepsilon_t, \quad E\left[\Lambda_t\right] = E\left[\varepsilon_t\right] = 0, \quad \Lambda_t, \varepsilon_t \sim \text{independent}, \quad \mathrm{Var}\left[R_t\right] = \sigma^2$$

All of the results presented here hold for linear multi-factor models as well. Factor exposures are assumed to be constant. As mentioned in Section III, Asness et al. (2001) find evidence of time-varying exposure to factor risk, though the perverse pattern of coefficients in up and down markets is hard to interpret. Fung and Hsieh (1997) also present evidence of time-varying exposure to factor risk using nonparametric techniques. Time-variation in exposure may be appropriate for some funds, and we accommodate it by using factors that reflect the time series behavior of dynamic trading strategies as discussed in Subsection C. We could incorporate time-varying exposures more explicitly by allowing $\beta$ to be a function of time or instrumental variables, and we leave this worthwhile extension for future research.

The return of a hedge fund's assets is known by the fund's manager. The reported return of the hedge fund, that which is observable by the econometrician, is denoted by $R_t^O$. Getmansky et al. (2004) assume that a fund manager reports a weighted average of the contemporaneous asset return and $k$ lags as follows:

$$
(2) \qquad \begin{aligned} R_t^O &= \sum_{j=0}^{k} \theta_j R_{t-j} \quad \text{where} \\ \theta_j &\in [0,1] \quad \text{for} \quad j = 0,\ldots,k \\ 1 &= \theta_0 + \theta_1 + \ldots + \theta_k \end{aligned}
$$

When the smoothing coefficients $\theta$ sum to one, as in (2), asset returns are eventually fully reflected in observed returns.

The specification (2) distorts several statistical properties of fund returns. The expected observed return, the variance of observed returns, and the covariance between current and lagged observed returns are equal to:

$$
(3) \qquad \begin{aligned} \mathrm{E}\left[R_t^O\right] &= \mu \\ \mathrm{Var}\left[R_t^O\right] &= \sigma^2 \sum_{j=0}^{k} \theta_j^2 \\ \mathrm{Cov}\left[R_t^O, R_{t-m}^O\right] &= \begin{cases} \sigma^2 \sum_{j=0}^{k-m} \theta_j \theta_{j+m} & \text{if} \quad 0 \le m \le k \\ 0 & \text{if} \quad m > k \end{cases} \end{aligned}
$$

Note that (3) implies that the variance of observed returns is lower than the variance of asset returns. In addition, and in advance of our screen for intentional smoothing, if we regress observed returns on their $m^{\text{th}}$ lag, the resulting slope coefficient is equal to the corresponding covariance in (3) divided by the variance of observed returns:

$$
(4) \qquad \begin{aligned} R_t^O &= a + b_m R_{t-m}^O + \eta_t \\ b_m &= \frac{\sum_{j=0}^{k-m} \theta_j \theta_{j+m}}{\sum_{j=0}^{k} \theta_j^2} \quad \text{for} \quad 1 \le m \le k \end{aligned}
$$

Equation (4) shows that smoothing the asset returns generates serial correlation in observed returns. Consider expressing the observed return by substituting the factor model (1) into the right-hand side of (2), i.e.

$$
(5a) \qquad R_t^O = \theta_0 \left(\mu + \beta \Lambda_t + \varepsilon_t\right) + \theta_1 \left(\mu + \beta \Lambda_{t-1} + \varepsilon_{t-1}\right) + \ldots + \theta_k \left(\mu + \beta \Lambda_{t-k} + \varepsilon_{t-k}\right)
$$

Now, suppose the identity of the factor is known, and consider regressing observed returns on contemporaneous and $k$ lagged values of the factor, i.e.

$$
(5b) \qquad R_t^O = \mu + \beta_0^O \Lambda_t + \beta_1^O \Lambda_{t-1} + \ldots + \beta_k^O \Lambda_{t-k} + \gamma_t
$$

Comparing (5a) to (5b) shows that smoothing reduces the measured exposure to the contemporaneous factor and induces exposures to the lagged factors, i.e. $\beta_j^O = \theta_j \beta$ for $0 \le j \le k$ .

The observed returns generated by (2) can reflect both conservatism when marking to market and intentionally dampening the observed return process to lower the fund's apparent risk. One must be able to distinguish between the two sources of smoothed returns to identify fraudulent reporting. In the unconditional model, however, the two sources of smoothed returns are observationally equivalent.

*B. Conditional smoothing*

The smoothing algorithm in equation (2) is unconditional in the sense that a fraction $\theta_0$ of the asset return is reported contemporaneously, with the remainder reflected in future fund returns, regardless of the value of the asset return. We conjecture that competition in the hedge fund industry and the standard compensation scheme for hedge fund managers provide an incentive for more complex behavior. The algorithm with which a manager converts asset returns to reported returns is unobservable; hence the only way to proceed is to employ a well-motivated specification as a proxy.

Managers have an incentive to affect the shape of the reported return distribution in order to make it more attractive to investors. During periods of large positive returns, managers likely fully report fund returns for fear of lagging competitors. During periods of large negative returns, managers may only partially report fund returns to mitigate capital flight. Though we focus exclusively on this behavioral pattern, other algorithms could be accommodated in our framework.

To illustrate, consider a simple decision rule that compares the asset return of the fund $R_t$ to a constant *c*. If the return falls below this critical level, then the manager reports some fraction $\theta_0$ of the return in the current period and reports the remainder over the following *k* periods using a set of weights $\theta_j$ for $j = 1, ..., k$ . If the return exceeds *c*, then the manager uses a different set of weights $\psi$ . We augment the smoothing algorithm (2) to include indicator variables which capture the conditional smoothing:

$$
\text{(6)} \quad \begin{aligned}
&R_t^O = \sum_{j=0}^{k} \left(\theta_j \left(1 - I_{t-j}\right) + \psi_j I_{t-j}\right) R_{t-j} \quad \text{where} \\
&I_{t-j} = 1 \quad \text{if} \quad R_{t-j} \geq c \quad \text{for} \quad j = 0, ..., k \\
&I_{t-j} = 0 \quad \text{if} \quad R_{t-j} < c \quad \text{for} \quad j = 0, ..., k
\end{aligned}
$$

We assume that both sets of smoothing coefficients, $\theta$ and $\psi$, sum to one, again ensuring that asset returns are eventually fully reflected in reported returns. When $R_t \geq c$ , we expect a larger fraction of the asset return to be reported contemporaneously, which implies that $\psi_0 > \theta_0$ , since managers have the incentive to reveal good performance when it occurs and to defer poor performance.[8] Naturally, this implies that a smaller fraction of a period's asset return will be reflected in future observed fund returns, i.e. $\sum_{j=1}^{k} \psi_j < \sum_{j=1}^{k} \theta_j$ . At the extreme, a manager could fully report in positive states, in which case $\psi_0 = 1$. By construction, our framework nests the managerial behavior suggested by Chandar and Bricker (2002) and Getmansky et al. (2004) in which managers underreport high positive returns in order to offset losses in other periods.

The conditional smoothing algorithm has specific implications for the time series properties of observed hedge fund returns. As with the unconditional case, reported returns have the same unconditional expected value as the underlying asset returns. The smoothing framework simply determines the timing of the revelation of asset returns. More importantly, it results in serial correlation in hedge fund returns that is conditional on the magnitude of asset returns. Consider regressing fund returns on a single lag, but incorporating an indicator variable for the lagged return as follows:

$$
\text{(7)} \quad R_t^O = a + \left(b_1^- \left(1 - I_{t-1}\right) + b_1^+ I_{t-1}\right) R_{t-1}^O + \eta_t
$$

where $I_{t-1} = 1$ if $R_{t-1} \geq c$ and zero otherwise. If a manager tends to defer reporting poor returns, then the relation between contemporaneous and lagged returns will be larger when the lagged returns are poor, i.e. $b_1^- > b_1^+$ . Figure 1 illustrates the difference between analytic values of the two coefficients $b_1^-$ and $b_1^+$ over a range of inputs, and shows that this is indeed the case.[9] This result motivates the following empirical prediction:

**Proposition 1** *If asset returns are generated by* $R_t = \mu + \beta\Lambda_t + \varepsilon_t$ *and observed returns are constructed as* $R_t^O = \left(\theta_0\left(1-I_t\right)+\psi_0 I_t\right)R_t + \left(\theta_1\left(1-I_{t-1}\right)+\psi_1 I_{t-1}\right)R_{t-1}$, *where* $I_{t-j}=1$ *if* $R_{t-j} \geq c$ for $j=0,1$ *and zero otherwise, then observed returns will display conditional serial correlation if* $\theta_1 \neq \psi_1$. *Conditional serial correlation can be detected by estimating parameters of* $R_t^O = a + \left(b_1^-\left(1-I_{t-1}\right)+b_1^+ I_{t-1}\right)R_{t-1}^O + \eta_t$, *and will result in* $b_1^- \neq b_1^+$.[10]

Conditional serial correlation is important because, under the assumptions stated above, it is a direct consequence of fraudulent reporting. Other reporting schemes may generate different time series patterns and would not be captured by a screen for conditional serial correlation. To address this issue of power, a collection of screens could be developed, each of which identifies patterns consistent with a specific reporting algorithm. What about the probability of selecting the wrong funds for examination? Perhaps there are innocuous explanations for the time series pattern identified by a specific screen. Getmansky et al. (2004), for example, discuss at length a variety of possible causes for unconditional serial correlation. In the context of our model, if liquidity drops when returns are low, then conditional serial correlation may also be a consequence of relying on model values rather than market prices in down states. We address this concern in our empirical analysis by applying our screen to a large number of fund indices, asset-based style factors, and mutual funds, which should be free of reporting distortions. In only very few cases do we find significantly greater serial correlation in down states. In contrast, we do find examples of conditional serial correlation in almost every category of hedge fund, suggesting that it is not solely a liquidity-related phenomenon.

If a manager applies conditional smoothing as in equation (6), then observed fund returns possess conditional exposures to contemporaneous and lagged values of the factor. To see this, substitute the factor model (1) into the right-hand side of (6) with one lag, i.e.

$$(8) \qquad R_t^O = \left(\theta_0\left(1-I_t\right)+\psi_0 I_t\right)\left(\mu+\beta\Lambda_t+\varepsilon_t\right)+\left(\theta_1\left(1-I_{t-1}\right)+\psi_1 I_{t-1}\right)\left(\mu+\beta\Lambda_{t-1}+\varepsilon_{t-1}\right)$$

Now consider the following regression:

$$(9) \qquad R_t^O = \alpha+\left(\beta_0^{O-}\left(1-I_t\right)+\beta_0^{O+} I_t\right)\Lambda_t+\left(\beta_1^{O-}\left(1-I_{t-1}\right)+\beta_1^{O+} I_{t-1}\right)\Lambda_{t-1}+\gamma_t$$

It is straightforward to show that conditional exposures are $\beta_j^{O-}=\theta_j\beta$ and $\beta_j^{O+}=\psi_j\beta$ for $j=0,1$. Thus, when true exposure is constant over time, a manager who smoothes conditionally will report fund returns that exhibit higher contemporaneous exposure to the factor when asset returns are high than when they are low.[11] Also, the exposure to the lagged factor would appear to be higher when lagged asset returns are lower. This formalizes the test for conditional smoothing in Asness et al. (2001).

### *C. Empirical implementation*

The conditional smoothing algorithm is a function of the fund's asset return, which is unobservable given the opacity of hedge fund holdings. To proceed, we assume that the manager compares the systematic component of the asset return, $\mu+\beta\Lambda_t$, to a threshold *c*. In the spirit of Chandar and Bricker (2002), the systematic component of a hedge fund's asset return can be interpreted as the return of liquid assets held by the portfolio, for which little discretionary valuation is possible. Thus, we redefine the indicator variable as $I_{t-j}=1$ if $\mu+\beta\Lambda_{t-j}\geq c$ for $j=0,1$ and $I_{t-j}=0$ otherwise. Proposition 1 is still valid under this alternative indicator variable.[12] The identity of the factor or factors $\Lambda$ is unknown and must be inferred from observed returns. Asness et al. (2001) face the same problem, and solve it by defining the indicator variable to equal one if the S&P 500 return is above its mean. While this indicator variable may be correlated with some managers' decision rules, for example when a fund follows a long-only equity strategy, it is likely to result in tests with low power for many funds. Further, as mentioned previously, Jagannathan and Korajczyk (1986) find that asymmetric factor exposures can occur when asset returns are more or less option-like than the chosen factors. Hedge fund return distributions feature option-like properties, as shown in Fung and Hsieh (1997), hence conditional factor exposures could be generated even when hedge fund returns are reported accurately. We address this issue by including factors constructed directly from option prices.

Estimating a factor model for hedge funds is challenging, given the multiple asset classes and dynamic trading strategies available to hedge fund managers. Two approaches are commonly used. Both approaches incorporate non-linear payoffs to capture the option-like feature of hedge fund returns. The first approach is to use a collection of hedge fund indices, each of which represents a particular strategy. See, for example, Agarwal and Naik (2004). As discussed by Fung and Hsieh (2004), the disadvantage of using hedge fund indices is that index returns are likely distorted by inaccuracies in constituent hedge fund returns, as well as biases in the databases from which the indices are constructed, including survivor-ship bias, back-fill bias, and selection bias. The advantage of using the indices, however, is that the constituent hedge fund returns capture the dynamic and nonlinear strategies employed by fund managers.

The second approach, pioneered by Fung and Hsieh (2001, 2002, 2004), is to use style factors constructed from market prices of assets chosen to represent different strategies. For example, managers of trend-following funds employing technical trading rules generate returns that are explained well by “look-back” options, the returns of which can be computed from observed market prices of vanilla options. The advantage of this approach is that the returns are free of the distortions described above, since they are computed directly from observed market prices. The disadvantage is that the approach allows for only those strategies specifically modeled by the econometrician, and does not permit time-variation in strategies.[13]

For robustness, we use a list of hedge fund indices, as well as Fung and Hsieh’s (2004) seven asset-based style factors, when estimating the systematic component of hedge fund returns. In both cases, we select for a given fund the subset of those factors whose contemporaneous and lagged observations maximize the explanatory power of the regression in equation (5b). Note that (5b) represents the relation between a fund’s observed returns and the underlying factors under the assumption that the smoothing is unconditional. It is straightforward to show that the resulting factor loadings at each lag are an average of the conditional factor loadings, so that the sum of the coefficient estimates from (5b) measure $\beta$, the true exposure of a fund's asset returns to the selected factors.

**V. Data**

We use the Center for International Securities and Derivatives Markets (CISDM) hedge fund database, maintained by the University of Massachusetts in cooperation with Managed Account Reports LLC, with data through December 2003. The CISDM database consists of two sets of files, one for live funds and one for dead funds. We include the dead funds in our analysis. Each set consists of a performance file, containing monthly observations of returns, total net assets, and net asset values, and a fund information file, containing fund name, strategy type, management fees, and other supplementary details. We discard funds with less than 24 months of returns to ensure some degree of accuracy in our empirical analysis. To provide contrast, we also examine the managed futures contained in the CISDM database. Managed futures invest in futures and forward contracts and are regulated primarily by the Commodity Futures Trading Commission (CFTC). The Commodity Exchange Act, which established the CFTC, also established the disclosure, record keeping, and reporting rules for managed futures. This Act provides several exemptions, the most relevant of which limits investment to qualified investors.

We categorize the CISDM hedge funds and managed futures using the database's two-tier classification scheme using classes and strategies. The hedge funds are separated into six classes, including hedge funds open to non-U.S. investors and U.S. investors ("HF-NON" and "HF-US" respectively), funds of funds open to non-U.S. investors and U.S. investors ("FOF-NON" and "FOF-US" respectively), CISDM hedge fund indices ("HED-IDX"), and the S&P 500 index ("IND-EX"). We group the "HF-NON" and "HF-US" funds together, as well as the "FOF-NON" and "FOF-US" funds. There are four managed futures classes in the database, including Commodity Trading Advisors (CTAs, “FUT-CTA” in the database), CISDM CTA indices (“FUT-IDX”), and public and private pools (“FUT-PUB” and “FUT-PVT” respectively). We keep the pools separate from the CTAs, since the pools generally invest in multiple CTAs, in the same way that funds of funds invest in multiple hedge funds, as described by Anson (2002). Furthermore, we maintain the distinction between public and private pools. Public pools are open to the general public and must register with the SEC, whereas private pools are exempt from registration since they are sold only to qualified investors. Managers of private pools, therefore, may have more discretion when valuing assets and so we may observe differences in the patterns of reported returns.

The hedge funds and CTAs are further separated by strategy. Table 3 lists summary statistics of the hedge funds, CTAs, and managed futures in the CISDM database. For each strategy, the table lists the number of funds and equally-weighted cross-sectional averages of each fund’s monthly average return, standard deviation, Sharpe ratio, skewness, excess kurtosis, and first-order serial correlation. Also listed is the number of funds with serial correlation coefficients significant at the 5% two-sided level, separated by sign. Live funds feature substantially higher Sharpe ratios than dead funds in almost all categories. This is no surprise, as anecdotal evidence suggests that hedge fund investors withdraw capital en masse following periods of poor performance, though Agarwal et al. (2003) report an asymmetric response to performance, i.e. poor performers are not punished to the same extent as good performers are rewarded. For most of the categories, live funds on average are more positively skewed than the corresponding dead funds. All categories feature substantial excess kurtosis. Thick tails in the return distributions of these funds are consistent with option-like payoffs in the strategies they employ, motivating Fung and Hsieh (2001, 2002, 2004) to use baskets of traded options to mimic strategy returns.

Approximately one quarter of the live hedge funds feature statistically significant and positive serial correlation, with an average coefficient of 0.13. The dead funds have similar properties. Average coefficients for categories range from a low of 0.05 for Global Macro and 0.06 for Long Only to a high of 0.19 for Global Emerging and 0.20 for Event-Distressed. These results are similar to those found in Getmansky et al. (2004), who report an average coefficient of 0.12. As mentioned, serial correlation could be caused by illiquidity in fund assets or by intentional managerial smoothing. The cross-sectional variation in correlation coefficients is consistent with the effects of illiquidity, since the categories associated with low liquidity, Event-Distressed and Global Emerging, tend to have higher levels of serial correlation. In contrast to the hedge funds, the CTAs and Managed Futures feature very few instances of significant serial correlation, likely resulting from the high liquidity of the futures contracts in which they invest. The Fund of Funds have an average coefficient of 0.22, which suggests that managers of these vehicles tend to select individual hedge funds with low liquidity, perhaps because they report high Sharpe ratios. Our test for conditional serial correlation attempts to distinguish the effect of illiquidity from conditional smoothing. As shown in Section IV, under the assumptions of our model, only purposeful conditional smoothing can generate conditional serial correlation.

Table 4 lists the $25^{th}$, $50^{th}$, and $75^{th}$ percentiles of the cross-sectional distributions of fund history lengths in the CISDM database, in months. Not surprisingly, the live funds have longer histories. The median live fund in the various hedge fund categories has history length ranging from 54 months for Sector funds to 88 months for Global International. For the dead funds, the medians range from 42 to 61 months. These history lengths are important because even the simple model of managerial behavior studied here will be somewhat data-intensive. As shown by Bollen and Busse (2001), studies of market timing in the mutual fund literature, which typically use monthly returns, tend to lack sufficient power to reject the null hypothesis of no market timing. Similarly, in our model of conditional serial correlation, the relatively short history lengths of hedge funds may limit the power of our screen.

Table 5 contains summary statistics of the CISDM hedge fund and CTA indices that are used as factors to characterize individual fund strategies. There are a handful of other CISDM indices that are discarded due to limited data histories. Also listed in Table 5 are summary statistics for the seven asset-based style factors of Fung and Hsieh (2004), which are used as an alternative set of variables to model hedge fund and CTA returns. The Wilshire Size factor equals the return of the Wilshire Small Cap 1750 minus the return of the Wilshire Large Cap 750, and is obtained from Wilshire. The three trend factors are the returns of three portfolios of options and are obtained from David Hsieh's website.[14] The changes in the 10-year Treasury yield and the credit spread, which is the corporate BAA yield minus the 10-year Treasury yield, are from *Datastream*.

Several statistics stand out. First, note that the first four categories of hedge fund have skewness of about negative 2 and excess kurtosis over 10. These results indicate significant non-normalities. Unreported analysis indicates that these results are highly sensitive to a few outliers, however. Second, all but two of the hedge fund indices feature statistically significant positive serial correlation. In contrast, most of the CTA indices and ABS factors are not serially correlated. Serial correlation in the hedge fund indices can cloud our inference regarding patterns of serial correlation in the individual funds; hence in most of the empirical work described in the next section we replace the indices' time series by the residuals of a regression of the index returns on their lags.

The returns of the hedge fund and CTA indices should be relatively free of distortions caused by fraudulent reporting, since fraud is presumably rare and its effects would be reduced through diversification in the construction of the indices. Since the ABS factors are constructed directly from market prices of liquid assets, their returns should also be free of reporting distortions. Thus, we can gain some confidence that our screen based on conditional serial correlation is not driven by the statistical properties of underlying assets by applying it to the indices and factors. Table 6 reports results of regressions of the form:

(10) $$R_t = a + b_1^+ R_{t-1} + b_1^- \left(1 - I_{t-1}\right) R_{t-1} + \eta_t$$

where $I_{t-1} = 1$ if the return in month $t-1$ is greater than its mean and zero otherwise. This specification is slightly different from equation (7). Here $b_1^-$ tests whether serial correlation is different when lagged returns are below a particular level. Two results are important. First, eight of ten hedge fund indices have a statistically significant positive serial correlation coefficient, $b_1^+$, as expected. Second, none of the indices or factors features significantly greater serial correlation in down states at a 5% two-sided level. Of the series with positive $b_1^-$ estimates, the Wilshire Size factor has the lowest *p*-value of 12.24%. In contrast, as described in the next section, there are individual funds from a wide variety of categories that display significantly higher serial correlation in down states.

We compare the frequency with which conditional serial correlation appears in our sample of individual hedge funds to the corresponding frequency in a sample of individual mutual funds. Since mutual fund managers have much less discretion over fund valuation and returns, we should see far fewer instances of conditional smoothing in mutual funds. We obtain mutual fund return data from the Center for Research in Security Prices (CRSP) Survivor-Bias Free U.S. Mutual Fund Database. To be included, a mutual fund must have at least five years of return data in the same 1994 to 2003 period that brackets the hedge fund data. Our sample includes 5,458 funds in six equity categories. We use Carhart's (1997) four factor model to estimate each fund's systematic return. We use the CRSP value-weighted equity index to proxy for the market factor, we use the 90-day U.S. Treasury Bill Discount from *Datastream* (code TBILL90) to proxy for the risk-free rate, and we obtain the size, value, and momentum factors from Kenneth French's website.[15]

## VI. Empirical analysis

This section describes our empirical analysis and interprets the results. Before running our test for conditional serial correlation on individual hedge funds, we gauge its small sample properties. This step is critical in evaluating the ability of regulators and institutional investors to use quantitative filters to select funds for in-depth examinations. If our test has a high probability of detecting conditional smoothing when it is absent, then this would result in needless audits and waste valuable resources. If our test has a low probability of detecting conditional smoothing when it is present, then this would allow fraudulent activity to continue. Subsections A and B describe our simulations and report the size and power associated with our screen for conditional serial correlation. We then turn to an analysis of individual hedge funds. Subsection C lists the actual frequency with which our screen identifies conditional serial correlation. Subsection D examines fund characteristics that are related to the probability of observing conditional serial correlation.

*A. Size*

We simulate hedge fund returns under the null hypothesis of unconditional smoothing to assess the size of our test. Our simulated data are calibrated to match the risk exposures of actual hedge funds. For each fund, we determine which single factor best describes the fund's observed returns using the regression $R_t^O = \alpha + \beta^O \Lambda_t + \varepsilon_t$, where $R^O$ denotes observed returns. We use the coefficients and the residual volatility, $\sigma_\varepsilon$, from this regression to simulate 20 sets of fund asset returns $R_t^A$. To simulate a series of length *n*, we draw *n* monthly observations with replacement from the selected factor's ten-year history, and generate *n* standard normal variates. We scale the factor return by $\beta^O$ and add $\alpha$ to construct the systematic return, and then add the normal variate scaled by $\sigma_\varepsilon$ to generate $R_t^A$. Then, we simulate unconditionally smoothed fund returns $R_t^S$ as follows:

(11) $$R_t^S = 0.5R_t^A + 0.5R_{t-1}^A.$$

Now, acting as the econometrician, we estimate conditional serial correlation in the simulated fund return by running the regression:

(12) $$R_t^S = a + b_1^+ R_{t-1}^S + b_1^- \left(1 - I_{t-1}\right) R_{t-1}^S + \eta_t$$

where $I_{t-1} = 1$ if the systematic component of the simulated return in month $t-1$ is greater than its mean and zero otherwise. To estimate (12), the econometrician needs to construct *I*. We run the simulation two ways. We assume the econometrician knows the systematic components from which returns are generated, and, more realistically, we assume he does not, and so must reconstruct them by determining which single factor best fits the simulated return history.

Table 7 reports the frequency with which the simulated returns produce a significant positive $b_1^-$ at a 5%, two-sided level for time series of lengths 120, 60 and 36 months. For virtually all fund types and at all history lengths the rejection rate is about 2%, indicating that the probability of falsely rejecting the null is in line with the significance level of the test.

*B. Power*

We gauge the power of our screen by generating random hedge fund returns that conform to the conditional smoothing algorithm specified in Section IV, and then determining the frequency with which our test identifies the resulting conditional serial correlation. We conduct the power analysis two ways. The first is under a controlled set of conditions; the second is under the actual conditions we face in the database.

For the controlled conditions, we generate asset returns using a single-factor model using the same procedure as described above for the size analysis. Then, we simulate conditionally smoothed fund returns $R_t^S$ as follows:

$$(13) \qquad R_t^S = \left(0.5\left(1 - I_t\right) + I_t\right) R_t^A + 0.5\left(1 - I_{t-1}\right) R_{t-1}^A$$

where $I_t = 1$ if the simulated systematic return in month $t$ is above its mean for the simulation and zero otherwise. As before, we run the simulation two ways, first assuming the econometrician knows $I$, and second requiring the econometrician to infer $I$ from the simulated data. In the latter case, we assume the econometrician knows that the manager is comparing the systematic return to its mean, but needs to estimate it from the data using a single-factor model. There are other interesting variations we could employ here: we could randomize the smoothing coefficients or we could set the smoothing coefficients equal to a function of the lagged systematic return, for example. In these cases, the power of our test will likely be lower than reported.

Table 8 shows the results for time series of lengths 120, 60 and 36 months. When the econometrician knows $I$, the power of the test is about 80% for a 120-month history. For a 60-month history, which is roughly what to expect with the funds in our database, the power drops to about 50%. For a 36-month history, it drops to about 33%. When the econometrician does not know $I$, the power is slightly lower. These results indicate that the power of our test is reasonable for funds with history lengths at or above the median.

To determine power under actual conditions, we use multi-factor models to generate asset returns, and use the actual history lengths of the funds in our sample. We regress observed returns of each hedge fund and fund of funds on contemporaneous and lagged returns of the CISDM hedge fund indices, and each CTA and managed futures on returns of the CISDM CTA indices, to determine which subset of the available proxies for the factors $\Lambda$ best captures the time-variation in a fund's returns. We also include the S&P 500 return as a factor. If managers smooth returns, residuals in the factor regressions will likely be serially correlated themselves. Serial correlation in residuals does not affect OLS estimates of factor exposures, but does affect their standard errors. Since we use significance levels to select the optimal subset of factors, we use feasible generalized least squares with one lag to account for serial correlation in residuals.[16]

Since we have a relatively large set of indices, and have little guidance regarding the number of lags to include, we start with contemporaneous observations of the indices. We use regression diagnostics to determine the relevant subset of factors for each fund. In particular, we use the IMSL subroutine RBEST to identify the subset of indices which maximize the adjusted R-squared of the regression. In order to avoid over-fitting, we determine the smallest subset for which the adjusted R-squared cannot be improved in a statistically significant manner by increasing the subset size. We add lagged factors to the regression only if they result in a significantly higher adjusted R-squared.

Panel A of Table 9 reports details of this procedure. Listed for each fund type is the number of funds, then the 25$^{th}$, 50$^{th}$, and 75$^{th}$ percentile of the cross-sectional distribution of adjusted R-squared when the regressions are limited to one contemporaneous factor, two contemporaneous factors, and an unlimited number of contemporaneous and lagged factors. For the three fund types, the distribution of adjusted R-squared shifts modestly to the right as the number of factors increases, though the levels are still quite low. When the number of factors is unconstrained, for example, the median adjusted R-squared is 45% for hedge funds and 31% for CTAs. This result indicates that a substantial portion of hedge fund portfolios are manager-specific and not captured by the indices.[17] Fung and Hsieh (1997), for comparison, find a median of about 25% in an earlier sample of 409 hedge funds and CTAs. Listed next is the average adjusted R-squared of the unconstrained regressions and the average number of factors and lagged factors included. The hedge funds on average have about three contemporaneous factors and one lagged factor, compared to 2.31 and 0.16, respectively, for CTAs, and 2.53 and 0.25, respectively, for managed futures.

Panel B of Table 9 reports the same regression details when the asset-based style factors are used instead of the CISDM indices. In all cases the distribution of adjusted R-squared is shifted substantially to the left. This result indicates that the set of factors is incomplete, as they do not capture much of the variation in fund returns. Also, note that the asset-based style factors represent a constant exposure to the underlying variables, whereas the CISDM indices reflect the time-varying strategy or set of strategies employed by the managers of the constituent funds.

The last step in computing estimates of the systematic component of fund returns is to compute a sum-product of factor returns and a fund's exposures, where the exposures equal the sum of contemporaneous and lagged coefficients. If a manager is smoothing unconditionally, then as shown in equation (5b) and the associated discussion, the true exposure to the factor equals the sum of the exposures of observed returns on contemporaneous and lagged factors, i.e.

$$\beta = \sum_{j=0}^{k} \beta_j^O \tag{14}$$

If a manager is smoothing conditionally, then the observed factor loadings $\beta^O$ are an average of the conditional factor loadings, and (14) is still a valid procedure for inferring the exposure of hedge fund asset returns to the selected factors.

We simulate 20 sets of asset returns for each hedge fund by reordering the residuals from the optimal regression, without replacement, thereby maintaining the actual history length for each fund. From here, we proceed in the same way as with controlled conditions. We construct the conditionally smoothed returns using (12), and then ask the econometrician to estimate the smoothing parameters. The parameters chosen in (12) correspond to fully reporting asset returns when the non-discretionary component of fund returns is above its mean, and averaging evenly over successive months otherwise.

Table 10 shows the results. Listed are the number of funds in each category, and the percentage of simulations which feature conditional serial correlation. At a 5%, two-sided significance level, the percentage of simulations with significant positive $b_1^-$

coefficients is about 33%. This result indicates that the relatively short history length of the funds in our sample, combined with the uncertainty regarding fund holdings and strategy, places some limits on our ability to infer managerial behavior. Note, though, that if the significance levels are relaxed, the power of the test improves markedly, to over 50% at a 20% two-sided significance level. In practice, the choice of significance level is a decision analysis problem, in which the costs of additional false positives, which in this case take the form of needless SEC investigations, are weighed against the benefit of reducing the number of fraud cases missed. The threshold for these power tests may be different than a standard econometric exercise. So long as the probability of identifying fraud is sufficiently high to deter fraudulent behavior, the test will serve its primary purpose.

*C. Actual rejection rates*

To see how many actual funds exhibit conditional serial correlation, we run the regression:

(15) $$R_t^O = a + b_1^+ R_{t-1}^O + b_1^- \left(1 - I_{t-1}\right) R_{t-1}^O + \eta_t$$

where $I_{t-1} = 1$ if the estimated systematic component of observed returns in month $t-1$ is greater than its mean and zero otherwise. The indicator variable is constructed using the procedure described in the power test under actual conditions.

Table 11 lists for each category the number of funds, and the number of funds that feature $b_1^+$ and $b_1^-$ coefficients significant at the 5% two-sided level when CISDM or ABS factors are used to infer the indicator variable. As predicted by the first-order serial correlation statistics reported in Table 3, many funds feature a significant positive $b_1^+$ coefficient. The Event-Distressed, Market Neutral, and Fund of Funds categories contain the highest number of funds with positive serial correlation. Illiquidity in the underlying assets, especially for the Event-Distressed funds, is a plausible explanation for this result. The number of funds that feature a significant positive $b_1^-$ coefficient is quite low, 178 out of 3,689 for the CISDM factors and 140 for the ABS factors, 4.83% and 3.80%, respectively. This low percentage is consistent with the low number of reported fraud cases, but above the 2.5% expected under the null hypothesis. Note that there are significant coefficients across the categories, indicating that properties of specific asset classes cannot be the sole explanation for conditional smoothing. Note also that for the ABS factors, there are actually more funds that have a significant negative $b_1^-$ coefficient than a significant positive one. For these funds, serial correlation is lower when returns are low, perhaps because asset sales create a breakpoint in any extrapolation from historical prices or marking the asset values to a model.

To examine the strength of the link between conditional serial correlation and the risk of fraud, we execute our filter on a subset of the SEC's 53 hedge fund fraud cases listed in Table 2. We study the 18 cases which appear in the CISDM database with at least 24 consecutive monthly observations prior to 2004. Of these, five feature statistically significantly higher serial correlation when their systematic returns are below average, i.e. significant positive $b_1^-$ coefficients. In other words, our screen identifies 28% of the fraud cases recently pursued by the SEC. While this number may seem low, note that it is substantially higher than the percentage of all funds that feature conditional serial correlation. Furthermore, additional screens could be developed, each of which focuses on a specific type of reporting irregularity, in order to increase the probability that fraudulent reporting is identified.

Table 12 reports the frequency of observing conditional serial correlation in a sample of equity mutual funds for comparison. For each fund with at least five years of returns in the same 1994 to 2003 period encompassing the hedge fund data, equation (15) is estimated. For the mutual funds, the indicator variable equals one in month $t-1$ if the systematic return that month, computed from Carhart's (1997) four factor model, is greater than its mean and zero otherwise. The number of funds that feature a positive $b_1^-$ coefficient significant at the 5% two-sided level is close to zero, only 0.92% of the sample. This result indicates that the frequency of observing conditional serial correlation in hedge funds, though quite low, is higher than expected using mutual fund returns as a benchmark.

*D. Determinants of conditional serial correlation*

We examine the cross-sectional properties of the flagged funds in our sample using logit regressions in which the dependent variable takes the value of one when a given fund is flagged and zero otherwise. The purpose of this analysis is to determine whether any fund characteristics are systematically related to the probability of observing conditional serial correlation.

Existing evidence regarding investor behavior suggests cross-sectional predictions of the likelihood that a fund manager is manipulating reported returns.[18] For example, investors in a young fund, or in a fund managed by an adviser with a limited track record, may be more sensitive to performance than other investors.[19] In the spirit of Berk and Green (2004), it would be rational for investors with more diffuse prior beliefs of managerial ability to respond more quickly to performance. Investors with short or unpredictable investment horizons are undesirable from a hedge fund manager's perspective, as they may limit her ability to exploit arbitrage opportunities or other trading strategies, as shown by Shliefer and Vishny (1997) and Liu and Longstaff (2004). Fund managers of young funds, or managers with limited track records, therefore, may be more compelled to discourage investor withdrawals than other managers. Indeed, Getmansky et al. (2004) find that funds that are open to new investors feature on average more smoothing than funds that are closed.

We consider a number of different regressors. The first set includes the following fund characteristics reported in the database: fund age, fund size, fee structure, whether the fund has been audited, the number of days investors are required to wait before they can withdraw their funds, and whether the fund is live or dead. The age variable is based on the first date of fund return data as reported in the database and is expressed in months. Fund size is measured by the maximum amount of assets under management during the history of the fund. The fee structure includes the management fee (annual percentage of fund assets) and incentive fee (annual percentage of fund profits above high water mark) separately. The second set of regressors includes the mean, standard deviation, and Sharpe ratio of reported returns. These measures proxy for investors' perceptions of past performance. The third set of regressors in our analyses includes the mean and standard deviation of investor fund flows. The fund flow variables may be the most informative predictors of managerial smoothing. They directly affect the manager's wealth by determining the size of the fund from which managerial fees are calculated. They can also constrain the fund manager's trading strategy by limiting the manager's ability to engage in arbitrage strategies. A fund manager who perceives a risk of fund outflows and/or volatile cash flows would have more of an incentive to smooth reported returns. With the information available to us, we can only measure realized flow variables, and use them as proxies for managerial expectations.

Fund flow is inferred from the time series of fund returns and assets under management reported in the database. Let $TNA_{i,t}$ denote the total net assets of a hedge fund at time $t$ and let $R_{i,t}$ denote the holding period return for a hedge fund investor in fund $i$ between times $t$ and $t-1$. Fund flow is estimated as:

$$DF_{i,t} = TNA_{i,t} - TNA_{i,t-1}\left(1 + R_{i,t}\right), \qquad (16)$$

where $DF_{i,t}$ denotes dollar flow. Since fund total net assets are often reported at irregular frequencies, we standardize $DF$ by dividing by the number of months over which it is computed. We then transform monthly dollar flows to percentage flows by dividing by $TNA_{i,t-1}$.

Variables such as fund size, wait period, and the volatility of cash flows are highly skewed. In order to improve the distributional properties of these measures, we apply a logarithmic transformation. All results are evaluated using robust Huber-White standard errors.

Table 13 lists the results for rejections when the CISDM and ABS factors are used in implementing the test for conditional serial correlation. Model I includes only the cash flow variables. Here, 40 funds are dropped from the analysis for lack of sufficient data. For this subset, 177 of the 3,649 funds feature conditional serial correlation using the CISDM indices, versus 138 using the ABS factors. In both cases, log cash flow volatility is significant and positive, and mean cash flow is negative, though it is not significant when using the ABS factors. These results are consistent with the intuition that managers facing a higher risk of capital flight are more likely to manipulate reported returns.[20]

Model II includes eight other explanatory variables. The sample size drops substantially to 2,058 though the rejection rate is the same in this subset as in the full sample. The cash flow variables remain important determinants in the presence of the additional explanatory variables. The log volatility of cash flows is positive and statistically significant in both cases; the mean cash flow is negative and statistically significant in both cases. The wait variable is positive and significant for both sets of factors. This result is consistent with Aragon (2004), who finds that restrictions on shareholder activity are more common in younger hedge funds, and are positively related to the level of smoothing. One interpretation is that some managers who face trigger-happy investors employ share restrictions and manipulate returns in order to limit capital flight. In addition, the audit variable is negative and significant using the ABS factors, suggesting that audited funds are less likely to exhibit asymmetric serial correlation. To measure the economic significance, we compute the fitted probability of observing conditional serial correlation evaluated at the mean of all variables and compare it to the fitted probability when one of the statistically significant variables is increased by one cross-sectional standard deviation. For the CISDM indices, the fitted probability is 4.98%. A one standard deviation increase in log cash flow volatility increases the probability to 6.40%. A one standard deviation increase in mean cash flow decreases the probability to 3.42%. In comparison, the fitted probability for the ABS factors is 3.48%. A one standard deviation increase in cash flow volatility increases the probability to 4.73% and a one standard deviation increase in mean cash flow decreases the probability to 3.87%. For audited and non-audited funds, the probabilities are 3.23% and 6.35%, respectively. The statistically significant variables are therefore also economically significant.

In our analysis, the flagged sub-sample comprises a small portion of the overall sample, as expected, since fraud is presumably a rare phenomenon. Therefore, for robustness, we estimate a rare events logit model in order to address potential biases introduced by the significantly unbalanced samples.[21] The rare event logit analysis does not change our results, which indicates that our sample size is large enough to obtain reliable coefficient estimates.

## VII. Summary

Recent regulatory action has sparked debate regarding the need for oversight of the hedge fund industry. Critics of the new registration requirement argue that it is unlikely that the SEC would be able to detect and prevent fraud without risk-based quantitative filters to guide their efforts. In addition, the new registration requirement does not provide investors with information on hedge fund investment strategies or holdings, which would allow them to directly monitor the accuracy of reported returns. Both regulators and investors require simple returns-based statistical screens to aid identification of suspicious activity. The SEC could use these screens to select funds for more in-depth examinations of fund operating procedures and asset valuations.

Our goal is to assess whether the returns reported by hedge fund managers contain enough information to infer fraudulent activity. We develop an econometric model of a representative managerial algorithm that converts asset returns to reported returns, and a methodology for inferring the parameters of the algorithm. Through analysis of a standard database of hedge fund returns, we find that the power of our test can reach 80% under controlled conditions, but is typically substantially lower, approximately 30% to 35%.

Our results suggest two directions for future research. First, perhaps there are other econometric approaches that would yield superior power. One possibility is jointly estimating the two stages of our methodology using nonlinear methods. Another is to group funds with similar assets in order to exploit restrictions on their cross-sectional correlations, rather than focus exclusively on the time series properties of individual funds. Second, we find that existing factors used to approximate the systematic components of hedge fund returns can only capture a portion of their variability. The power of our test may be improved with a more complete set of hedge fund factors. Alternatively, it may be necessary to incorporate time-variation in factor loadings to more accurately capture the systematic component of hedge fund returns.

**References**


Ackerman, C., R. McEnally, and D. Ravenscraft. "The Performance of Hedge Funds: Risk, Return, and Incentives." *Journal of Finance* 54 (1999), 833-874.

Agarwal, V., N. D. Daniel, and N. Y. Naik. "Flows, Performance, and Managerial Incentives in the Hedge Fund Industry." Working paper (2003), Georgia State University.

Agarwal, V., and N. Y. Naik. "Multi-period Performance Persistence Analysis of Hedge Funds." *Journal of Financial and Quantitative Analysis* 35 (2000), 327-342.

Agarwal, V., and N. Y. Naik. "Risks and Portfolio Decisions Involving Hedge Funds." *Review of Financial Studies* 17 (2004), 63-98.

Anson, M. *Handbook of Alternative Assets*, John Wiley & Sons, New York, NY (2002).

Aragon, G. "Share Restrictions and Asset Pricing: Evidence from the Hedge Fund Industry." Working paper (2004), Boston College.

Asness, C., R. Krail, and J. Liew. "Do Hedge Funds Hedge?" *Journal of Portfolio Management* 28 (2001), 6-19.

Atkins, P. S., and C. A. Glassman. "Dissent of Commissioners Glassman and Atkins to Proposing Release No. IA-2266." www.sec.gov/rules/proposed/ia-2266.htm#dissent (2004).

Berk, J., and R. Green. "Mutual Fund Flows and Performance in Rational Markets." *Journal of Political Economy* 112 (2004), 1269-1295.

Bollen, N.P.B., and J. A. Busse. "On the Timing Ability of Mutual Fund Managers." *Journal of Finance* 56 (2000), 1075-1094.

Boudoukh, J., M. Richardson, M. Subrahmanyam, and R. F. Whitelaw. "Stale Prices and Strategies for Trading Mutual Funds." *Financial Analysts Journal*, 58 (2002), 53-71.

Brown, S. J., D. R. Gallagher, O. W. Steenbeek, and P. L. Swan. "Double or Nothing: Patterns of Equity Fund Holdings and Transactions." Working paper (2004), New York University.

Brown, S. J., W. N. Goetzmann, and R. G. Ibbotson. "Offshore Hedge Funds: Survival and Performance 1989-95." *Journal of Business* 72 (1999), 91-117.

Brown, K. C., W. V. Harlow, and L. T. Starks. "Of Tournaments and Temptations: An Analysis of Managerial Incentives in the Mutual Fund Industry." *Journal of Finance* 51 (1996), 85-110.

Burns, N., and S. Keida. "The impact of performance-based compensation on misreporting." *Journal of Financial Economics* 79 (2006), 35-67.

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

Carhart, M. "On the Persistence of Mutual Fund Performance." *Journal of Finance* 52 (1997), 57-82.

Chandar, N., and R. Bricker. "Incentives, Discretion, and Asset Valuation in Closed-end Mutual Funds." *Journal of Accounting Research* 40 (2002), 1037-1070.

Chevalier, J., and G. Ellison. "Risk Taking by Mutual Funds as a Response to Incentives." *Journal of Political Economy* 105 (1997), 1167-1200.

Dimson, E. "Risk Measurement when Shares are Subject to Infrequent Trading." *Journal of Financial Economics* 7 (1979), 197-226.

Fung, W., and D. A. Hsieh. "Empirical Characteristics of Dynamic Trading Strategies: The Case of Hedge Funds." *Review of Financial Studies* 10 (1997), 275-302.

Fung, W., and D. A. Hsieh. "The Risk in Hedge Fund Strategies: Theory and Evidence from Trend Followers." *Review of Financial Studies* 14 (2001), 313-341.

Fung, W., and D. A. Hsieh. "Asset-based Style Factors for Hedge Funds." *Financial Analysts Journal* 58 (2002), 16-27.

Fung, W., and D. A. Hsieh. "Hedge Fund Benchmarks: A Risk Based Approach." *Financial Analysts Journal* 60 (2004), 65-80.

Getmansky, M., A. W. Lo, and I. Makarov. "An Econometric Model of Serial Correlation and Illiquidity in Hedge Fund Returns." *Journal of Financial Economics* 74 (2004), 529-609.

Greene, W. *Econometric analysis*, Fifth Edition, Prentice Hall (2003).

Investment Company Institute. *2004 Mutual Fund Fact Book*, Investment Company Institute, Washington, D.C. (2004).

Jagannathan, R., and R. A. Korajczyk. "Assessing the Market Timing Performance of Managed Portfolios." *Journal of Business* 59 (1986), 217-235.

King, G., and L. Zeng. "Logistic Regression in Rare Events Data." Working paper (1999), Department of Government, Harvard University.

Koski, J., and J. Pontiff. "How are Derivatives Used? Evidence from the Mutual Fund Industry." *Journal of Finance* 54 (1999), 791-816.

Liu, J., and F. A. Longstaff. "Losing Money on Arbitrages: Optimal Dynamic Portfolio Choice in Markets with Arbitrage Opportunities." *Review of Financial Studies* 17 (2004), 611-641.

PricewaterhouseCoopers. *Investigation into Foreign Exchange Losses at the National Australia Bank*, PricewaterhouseCoopers, (2004).

Scholes, M., and J. Williams. "Estimating Betas from Non-synchronous Data." *Journal of Financial Economics* 5 (1977), 309-327.

Securities and Exchange Commission. *Implications of the Growth of Hedge Funds*, United States Securities and Exchange Commission, Washington, D.C. (2003).

Securities and Exchange Commission. *Registration Under the Advisers Act of Certain Hedge Fund Advisers*, Proposed Rule, File No. S7-30-04, United States Securities and Exchange Commission, Washington, D.C. (2004).

Securities and Exchange Commission. *Registration Under the Advisers Act of Certain Hedge Fund Advisers*, Final Rule, File No. S7-30-04, United States Securities and Exchange Commission, Washington, D.C. (2005)

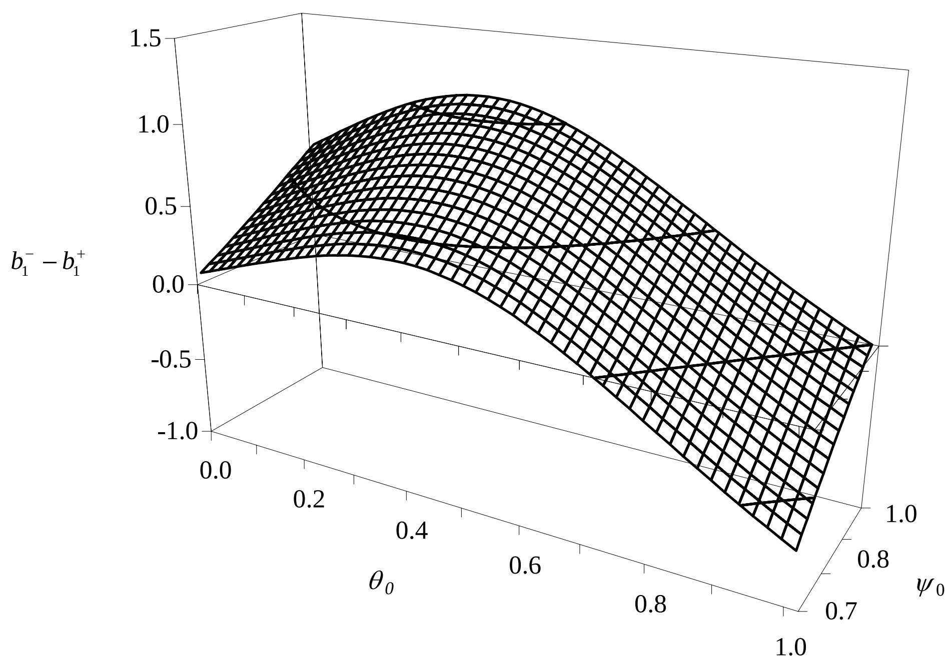
Sharpe, W. F. "Asset Allocation: Management Style and Performance Measurement." *Journal of Portfolio Management* 18 (1992), 7-19.

Shleifer, A., and R. Vishny. "The Limits of Arbitrage." *Journal of Finance* 52 (1997), 35-55.

Zitzewitz, E. "Who Cares About Shareholders? Arbitrage-Proofing Mutual Funds." *Journal of Law Economics & Organization* 19 (2003), 245-280.

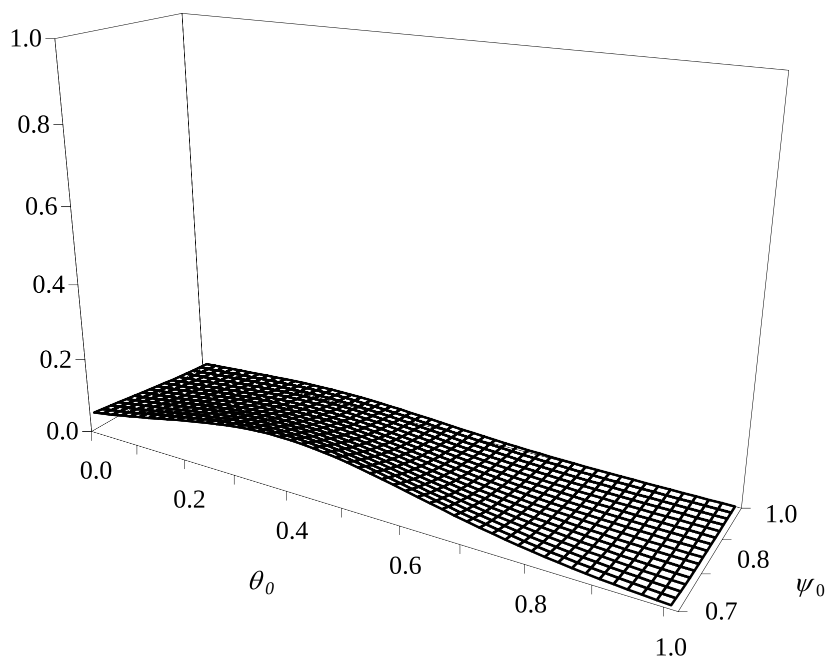
## Figure 1. Difference between conditional betas

Depicted on the vertical axis is the analytic difference between the conditional betas in a regression of the observed hedge fund return, $R_t^O$, on its lag: $R_t^O = a + \left(b_1^- \left(1 - I_{t-1}\right) + b_1^+ I_{t-1}\right) R_{t-1}^O + \eta_t$, where $I_{t-1}$ equals one if the lagged return of the hedge fund's assets, $R_{t-1}$, is above a specified level *c* and zero otherwise. The observed hedge fund return is related to the return of the hedge fund's assets through the following conditional smoothing algorithm: $R_t^O = \left(\theta_0 \left(1 - I_t\right) + \psi_0 I_t\right) R_t + \left(\theta_1 \left(1 - I_{t-1}\right) + \psi_1 I_{t-1}\right) R_{t-1}$. Asset returns are generated by the one-factor model $R_t = \mu + \beta \Lambda_t + \varepsilon_t$, $\Lambda_t \sim N\left(0, \sigma_\Lambda^2\right)$, $\varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right)$, with $\Lambda_t$ and $\varepsilon_t$ independent. Parameters are calibrated by averaging coefficients from 1,616 single-factor model regressions using the monthly returns of hedge funds from the CISDM database. For each hedge fund, a factor is selected to maximize the regression adjusted R-squared. Parameters are set to the following values: $\mu = 0.000$, $\beta = 1.250$, $\sigma_\Lambda = 0.025$, $\sigma_\varepsilon = 0.040$, $c = 0.000$, $\theta_1 = 1 - \theta_0$, and $\psi_1 = 1 - \psi_0$.



## Figure 2. Residual serial correlation

Depicted on the vertical axis is the analytic serial correlation of the residuals from a regression of the observed hedge fund return, $R_t^O$, on its lag: $R_t^O = a + \left(b_1^- \left(1 - I_{t-1}\right) + b_1^+ I_{t-1}\right) R_{t-1}^O + \eta_t$, where $I_{t-1}$ equals one if the lagged return of the hedge fund's assets, $R_{t-1}$, is above a specified level $c$ and zero otherwise. The observed hedge fund return is related to the return of the hedge fund's assets through the following conditional smoothing algorithm: $R_t^O = \left(\theta_0 \left(1 - I_t\right) + \psi_0 I_t\right) R_t + \left(\theta_1 \left(1 - I_{t-1}\right) + \psi_1 I_{t-1}\right) R_{t-1}$. Asset returns are generated by the one-factor model $R_t = \mu + \beta \Lambda_t + \varepsilon_t$, $\Lambda_t \sim N\left(0, \sigma_\Lambda^2\right)$, $\varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right)$, with $\Lambda_t$ and $\varepsilon_t$ independent. Parameters are calibrated by averaging coefficients from 1,616 single-factor model regressions using the monthly returns of hedge funds from the CISDM database. For each hedge fund, a factor is selected to maximize the regression adjusted R-squared. Parameters are set to the following values: $\mu = 0.000$, $\beta = 1.250$, $\sigma_\Lambda = 0.025$, $\sigma_\varepsilon = 0.040$, $c = 0.000$, $\theta_1 = 1 - \theta_0$, and $\psi_1 = 1 - \psi_0$.



## Table 1. Actual versus reported profits from currency options trading at NAB

Listed are profits and losses in thousands of AUD associated with the trading activity of the currency options trading desk of National Australia Bank (NAB). Data are from PricewaterhouseCoopers (2004).

| | | Monthly actual profit/(loss) | Monthly reported profits | (Under)/Overstatement of reported profits | Cumulative overstatement of portfolio value |
|---|---|---|---|---|---|
| 2002 | October | 8,946 | 974 | (7,972) | 0 |
| | November | 3,365 | 3,365 | 0 | 0 |
| | December | 2,837 | 2,837 | 0 | 0 |
| 2003 | January | 2,792 | 3,678 | 886 | 886 |
| | February | 2,559 | 2,650 | 91 | 977 |
| | March | 2,774 | 1,797 | (977) | 0 |
| | April | (10) | 2,567 | 2,577 | 2,577 |
| | May | (1,292) | 4,372 | 5,664 | 8,241 |
| | June | 3,390 | 4,558 | 1,168 | 9,409 |
| | July | 12,556 | 7,165 | (5,391) | 4,018 |
| | August | (169) | 1,323 | 1,492 | 5,510 |
| | September | (34,780) | 1,761 | 36,541 | 42,051 |
| | October | 13,871 | 5,774 | (8,097) | 33,954 |
| | November | 3,993 | 7,421 | 3,428 | 37,382 |
| | December | (49,106) | 5,272 | 54,378 | 91,760 |

## Table 2. SEC fraud cases

We examine SEC litigation releases involving 53 hedge fund fraud cases. Listed is the number of cases (# Cases) categorized by their primary offense as described in the SEC litigation releases. Also listed is the number of cases associated with of each type of offense (# Relevant) allowing for more than one offense for each case. “Misappropriation” includes cases in which a fund manager diverts investor capital for personal expenses. This is always categorized as the primary offense if it is associated with a case. “Misrepresent returns” includes cases in which a fund manager overstates or otherwise distorts fund asset values. “Misrepresent strategy” involves cases in which a fund manager misleads investors regarding the types of securities in which the fund invests or trading strategies used by the manager. “Fraudulent offerings” include cases in which the fund is advertised to new investors based on false information, including failure to disclose any disciplinary history. “Ponzi schemes” include cases for which the litigation report specifically mentions the terminology. We place funds in the “Other” category if they do not fit any of the first five groups, including, for example, illegal short sale and market timing activities.

| Fraud type | # Cases | # Relevant |
|---|---|---|
| Misappropriation | 21 | 21 |
| Misrepresent returns | 15 | 34 |
| Misrepresent strategy | 1 | 4 |
| Fraudulent offerings | 5 | 10 |
| Ponzi schemes | 6 | 6 |
| Other | 5 | 5 |
| *Total* | 53 | 80 |

## Table 3. Summary statistics of CISDM funds

Listed are summary statistics of the returns of hedge funds, CTAs, and managed futures in the December 2003 CISDM database. Live funds are in existence as of December 2003. Dead funds ceased reporting sometime prior to December 2003. Listed are the number of funds (#), equally-weighted average monthly return ($\mu$), standard deviation of returns ($\sigma$), Sharpe ratio ($S$), skewness ($W$), excess kurtosis ($K$), first-order serial correlation ($\rho$), and the number of funds with significant positive and significant negative serial correlation. 'E-D' denotes Event-Driven, 'G' denotes Global, and 'M-N' denotes Market Neutral.

**Panel A. Live Funds**

| | # | $\mu$ | $\sigma$ | $S$ | $W$ | $K$ | $\rho$ | # Pos | # Neg |
|---|---|---|---|---|---|---|---|---|---|
| *Hedge Funds* | | | | | | | | | |
| E-D | 139 | 0.98 | 2.77 | 0.33 | -0.24 | 3.98 | 0.20 | 62 | 0 |
| G Emerging | 96 | 1.63 | 6.70 | 0.35 | -0.07 | 4.79 | 0.19 | 29 | 0 |
| G Established | 288 | 1.33 | 5.08 | 0.24 | 0.41 | 2.72 | 0.13 | 61 | 0 |
| G International | 37 | 1.13 | 5.08 | 0.17 | 0.16 | 2.71 | 0.15 | 16 | 0 |
| G Macro | 46 | 1.11 | 4.39 | 0.25 | 0.23 | 2.08 | 0.05 | 6 | 0 |
| Long Only | 12 | 1.31 | 9.21 | 0.14 | 0.13 | 1.38 | 0.06 | 2 | 0 |
| M-N | 344 | 1.06 | 2.82 | 0.41 | -0.07 | 4.08 | 0.19 | 129 | 0 |
| Sector | 108 | 1.46 | 6.18 | 0.25 | 0.55 | 3.83 | 0.11 | 22 | 0 |
| Short-Sellers | 20 | 0.60 | 7.14 | 0.05 | -0.04 | 1.92 | 0.07 | 1 | 0 |
| Fund of Funds | 425 | 0.76 | 2.01 | 0.33 | -0.16 | 3.78 | 0.22 | 185 | 0 |
| *CTAs* | | | | | | | | | |
| Agriculture | 14 | 1.43 | 5.78 | 0.19 | 0.67 | 2.41 | 0.00 | 1 | 0 |
| Currency | 34 | 1.15 | 5.02 | 0.19 | 0.92 | 4.47 | 0.08 | 8 | 0 |
| Diversified | 143 | 1.46 | 6.94 | 0.16 | 0.57 | 2.29 | -0.05 | 0 | 8 |
| Energy | 2 | 1.23 | 4.34 | 0.25 | 1.01 | 1.35 | -0.01 | 0 | 0 |
| Financial | 46 | 1.37 | 6.05 | 0.18 | 0.64 | 2.59 | 0.01 | 5 | 0 |
| Stock Index | 22 | 1.25 | 6.10 | 0.13 | 0.42 | 5.58 | 0.04 | 3 | 1 |
| *Managed Futures* | | | | | | | | | |
| Public Pools | 161 | 1.15 | 5.58 | 0.17 | 0.45 | 2.20 | 0.03 | 11 | 6 |
| Private Pools | 103 | 1.20 | 6.57 | 0.15 | 0.48 | 3.57 | 0.01 | 5 | 6 |

## Table 3. Summary statistics of CISDM funds (continued)

**Panel B. Dead Funds**

| | # | $\mu$ | $\sigma$ | $S$ | $W$ | $K$ | $\rho$ | # Pos | # Neg |
|---|---|---|---|---|---|---|---|---|---|
| *Hedge Funds* | | | | | | | | | |
| E-D | 85 | 1.02 | 4.75 | 0.18 | -0.20 | 4.89 | 0.13 | 22 | 0 |
| G Emerging | 41 | 0.38 | 8.21 | 0.05 | -0.85 | 5.89 | 0.16 | 8 | 0 |
| G Established | 258 | 1.24 | 6.93 | 0.16 | 0.01 | 4.86 | 0.10 | 45 | 2 |
| G International | 21 | 1.22 | 6.22 | 0.19 | 0.10 | 4.86 | 0.20 | 6 | 0 |
| G Macro | 53 | 0.79 | 5.63 | 0.11 | 0.02 | 3.58 | 0.10 | 10 | 0 |
| Long Only | 12 | 0.72 | 9.06 | 0.04 | -0.25 | 1.28 | -0.01 | 0 | 0 |
| M-N | 175 | 0.88 | 3.45 | 0.21 | -0.42 | 4.64 | 0.15 | 50 | 0 |
| Sector | 52 | 1.88 | 8.80 | 0.21 | 0.39 | 1.55 | 0.11 | 5 | 0 |
| Short-Sellers | 8 | 0.50 | 6.77 | 0.04 | 0.03 | 0.55 | 0.09 | 0 | 0 |
| Fund of Funds | 151 | 0.64 | 3.73 | 0.12 | -0.15 | 3.08 | 0.19 | 58 | 1 |
| *CTAs* | | | | | | | | | |
| Agriculture | 18 | 2.86 | 14.14 | 0.16 | 0.89 | 4.89 | 0.03 | 2 | 2 |
| Currency | 57 | 0.97 | 4.69 | 0.05 | 0.92 | 3.75 | 0.00 | 5 | 9 |
| Diversified | 177 | 1.20 | 7.71 | 0.08 | 0.77 | 3.52 | -0.03 | 0 | 19 |
| Energy | 6 | 0.09 | 13.12 | -0.02 | 1.33 | 7.99 | -0.07 | 0 | 0 |
| Financial | 70 | 0.86 | 5.59 | 0.08 | 0.44 | 2.05 | -0.01 | 2 | 4 |
| Stock Index | 27 | 0.87 | 7.79 | 0.06 | -0.02 | 2.98 | 0.03 | 5 | 0 |
| *Managed Futures* | | | | | | | | | |
| Public Pools | 336 | 0.48 | 5.15 | 0.01 | 0.29 | 2.81 | -0.01 | 4 | 24 |
| Private Pools | 174 | 0.81 | 6.78 | 0.07 | 0.41 | 3.38 | -0.02 | 3 | 17 |

## Table 4. History lengths of CISDM funds

Listed are the number of funds (#) and the 25th, 50th, and 75th percentiles of the cross-sectional distributions of history lengths, in months, for hedge funds, CTAs, and managed futures in the December 2003 CISDM database. Live funds are in existence as of December 2003 whereas Dead funds were liquidated in a prior month. 'E-D' denotes Event-Driven, 'G' denotes Global, and 'M-N' denotes Market Neutral.

| | **Live Funds** | | | | **Dead Funds** | | | |
|---|---|---|---|---|---|---|---|---|
| | # | 25th | 50th | 75th | # | 25th | 50th | 75th |
| *Hedge Funds* | | | | | | | | |
| E-D | 139 | 48 | 78 | 108 | 85 | 36 | 53 | 86 |
| G Emerging | 96 | 50 | 73 | 94 | 41 | 30 | 42 | 69 |
| G Established | 288 | 40 | 67 | 96 | 258 | 38 | 53 | 92 |
| G International | 37 | 62 | 88 | 120 | 21 | 47 | 59 | 73 |
| G Macro | 46 | 43 | 79 | 104 | 53 | 36 | 48 | 73 |
| Long Only | 12 | 63 | 80 | 92 | 12 | 32 | 48 | 78 |
| M-N | 344 | 36 | 63 | 89 | 175 | 36 | 55 | 72 |
| Sector | 108 | 42 | 54 | 88 | 52 | 32 | 47 | 64 |
| Short-Sellers | 20 | 51 | 79 | 86 | 8 | 34 | 61 | 66 |
| Fund of Funds | 425 | 43 | 71 | 108 | 151 | 35 | 55 | 74 |
| *CTAs* | | | | | | | | |
| Agriculture | 14 | 52 | 65 | 114 | 18 | 38 | 58 | 69 |
| Currency | 34 | 87 | 108 | 152 | 57 | 39 | 59 | 102 |
| Diversified | 143 | 60 | 101 | 142 | 177 | 41 | 65 | 102 |
| Energy | 2 | 36 | 36 | 52 | 6 | 35 | 47 | 73 |
| Financial | 46 | 76 | 98 | 122 | 70 | 39 | 66 | 95 |
| Stock Index | 22 | 52 | 81 | 116 | 27 | 43 | 53 | 69 |
| *Managed Futures* | | | | | | | | |
| Public Pools | 161 | 58 | 87 | 126 | 336 | 37 | 56 | 85 |
| Private Pools | 103 | 71 | 107 | 153 | 174 | 40 | 62 | 118 |

## Table 5. Summary statistics of CISDM indices and asset-based style factors

Listed are summary statistics of the returns of hedge fund indices and CTA indices in the December 2003 CISDM database, as well as the asset-based style (ABS) factors developed by Fung and Hsieh (2004). Data are from 1994 – 2003. Statistics include the average monthly return ($\mu$), standard deviation of returns ($\sigma$), Sharpe ratio ($S$), skewness ($W$), excess kurtosis ($K$), first-order serial correlation ($\rho$), and the associated two-sided $p$-value. 'E-D' denotes Event-Driven, 'G' denotes Global, and 'M-N' denotes Market Neutral.

| | $\mu$ | $\sigma$ | $S$ | $W$ | $K$ | $\rho$ | $p$-value |
|---|---|---|---|---|---|---|---|
| *HF Indices* | | | | | | | |
| E-D Dist | 0.91 | 1.66 | 0.34 | -2.05 | 10.54 | 0.36 | 0.00 |
| E-D | 0.81 | 1.17 | 0.40 | -2.29 | 14.40 | 0.28 | 0.00 |
| E-D Arb | 0.82 | 1.05 | 0.45 | -1.91 | 10.32 | 0.30 | 0.00 |
| G Emerg | 0.69 | 4.08 | 0.08 | -2.42 | 14.78 | 0.25 | 0.01 |
| G Est | 1.11 | 2.47 | 0.31 | -0.17 | 2.56 | 0.21 | 0.02 |
| G Macro | 0.63 | 1.55 | 0.19 | 0.52 | 4.08 | 0.04 | 0.70 |
| M-N Arb | 1.02 | 2.27 | 0.30 | 3.87 | 20.60 | 0.26 | 0.00 |
| M-N L/S | 0.79 | 0.57 | 0.78 | 0.26 | 1.42 | 0.29 | 0.00 |
| M-N | 0.79 | 0.39 | 1.13 | -0.27 | 0.47 | 0.50 | 0.00 |
| Short | 0.39 | 5.20 | 0.01 | 0.56 | 1.83 | 0.13 | 0.15 |
| *CTA Indices* | | | | | | | |
| Currency | 0.49 | 2.11 | 0.07 | 0.81 | 3.06 | 0.09 | 0.33 |
| Disc | 0.72 | 1.45 | 0.26 | 0.36 | 0.22 | 0.15 | 0.09 |
| Divers | 0.80 | 3.10 | 0.15 | 0.35 | 0.13 | 0.01 | 0.90 |
| Finl | 0.84 | 3.35 | 0.15 | 0.53 | 0.31 | 0.11 | 0.23 |
| Stock | 0.32 | 2.89 | -0.01 | -0.46 | 1.26 | -0.01 | 0.95 |
| System | 0.62 | 2.78 | 0.10 | 0.30 | 0.28 | 0.02 | 0.82 |
| Trend | 0.87 | 4.14 | 0.13 | 0.29 | -0.12 | 0.05 | 0.57 |
| *ABS Factors* | | | | | | | |
| S&P 500 | 0.98 | 4.57 | 0.14 | -0.59 | 0.20 | -0.01 | 0.94 |
| Wilshire Size | 0.04 | 3.50 | -0.09 | 0.51 | 3.71 | -0.11 | 0.23 |
| Bond Trend | 1.50 | 19.12 | 0.06 | 1.41 | 2.10 | 0.13 | 0.17 |
| Currency Trend | -0.16 | 19.39 | -0.03 | 1.40 | 3.35 | 0.00 | 0.99 |
| Commodity Trend | -1.14 | 12.62 | -0.12 | 1.59 | 5.16 | -0.15 | 0.09 |
| D 10yr Treasury | -0.01 | 0.24 | -1.48 | 0.35 | -0.39 | 0.26 | 0.00 |
| D Credit Spread | 0.00 | 0.13 | -2.62 | 0.90 | 1.84 | 0.38 | 0.00 |

## Table 6. Conditional serial correlation of CISDM indices and asset-based style factors

Listed are conditional serial correlation regression statistics for hedge fund indices and CTA indices in the December 2003 CISDM database, as well as the asset-based style (ABS) factors developed by Fung and Hsieh (2004). Data are from 1994 – 2003. Regressions are of the form: $R_t = a + b_1^+ R_{t-1} + b_1^- \left(1 - I_{t-1}\right) R_{t-1} + \eta_t$, where $I_{t-1} = 1$ if the return in month $t-1$ is greater than its mean and zero otherwise.

| | R-sq | $a$ | $p$-value | $b_1^+$ | $p$-value | $b_1^-$ | $p$-value |
|---|---|---|---|---|---|---|---|
| *Hedge Fund Indices* | | | | | | | |
| E-D Dist | 0.1152 | 0.0050 | 0.0124 | 0.4158 | 0.0013 | -0.1305 | 0.5118 |
| E-D | 0.0633 | 0.0056 | 0.0001 | 0.2906 | 0.0110 | -0.0290 | 0.8660 |
| E-D Arb | 0.0874 | 0.0051 | 0.0001 | 0.3692 | 0.0008 | -0.1827 | 0.2938 |
| G Emerg | 0.0516 | 0.0019 | 0.7135 | 0.3606 | 0.0445 | -0.1815 | 0.4667 |
| G Est | 0.0289 | 0.0077 | 0.0128 | 0.2530 | 0.0416 | -0.1377 | 0.5776 |
| G Macro | -0.0108 | 0.0066 | 0.0002 | -0.0159 | 0.8900 | 0.1783 | 0.4457 |
| M-N Arb | 0.0554 | 0.0076 | 0.0009 | 0.2379 | 0.0131 | 0.1904 | 0.4756 |
| M-N L/S | 0.0839 | 0.0050 | 0.0000 | 0.3055 | 0.0008 | 0.2273 | 0.1782 |
| M-N | 0.2578 | 0.0048 | 0.0000 | 0.4426 | 0.0000 | -0.1996 | 0.0850 |
| Short | 0.0012 | 0.0027 | 0.7058 | 0.1586 | 0.2874 | -0.0615 | 0.8250 |
| *CTA Indices* | | | | | | | |
| Currency | -0.0009 | 0.0067 | 0.0179 | -0.0047 | 0.9721 | 0.2739 | 0.3340 |
| Disc | 0.0141 | 0.0070 | 0.0001 | 0.0923 | 0.4142 | 0.2514 | 0.3504 |
| Divers | -0.0109 | 0.0054 | 0.2292 | 0.1024 | 0.4715 | -0.2619 | 0.3985 |
| Finl | 0.0025 | 0.0048 | 0.2929 | 0.1978 | 0.1373 | -0.2695 | 0.3626 |
| Stock | -0.0144 | 0.0021 | 0.5989 | 0.0775 | 0.6518 | -0.1541 | 0.5708 |
| System | -0.0166 | 0.0061 | 0.1277 | 0.0367 | 0.8008 | -0.0420 | 0.8887 |
| Trend | 0.0034 | 0.0015 | 0.8171 | 0.2236 | 0.1389 | -0.4639 | 0.1531 |
| *ABS Factors* | | | | | | | |
| S&P 500 | -0.0123 | 0.0055 | 0.4323 | 0.1045 | 0.5534 | -0.2246 | 0.4560 |
| Wilshire Size | 0.0156 | 0.0061 | 0.2059 | -0.3043 | 0.0515 | 0.4282 | 0.1224 |
| Bond Trend | 0.0464 | -0.0393 | 0.1668 | 0.3501 | 0.0079 | -0.8187 | 0.0178 |
| Currency Trend | -0.0149 | -0.0110 | 0.7086 | 0.0564 | 0.6924 | -0.1631 | 0.6092 |
| Commodity Trend | 0.0090 | -0.0191 | 0.3111 | -0.0942 | 0.5093 | -0.1576 | 0.6023 |
| D 10yr Treasury | 0.0529 | -0.0002 | 0.5349 | 0.3138 | 0.0691 | -0.1177 | 0.7109 |
| D Credit Spread | 0.1303 | 0.0001 | 0.7468 | 0.3587 | 0.0078 | 0.0535 | 0.8376 |

## Table 7. Size analysis of conditional serial correlation

Simulated hedge fund returns are based on the hedge funds, CTAs, and managed futures in the December 2003 CISDM database. For each fund in the sample, 20 sets of simulated hedge fund returns are generated. First, for each fund, parameters of a single-factor model $R_t^O = \alpha + \beta^O \Lambda_t + \varepsilon_t$ are estimated, where $R_t^O$ is the observed hedge fund return and $\Lambda_t$ is the return of the CISDM index that maximizes the adjusted R-squared of the regression. Second, simulated asset returns are constructed as $R_t^A = \alpha + \beta^O \Lambda_t^E + \xi_t$ , where $\Lambda_t^E$ is randomly drawn from the empirical return distribution of the selected index, and $\xi_t$ is a randomly generated mean-zero normal variate, with standard deviation calibrated to the original single-factor model. Third, simulated hedge fund returns are constructed by smoothing simulated asset returns as follows: $R_t^S = 0.5R_t^A + 0.5R_{t-1}^A$. Three sets of simulations are conducted using lengths of 120, 60, and 36 months. Listed is the percentage of simulated hedge funds for which conditional smoothing is detected using the following regression: $R_t^S = a + b_1^+ R_{t-1}^S + b_1^- \left(1 - I_{t-1}\right) R_{t-1}^S + \eta_t$ , where $R_t^S$ is the simulated fund return in month *t* and $I_{t-1}$ equals one if the simulated fund's systematic return from an optimal factor model at month *t* – 1 is greater than the mean systematic return. Listed for each fund type are the number of funds (#) and the percentage of simulations with significant positive $b_1^-$ coefficients evaluated at the two-sided 5% level. Results are listed for when the econometrician knows the factor when computing $I_t$, and for when the econometrician must infer the factor.

| | | Known Factor | | | Unobservable Factor | | |
|---|---|---|---|---|---|---|---|
| | | *Months* | | | *Months* | | |
| | # | 120 | 60 | 36 | 120 | 60 | 36 |
| *Hedge Funds* | | | | | | | |
| E-D | 210 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 |
| G Emerging | 137 | 0.04 | 0.04 | 0.03 | 0.04 | 0.04 | 0.03 |
| G Established | 534 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 |
| G International | 58 | 0.02 | 0.02 | 0.03 | 0.02 | 0.02 | 0.02 |
| G Macro | 92 | 0.02 | 0.02 | 0.02 | 0.03 | 0.02 | 0.02 |
| Long Only | 24 | 0.04 | 0.01 | 0.01 | 0.04 | 0.02 | 0.02 |
| M-N | 511 | 0.01 | 0.02 | 0.01 | 0.02 | 0.02 | 0.01 |
| Sector | 160 | 0.01 | 0.01 | 0.01 | 0.01 | 0.02 | 0.02 |
| Short-Sellers | 25 | 0.01 | 0.02 | 0.01 | 0.01 | 0.02 | 0.01 |
| Fund of Funds | 567 | 0.02 | 0.01 | 0.01 | 0.02 | 0.01 | 0.01 |
| *CTAs* | | | | | | | |
| Agriculture | 32 | 0.01 | 0.02 | 0.02 | 0.01 | 0.02 | 0.03 |
| Currency | 90 | 0.01 | 0.02 | 0.02 | 0.01 | 0.01 | 0.02 |
| Diversified | 319 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 |
| Energy | 8 | 0.01 | 0.01 | 0.01 | 0.02 | 0.02 | 0.01 |
| Financial | 116 | 0.01 | 0.01 | 0.01 | 0.01 | 0.02 | 0.02 |
| Stock Index | 49 | 0.02 | 0.03 | 0.01 | 0.02 | 0.02 | 0.02 |
| *Managed Futures* | | | | | | | |
| Public Pools | 489 | 0.02 | 0.01 | 0.01 | 0.02 | 0.02 | 0.01 |
| Private Pools | 268 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 |

## Table 8. Power analysis of conditional serial correlation under controlled conditions

Simulated hedge fund returns are based on the hedge funds, CTAs, and managed futures in the December 2003 CISDM database. For each fund in the sample, 20 sets of simulated hedge fund returns are generated. First, for each fund, parameters of a single-factor model $R_t^O = \alpha + \beta^O \Lambda_t + \varepsilon_t$ are estimated, where $R_t^O$ is the observed hedge fund return and $\Lambda_t$ is the return of the CISDM index that maximizes the adjusted R-squared of the regression. Second, simulated asset returns are constructed as $R_t^A = \alpha + \beta^O \Lambda_t^E + \xi_t$ , where $\Lambda_t^E$ is randomly drawn from the empirical return distribution of the selected index, and $\xi_t$ is a randomly generated mean-zero normal variate, with standard deviation calibrated to the original single-factor model. Third, simulated hedge fund returns are constructed by smoothing simulated asset returns as follows: $R_t^S = \left(0.5\left(1 - I_t\right) + I_t\right) R_t^A + 0.5\left(1 - I_{t-1}\right) R_{t-1}^A$ where $I_t$ equals one if $\alpha + \beta^O \Lambda_t^E$ is above its mean and zero otherwise. Three sets of simulations are conducted using lengths of 120, 60, and 36 months. Listed is the percentage of simulated hedge funds for which conditional smoothing is detected using the following regression: $R_t^S = a + b_1^+ R_{t-1}^S + b_1^- \left(1 - I_{t-1}\right) R_{t-1}^S + \eta_t$ , where $R_t^S$ is the simulated fund return in month *t* and $I_{t-1}$ equals one if the simulated fund's systematic return from an optimal factor model at month *t* – 1 is greater than the mean systematic return. Listed for each fund type are the number of funds (#) and the percentage of simulations with significant positive $b_1^-$ coefficients evaluated at the two-sided 5% level. Results are listed for when the econometrician knows the factor used to compute $I_t$ , and for when the econometrician must infer the factor.

| | | Known Factor | | | Unobservable Factor | | |
|---|---|---|---|---|---|---|---|
| | | *Months* | | | *Months* | | |
| | # | 120 | 60 | 36 | 120 | 60 | 36 |
| *Hedge Funds* | | | | | | | |
| E-D | 210 | 0.81 | 0.53 | 0.33 | 0.77 | 0.44 | 0.25 |
| G Emerging | 137 | 0.84 | 0.56 | 0.32 | 0.78 | 0.46 | 0.24 |
| G Established | 534 | 0.79 | 0.50 | 0.31 | 0.74 | 0.43 | 0.23 |
| G International | 58 | 0.84 | 0.55 | 0.34 | 0.76 | 0.44 | 0.23 |
| G Macro | 92 | 0.84 | 0.54 | 0.34 | 0.75 | 0.41 | 0.23 |
| Long Only | 24 | 0.77 | 0.46 | 0.29 | 0.76 | 0.41 | 0.24 |
| M-N | 511 | 0.86 | 0.59 | 0.38 | 0.77 | 0.45 | 0.25 |
| Sector | 160 | 0.80 | 0.49 | 0.32 | 0.76 | 0.43 | 0.25 |
| Short-Sellers | 25 | 0.79 | 0.48 | 0.28 | 0.76 | 0.44 | 0.23 |
| Fund of Funds | 567 | 0.77 | 0.48 | 0.30 | 0.74 | 0.43 | 0.25 |
| *CTAs* | | | | | | | |
| Agriculture | 32 | 0.88 | 0.59 | 0.38 | 0.63 | 0.33 | 0.20 |
| Currency | 90 | 0.85 | 0.56 | 0.34 | 0.76 | 0.44 | 0.24 |
| Diversified | 319 | 0.78 | 0.48 | 0.30 | 0.71 | 0.38 | 0.22 |
| Energy | 8 | 0.86 | 0.54 | 0.42 | 0.60 | 0.29 | 0.18 |
| Financial | 116 | 0.81 | 0.49 | 0.32 | 0.74 | 0.38 | 0.22 |
| Stock Index | 49 | 0.86 | 0.58 | 0.35 | 0.73 | 0.39 | 0.21 |
| *Managed Futures* | | | | | | | |
| Public Pools | 489 | 0.76 | 0.47 | 0.29 | 0.70 | 0.39 | 0.22 |
| Private Pools | 268 | 0.78 | 0.48 | 0.30 | 0.70 | 0.39 | 0.21 |

## Table 9. Factor model regressions

Listed are details of several factor model regressions using hedge funds, CTAs, and managed futures in the December 2003 CISDM database. For each fund, a subset of available factors is selected to maximize the adjusted R-squared of the regression, subject to the criterion that simpler regressions are favored if additional factors do not statistically significantly improve the fit. Panel A shows results when available factors are the S&P 500 index return and returns of the CISDM hedge fund indices and CTA indices. Listed within each fund category are the number of funds (#) and the 25$^{th}$, 50$^{th}$, and 75$^{th}$ percentiles of the cross-sectional distributions of the adjusted $R^2$ when the number of factors is one, two, or unconstrained. Also listed are the average adjusted R-squared, the average number of contemporaneous factors, and the average number of lagged factors in the unconstrained regressions. Panel B shows results when available factors are the S&P 500 index return and the returns of the seven asset-based style factors developed by Fung and Hsieh (2004).

**Panel A. CISDM**

| | | 1 Factor | | | 2 Factors | | | Unconstrained | | | Unconstrained | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | # | 25$^{th}$ | 50$^{th}$ | 75$^{th}$ | 25$^{th}$ | 50$^{th}$ | 75$^{th}$ | 25$^{th}$ | 50$^{th}$ | 75$^{th}$ | Avg. | # Cont | # Lag |
| Hedge Funds | 2,318 | 0.15 | 0.31 | 0.47 | 0.21 | 0.37 | 0.53 | 0.28 | 0.46 | 0.62 | 0.45 | 3.41 | 1.10 |
| CTAs | 614 | 0.05 | 0.18 | 0.43 | 0.08 | 0.23 | 0.47 | 0.09 | 0.26 | 0.49 | 0.31 | 2.31 | 0.16 |
| Managed Futures | 757 | 0.12 | 0.34 | 0.59 | 0.16 | 0.39 | 0.62 | 0.18 | 0.42 | 0.64 | 0.42 | 2.53 | 0.25 |

**Panel B. ABS**

| | | 1 Factor | | | 2 Factors | | | Unconstrained | | | Unconstrained | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | # | 25$^{th}$ | 50$^{th}$ | 75$^{th}$ | 25$^{th}$ | 50$^{th}$ | 75$^{th}$ | 25$^{th}$ | 50$^{th}$ | 75$^{th}$ | Avg. | # Cont | # Lag |
| Hedge Funds | 2,318 | 0.07 | 0.15 | 0.27 | 0.10 | 0.22 | 0.38 | 0.12 | 0.26 | 0.42 | 0.29 | 2.13 | 0.50 |
| CTAs | 614 | 0.05 | 0.11 | 0.18 | 0.07 | 0.15 | 0.25 | 0.07 | 0.17 | 0.29 | 0.19 | 2.06 | 0.20 |
| Managed Futures | 757 | 0.07 | 0.13 | 0.21 | 0.10 | 0.20 | 0.28 | 0.11 | 0.23 | 0.35 | 0.25 | 2.37 | 0.36 |

## Table 10. Power analysis of conditional serial correlation under actual conditions

Simulated hedge fund returns are based on the hedge funds, CTAs, and managed futures in the December 2003 CISDM database. For each fund in the sample, 20 sets of simulated hedge fund returns are generated and tested for conditional serial correlation. First, simulated asset returns $R_t^A$ are constructed by reordering the residuals from a regression of actual fund returns on the subset of factors and their lags that maximizes the regression's adjusted R-squared. Second, simulated hedge fund returns are constructed by smoothing the simulated asset returns as follows: $R_t^S = \left(0.5\left(1-I_t\right)+I_t\right)R_t^A + 0.5\left(1-I_{t-1}\right)R_{t-1}^A$ where $I_t$ equals one if the month *t* systematic return from the optimal factor model is above its mean and zero otherwise. Third, simulated hedge fund returns are regressed on their lag: $R_t^S = a + b_1^+ R_{t-1}^S + b_1^- \left(1-I_{t-1}\right)R_{t-1}^S + \eta_t$, where $R_t^S$ is the simulated fund return in month *t* and $I_{t-1}$ equals one if the econometrician's estimate of the simulated fund's systematic return at month *t* – 1 is greater than the mean systematic return. Listed for each fund type are the number of funds (#) and the percentage of simulations with significant positive $b_1^-$ coefficient evaluated at the two-sided 5%, 10%, and 20% levels. Listed are results when CISDM indices, as well as the asset-based style (ABS) factors developed by Fung and Hsieh (2004), are used as available factors in determining $I_{t-1}$.

| | | CISDM | | | ABS | | |
|---|---|---|---|---|---|---|---|
| | # | 5% | 10% | 20% | 5% | 10% | 20% |
| *Hedge Funds* | | | | | | | |
| E-D | 210 | 0.36 | 0.45 | 0.56 | 0.39 | 0.48 | 0.61 |
| G Emerging | 137 | 0.37 | 0.47 | 0.59 | 0.37 | 0.47 | 0.59 |
| G Established | 534 | 0.28 | 0.38 | 0.51 | 0.34 | 0.44 | 0.56 |
| G International | 58 | 0.35 | 0.45 | 0.57 | 0.43 | 0.52 | 0.63 |
| G Macro | 92 | 0.27 | 0.36 | 0.47 | 0.32 | 0.41 | 0.53 |
| Long Only | 24 | 0.39 | 0.48 | 0.61 | 0.37 | 0.48 | 0.61 |
| M-N | 511 | 0.28 | 0.37 | 0.49 | 0.34 | 0.43 | 0.53 |
| Sector | 160 | 0.26 | 0.36 | 0.49 | 0.33 | 0.43 | 0.56 |
| Short-Sellers | 25 | 0.36 | 0.48 | 0.60 | 0.46 | 0.57 | 0.70 |
| Fund of Funds | 567 | 0.35 | 0.45 | 0.58 | 0.39 | 0.49 | 0.61 |
| *CTAs* | | | | | | | |
| Agriculture | 32 | 0.30 | 0.39 | 0.48 | 0.27 | 0.35 | 0.43 |
| Currency | 90 | 0.34 | 0.43 | 0.55 | 0.33 | 0.42 | 0.54 |
| Diversified | 319 | 0.34 | 0.45 | 0.58 | 0.32 | 0.41 | 0.53 |
| Energy | 8 | 0.13 | 0.22 | 0.35 | 0.16 | 0.23 | 0.31 |
| Financial | 116 | 0.32 | 0.42 | 0.55 | 0.31 | 0.41 | 0.53 |
| Stock Index | 48 | 0.28 | 0.36 | 0.47 | 0.29 | 0.39 | 0.50 |
| *Managed Futures* | | | | | | | |
| Public Pools | 489 | 0.29 | 0.40 | 0.52 | 0.26 | 0.36 | 0.48 |
| Private Pools | 268 | 0.31 | 0.42 | 0.54 | 0.30 | 0.39 | 0.51 |

## Table 11. Frequency of conditional serial correlation

Regressions take the following form: $R_t^O = a + b_1^+ R_{t-1}^O + b_1^- \left(1 - I_{t-1}\right) R_{t-1}^O + \eta_t$, where $R_t^O$ is the observed fund return in month $t$ and $I_{t-1}$ equals one if the fund's systematic return from an optimal factor model at month $t$ – 1 is greater than the mean systematic return. Listed for each type of hedge fund, CTA, and managed futures are the number of funds and the number of funds with significant positive and significant negative coefficients evaluated at the two-sided 5% level. Data are from the December 2003 CISDM database. Listed are results when CISDM indices, as well as the asset-based style (ABS) factors developed by Fung and Hsieh (2004), are used as available factors in determining $I_{t-1}$.

| | | CISDM | | | | ABS | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | $b_1^+$ | | $b_1^-$ | | $b_1^+$ | | $b_1^-$ | |
| | # | Pos | Neg | Pos | Neg | Pos | Neg | Pos | Neg |
| *Hedge Funds* | | | | | | | | | |
| E-D | 210 | 56 | 4 | 14 | 10 | 60 | 4 | 10 | 7 |
| G Emerging | 137 | 22 | 2 | 7 | 4 | 24 | 0 | 1 | 5 |
| G Established | 534 | 65 | 9 | 30 | 16 | 64 | 7 | 14 | 24 |
| G International | 58 | 17 | 0 | 1 | 4 | 19 | 0 | 3 | 4 |
| G Macro | 92 | 11 | 3 | 3 | 3 | 11 | 4 | 4 | 5 |
| Long Only | 24 | 3 | 1 | 3 | 1 | 2 | 1 | 3 | 1 |
| M-N | 511 | 147 | 7 | 20 | 38 | 144 | 15 | 25 | 35 |
| Sector | 160 | 20 | 2 | 16 | 5 | 26 | 1 | 10 | 6 |
| Short-Sellers | 25 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 |
| Fund of Funds | 567 | 168 | 6 | 27 | 17 | 191 | 4 | 16 | 24 |
| *CTAs* | | | | | | | | | |
| Agriculture | 32 | 1 | 3 | 7 | 1 | 1 | 1 | 5 | 0 |
| Currency | 90 | 13 | 7 | 5 | 2 | 10 | 5 | 2 | 3 |
| Diversified | 319 | 11 | 18 | 11 | 17 | 17 | 16 | 14 | 21 |
| Energy | 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| Financial | 116 | 7 | 5 | 2 | 8 | 8 | 5 | 3 | 8 |
| Stock Index | 49 | 2 | 3 | 2 | 2 | 5 | 2 | 2 | 6 |
| *Managed Futures* | | | | | | | | | |
| Public Pools | 489 | 39 | 18 | 21 | 25 | 36 | 17 | 15 | 29 |
| Private Pools | 268 | 12 | 14 | 9 | 9 | 10 | 10 | 13 | 7 |
| *Totals* | 3,689 | 595 | 102 | 178 | 163 | 629 | 92 | 140 | 187 |
| *% Rejections* | | 16.13% | 2.76% | 4.83% | 4.42% | 17.05% | 2.49% | 3.80% | 5.07% |

## Table 12. Frequency of conditional serial correlation in mutual funds

Regressions take the following form: $R_t^O = a + b_1^+ R_{t-1}^O + b_1^- \left(1 - I_{t-1}\right) R_{t-1}^O + \eta_t$, where $R_t^O$ is the observed fund return in month *t* and $I_{t-1}$ equals one if the fund's systematic return from Carhart's (1997) four factor model at month *t* – 1 is greater than the mean systematic return. Listed for each type of equity mutual fund are the number of funds and the number of funds with significant positive $b_1^-$ coefficients evaluated at the two-sided 5% level. Data are from the CRSP Survivor-Bias Free U.S. Mutual Fund Database. To be included, a mutual fund must have at least 5 years of return data in the 1994 to 2003 period.

| | # Funds | # Rejections |
|---|---|---|
| Aggressive Growth | 1,136 | 18 |
| Global Equity | 348 | 0 |
| Growth and Income | 866 | 3 |
| Intl Equity | 1,133 | 9 |
| Long Term Growth | 1,514 | 14 |
| Sector Funds | 461 | 6 |
| *Total* | 5,458 | 50 |
| *% Rejections* | | 0.92% |

## Table 13. Cross-sectional analysis of flagged hedge funds

Listed are details of logit regressions in which the dependent variable equals one if a fund features statistically significant conditional serial correlation and zero otherwise. Conditional serial correlation is measured in the regression: $R_t^O = a + b_1^+ R_{t-1}^O + b_1^- \left(1 - I_{t-1}\right) R_{t-1}^O + \eta_t$, where $R_t^O$ is the observed fund return in month $t$ and $I_{t-1}$ equals one if the fund's systematic return from an optimal factor model at month $t$ – 1 is greater than the mean systematic return. Listed are results when CISDM indices, as well as the asset-based style (ABS) factors developed by Fung and Hsieh (2004), are used as available factors in determining $I_{t-1}$. Independent variables in Panel A are ln(Cfvol), the natural logarithm of the volatility of investor cash flows as a percentage of fund assets, and Cfmu, the monthly mean investor cash flow as a percentage of fund assets. Panel B adds E[$r$], fund mean monthly return, Fee, management fee, Incent, percentage of fund profits, Live, an indicator variable that equals one if the fund is live as of December 2003, ln(Size), the natural logarithm of the maximum size of the fund in the sample, Audit, an indicator variable that equals one if the fund has been audited, Age, the age of the fund in months, and ln(Wait), the natural logarithm of the number of days an investor must wait before capital is returned upon demand.

| | **Panel A. Model I** | | | | **Panel B. Model II** | | | |
|---|---|---|---|---|---|---|---|---|
| | CISDM | | ABS | | CISDM | | ABS | |
| | Coefficient | *p*-value | Coefficient | *p*-value | Coefficient | *p*-value | Coefficient | *p*-value |
| Constant | -2.4207 | 0.0000 | -2.7607 | 0.0000 | -2.5149 | 0.0000 | -1.8086 | 0.0052 |
| ln(Cfvol) | 0.2203 | 0.0287 | 0.2177 | 0.0502 | 0.3326 | 0.0201 | 0.4161 | 0.0117 |
| Cfmu | -4.9885 | 0.0021 | -1.0530 | 0.5365 | -5.7388 | 0.0316 | -4.5155 | 0.0593 |
| E[r] | | | | | 12.0490 | 0.2597 | 11.1763 | 0.3798 |
| Fee | | | | | -0.1135 | 0.4847 | 0.0600 | 0.7527 |
| Incent | | | | | 0.0127 | 0.3005 | 0.0053 | 0.7366 |
| Live | | | | | -0.2937 | 0.1656 | -0.2086 | 0.4197 |
| ln(Size) | | | | | 0.0025 | 0.9082 | -0.0220 | 0.3594 |
| Audit | | | | | -0.1846 | 0.5595 | -0.7472 | 0.0212 |
| Age | | | | | 0.0035 | 0.1552 | 0.0016 | 0.5121 |
| ln(wait) | | | | | 0.1077 | 0.0672 | 0.0903 | 0.0181 |
| LR statistic | | 8.9597 | | 3.4460 | | 15.2337 | | 13.7187 |
| Probability(LR stat) | | 0.0113 | | 0.1785 | | 0.1238 | | 0.1862 |
| McFadden R-squared | | 0.0113 | | 0.0029 | | 0.0177 | | 0.0217 |
| # obs | | 3,649 | | 3,649 | | 2,058 | | 2,058 |
| # obs red-flagged | | 177 | | 138 | | 110 | | 73 |
| Frequency | | 0.0485 | | 0.0378 | | 0.0534 | | 0.0355 |

## Endnotes

[1] See, for example, "The care and feeding of managers," (New York Times, July 8, 2001, p.3.13), and "Hedge funds snag London traders," (Wall Street Journal, June 14, 2004, p.A14).
[2] See "The sleaziest show on earth," (Forbes, May 24, 2004, p.110).
[3] The SEC (2004), for example, emphasizes that valuation problems often arise "when hedge fund advisers overstate assets in order to cover trading losses." Asness et al. (2001) suggest some hedge fund managers smooth returns conditional on performance and examine the impact on observed lagged market exposure. Chandar and Bricker (2002) study a similar issue in a multi-period model of incentives for closed-end mutual fund managers. In the model, managers are predicted to overvalue illiquid securities when the return of liquid assets falls slightly below a benchmark, and to undervalue illiquid securities when the return of liquid assets is extremely high or low.
[4] The authors thank Stephen Brown, the editor, for suggesting this example.
[5] Similarly, Burns and Kedia (2006) report that 93% of restated earnings of S&P 1,500 companies from 1995 – 2002 involved earnings that were originally overstated.
[6] Ackerman et al. (1999) find a similar result.
[7] See, for example, "Hedge fund values: Stop the fudging," (Business Week, May 20, 2004, p.106).
[8] High water marks in managerial incentive contracts could be modeled by a path-dependent hurdle return $c$.
[9] By construction, the residuals of the smoothing regressions are serially correlated. Consequently, OLS coefficient estimates may be biased since the lagged dependent variable is the regressor. Greene (2003) derives the bias and shows that it is an increasing function of the serial correlation in the residual. We derive analytic expressions for the serial correlation in the residuals of equations (4) and (7), available upon request. Figure 2 graphs the first order serial correlation in the residuals of the conditional model in equation (7) as a function of the smoothing coefficients. For the relevant regions of the parameter space, the residual serial correlation is quite small. Simulations confirm that the regression's estimated coefficients are quite close to analytical values. The residual serial correlation and resulting bias are even smaller in the unconditional model in equation (4). Therefore, we estimate coefficients using OLS to avoid numerical optimization.
[10] A detailed proof is available from the authors.
[11] In contrast, Brown et al. (2004) document a concave relation between the returns of Australian equity funds and an equity index. However, as the authors point out, it is unclear whether their result is relevant to the case of hedge funds, which operate with much less oversight and transparency.
[12] A detailed proof is available from the authors.
[13] Fung and Hsieh (2004) acknowledge this issue and model time-variation in risk factors by searching for breakpoints in the data indicating significant changes in factor loadings.
[14] http://faculty.fuqua.duke.edu/~dah7
[15] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.
[16] In particular, we run OLS, save the residuals, and regress them on their lag to estimate the serial correlation $\rho$ in residuals. Lastly, we transform the original observations by subtracting $\rho$ times their lag, and rerun OLS.
[17] In contrast, Sharpe (1992) finds style factors can explain 90% of the variation in returns for equity mutual funds.
[18] Getmansky et al. (2004) report adjusted $R^2$ of 17.7% in a cross-sectional regression of fund smoothing coefficients on a number of descriptive indicator variables.
[19] Chevalier and Ellison (1997) find that the flow-performance relation is strongest for young mutual funds.
[20] Endogeneity may be a concern for the cash flow variables. For robustness, we employ two stage methods. Instruments for the cash flow variables include the market return, the variability of the market's return, and fund performance measures. Results are qualitatively unchanged.
[21] King and Zeng (1999) perform simulations to examine the performance of logit models in rare event analyses. When the dependent variable of the logit model takes the value of one only for a small fraction of the sample, the estimated probability of the event is downward biased. We use the following software: Michael Tomz, Gary King, and Langche Zeng. 1999. RELOGIT: Rare Events Logistic Regression, Version 1.1 Cambridge, MA: Harvard University, October 1, http://gking.harvard.edu