Mind the Gap: Unmapped Holdings and the Performance of U.S. Equity Mutual Funds

David L. Hunter

February 13, 2008

Abstract

This paper investigates a newly available dataset of assets that are held by U.S. equity mutual funds, but are not U.S. equities ('unmapped holdings'). Past research has generally disregarded these holdings and assumed their existence to be inconsequential. This paper shows their widespread presence and investigates how they are used within mutual fund portfolios. Funds can be classified by their unmapped holdings characteristics. Unmapped holdings have a statistically significant effect upon portfolio risk and return, and actively managed funds tend to use unmapped holdings to reduce portfolio volatility while index funds show marginally increased volatility. Unmapped holdings represent an unmodeled risk factor in portfolio regressions, and this paper tests for mutual fund return predictability while controlling for unmapped holdings related risks. Return predictability is improved when unmapped holdings characteristics are controlled.

1 Introduction

A majority of U.S. equity mutual funds do not restrict their investments exclusively to U.S. equities. In fact, only 16.6% of domestic equity mutual funds are actually invested with 100% 'domestic equity'. What remains is a pool of investments that includes cash, international equities, bonds, convertibles and preferreds, and even derivatives. Despite this widespread practice, researchers have typically performed their analysis of these funds as if all funds were fully invested in U.S. equities.

'Unmapped holdings' or 'unmapped assets' are defined herein as any assets held by U.S. equity mutual funds that are not U.S. equities. By this definition, cash, commercial paper, bonds, futures, options, warrants, preferreds, and international stocks and bonds all qualify as 'unmapped holdings'.

The reason for mutual funds to hold unmapped holdings is not obvious. Arguments for their existence could include diversification that can alter a fund's reward to risk ratio (Chevalier[6]), liquidity that reduces the cost of unpredictable investment flows (Edelen[8]), illiquidity that facilitates a fund's ability to manipulate prices (Carhart et al.[5]), leverage, and return enhancement through mispricing opportunities. Although the purpose for holding unmapped assets is not distinguished in this paper, funds are grouped into three general styles of unmapped holdings management. Given the designation that these funds are 'domestic equity' mutual funds, it is unclear if unmapped holdings should exist within these funds at all.

A domestic equity fund's use of unmapped holdings is not prohibited by U.S. securities law. Funds that invest in unmapped holdings clearly represent this fact in their

prospectus. Likewise, unmapped holdings information has been regularly disclosed for decades together with mapped holdings data in prospectus filings and/or regulated holdings disclosures to the Securities and Exchange Commission. Despite this fact, access to this data has been difficult in the past to obtain and holdings datasets such as Thompson/CDA limited their data exclusively to long-only positions of U.S. equity assets. Works such as Grinblatt and Titman[14][13], Daniel et al.[7], and Wermers[23] were all constrained in their analysis to this filtered dataset. This paper uses the newly available mutual fund holdings database in CRSP that provides more extensive information about unmapped holdings in mutual funds. Holdings such as derivatives, bonds, international stocks, convertibles, and preferreds are present in this database, although they still are not mapped into any other securities datasets.

Diversity in the way funds manage unmapped holdings can obscure performance comparisons between funds. Consider the tools an investor may use to evaluate the performance of two funds. One fund invests exclusively in U.S. equities, while the other holds 90% of its portfolio value and the remaining 10% in unmapped holdings. The investor may choose to construct an estimate of Jensen's α using a factor model such as the CAPM, the Fama-French 3 factor model, or the Carhart 4-factor model. With any of these regressions, one assumes U.S. equity related factors to be the only relevant factors in the analysis. For the first fund this estimate wpuld be appropriate, but the same regression in the second fund would suffer from estimation error due to omitted variable bias. Although the direction of the bias is unclear, the estimate may be too high or low depending upon the additional factors introduced by unmapped assets. Alternatively, suppose the investor employs a simple performance-to-benchmark

return comparison as is common in industry. Again the effects of unmapped holdings may positively or negatively bias inferences about the fund manager's skill. If the investor evaluates performance based upon characteristic based benchmarks (Daniel et. al. (1997[7])), or uses Grinblatt and Titman's performance without benchmarks (Grinblat and Titman (1993)[13]), the effects of unmapped holdings upon performance are ignored entirely. Even comparisons of each fund's Sharpe ratio can be altered through a fund's use of unmapped holdings.

In this paper, mutual funds are categorized by their use of unmapped holdings. This categorization identifies distinct types of funds with clear differences in the effects of the unmapped holdings they contain. This categorization is used to show that unmapped holdings generally have a statistically significant effect upon portfolio risk and returns. This categorization also produces a clear distinction between actively managed funds and index funds.

The economic significance of unmapped holdings to mutual fund risk and return has rarely been questioned. Most studies have assumed unmapped assets to be insignificant. In one exception, Koski and Pontiff[17] investigate the risk and return characteristics of funds that use derivatives and argue that those funds are indistinguishable from funds that do not use derivatives. Their conclusion is drawn from telephone survey data where funds were asked to disclose their derivatives use.

Understanding the role of unmapped holdings in mutual fund portfolios can be instrumental in understanding the viability of active management. Wermers[23] indirectly calculates the average return of unmapped holdings using portfolio returns and mapped holdings. He infers that the average mutual fund loses 70 basis points per year

due to investment in unmapped holdings. He shows this after attributing mutual fund returns to stock characteristics, asset selection, transaction costs, and fees. His results suggest that if managers had exclusively held U.S. stock holdings, then their asset selection and asset characteristics would very nearly offset the expenses and fees they incur. Instead, the performance costs of unmapped holdings make active management more costly than passive alternatives. His result seems to suggest that somehow fund managers skillfully invest in U.S. equities, but then destroy this through their use of unmapped holdings. From our classification of funds by unmapped holdings, we find that active managers tend to reduce portfolio volatility using unmapped holdings while index funds marginally increase in portfolio volatility. This fact allows a different interpretation of the Wermers result: unmapped holdings reduce portfolio risk rendering a comparison between active management and index funds unfair.

The question of if fund managers have skill is a question dating back to Jensen[15]. Recent research has argued that active fund managers are successful at finding good investments. In addition to Wermers, Baks et al.[1] find that mean-variance investors who are skeptical about active management skills can identify mutual funds that generate ex-ante positive alphas. More recently, Kacperczyk, Sialm, and Zheng[16] also produce a measure called the return gap that they show has strong ability to predict fund returns due to unobserved manager actions.

If many active fund managers use unmapped holdings to reduce their portfolio risk, a direct return comparison between active and passive management styles is misleading. Likewise if unmapped holdings are not represented by U.S. equity risk factors, then a regression that uses only U.S. equity risk factors would fail to adequately control for

the risk effects of unmapped holdings. In this paper, a technique by Hunter, Kandel, Kandel, and Wermers is applied to control for unmodeled risk factors contained within unmapped holdings and this technique is used to investigate predictability of mutual fund performance when these factors are controlled.

This investigation of mutual fund performance predictability will consider the recent article by Kacperczyk, Sialm, and Zheng[16]. Kacperczyk et al. used domestic equity mutual funds to investigate the return gap, a measure of the difference between reported fund returns and the return on a hypothetical portfolio that invests in previously disclosed portfolio holdings. They find that the return gap strongly predicts fund performance and they attribute the strength of this measure to evidence of manager skill. The Kacperczyk et al. paper is particularly relevant to our study of unmapped holdings because the size of the return gap is directly related to the portfolio weight that a fund invests in unmapped holdings. In other words, both a large positive and a large negative return gap coincide with a large investment in unmapped holdings. A fundamental assumption of the Kacperczyk et al. argument is that the return gap does not proxy for risk. Due to the types of assets contained within unmapped holdings, we propose that the return gap does alter risk, but this risk may not be well represented by standard U.S. equity factors. Kacperczyk et al. builds their inferences using portfolio sorts on a fund's past return gap. We find that a sort on the return gap ranks funds with large allocations to unmapped holdings into the extreme top or bottom ranks of the return gap. Because of the correspondence between unmapped holdings and the return gap, the predictability suggested by Kacperczyk et al. instead represents an allocation to unmodeled risk factors contained within the different ways mutual funds

use unmapped holdings.

To quantify unmapped holdings characteristics, implied measures of unmapped holdings return, volatility, and covariance with mapped holdings are derived. Mutual funds are then classified based upon their use of unmapped holdings. The statistics used for classification are the portfolio weight a fund allocates to unmapped holdings, the correlation of unmapped holdings to mapped holdings, and the direction that portfolio risk changes with the inclusion of unmapped holdings. This categorization strongly separates actively managed mutual funds from index mutual funds. The properties of the return gap and its consistencies with unmapped holdings are investigated, and a test for mutual fund performance predictability is constructed.

This paper will proceed as follows: section 2 describes the data and construction of endogenous benchmarks to control for unmodeled risk factors; in section 3, the variance of unmapped holdings and their covariance with mapped assets is derived; These measures are then used to distinguish mutual funds by their unmapped holdings characteristics; section 4 presents results; and section 5 concludes.

2 Data

The results of this paper are constructed using newly available holdings data in the CRSP mutual funds database. Since 2003, CRSP has been updating their database with mutual fund holdings provided to them by S&P. Holdings data are voluntarily provided by participating mutual funds to Standard & Poor's, which was recently acquired by Morningstar. CRSP obtains this data from Standard & Poor's and proceeds

to map equity holdings into their U.S. stock database. S&P has a relationship with the majority of the companies from which it collects its holdings data and they collect it regularly. In situations where CRSP does not obtain some information, they attempt to pursue it as may be necessary.

One particular advantage to this dataset is that it includes information on bond, international stock, and derivative holdings for a large number of funds. Such data has always been absent from the data in the Thompson/CDA dataset. This is one reason for the omission of unmapped holdings from much of the research in the past. An additional advantage to this new dataset is that holdings data are available at greater frequency than the typical semiannual mandatory reporting required by the SEC. The frequency of reporting varies across funds, but the new dataset typically provides a much denser view of mutual fund holdings over time.

One shortcoming of this dataset is that the data is voluntarily offered by each fund company and there is no assurance that holdings are 100% complete and accurate. During measurement and analysis of this dataset, one observes that statistical properties that are constructed from return data that are generally consistent with the holdings that are reported by the funds, such as correlations and variances. However, one finds instances where unmapped holdings statistics would imply that particular types of holdings may have been omitted from the data, particularly derivative positions. Although such funds may have constructed synthetic derivative exposure through trading, it is also possible that the unmapped asset data reflects a potentially smaller sample than what is actually allocated in the entire population of funds. Just the same, the sample of funds in this population that did report derivative holdings tend to show

variance and covariance statististics that would be consistent with such holdings, and therefore appear to have been reported correctly.

In a few instances CRSP incorrectly populated its dataset without identifying mapped and unmapped holdings differently. For this reason, all holdings data that showed a 100% investment in unmapped assets were omitted from the dataset.

This analysis uses quarterly holdings reports of funds that are defined as domestic equity mutual funds. The dataset is filtered using self-reported domestic equity investment styles. The criteria applied is generally consistent with the criteria applied in other studies of U.S. domestic equity mutual funds such as Kacperczyk et al[16]. One conventional criteria that is intentionally omitted is a filter which restricts the analysis to only funds where the aggregate market value of U.S. domestic equity assets is within a margin of error to the reported total net asset value of the funds. Such a criteria generally restricts analyses to include only funds that have a reported holdings market value within 10% or 20% of the reported net asset value. For this paper's purposes, this convention arbitrarily eliminates an important set of funds from the sample since these funds are likely to hold greater exposure to unmapped holdings. The findings of this paper are robust to the application of such a filter because a significant number of funds hold unmapped holdings within such arbitrary cutoff levels. In fact, the majority of mutual funds hold unmapped asset exposures within the typical cutoffs in this criterium. Table 1 shows that most funds hold unmapped assets in levels less than a typical10% filter restriction.

Table 1 presents summary statistics for funds with different percentage allocations to unmapped holdings. The first group of data, labeled "Extent of Unmapped Assets"

presents the percentage of the fund population that holds different market weights of unmapped assets. 16.6% of all mutual funds had no exposure to unmapped assets at least once during the measurement period, and the majority of funds (62.7%) held between 2 and 5% of their market value in unmapped assets at least once during the measurement period. On average, most funds (34.5%) held between 2 and 5% of their assets in unmapped holdings, but a substantial 18.12% (11.28% + 6.33% + 0.51%) of the fund population held over 10% of their assets in unmapped holdings on average.

One does not find substantial variation in fund characteristics as unmapped asset allocation varies. The second group of data in table 1 presents the characteristics of funds that have different allocations to unmapped assets. The smallest funds (avg. $1.43 mil) tend to hold between a 2 and 5% allocation to unmapped holdings, and the largest funds (avg. $54.07 mil.) tend to have the greatest allocation to unmapped assets. This relationship is not monotonic, as slightly larger funds (avg. $6.94 mil) also tend to hold no unmapped assets. The number of issues held by a fund does not appear to correlate with a fund's unmapped holdings allocation.

There is also little distinction in the asset investment across different allocations to unmapped assets. The third group of data in table 1 shows the average percentage market value allocation that funds invest in different types of unmapped assets. One observes generally increasing values in all of the asset types as the allocation to unmapped assets increases.

Unmapped assets are not distinguishable by investment styles. Table 2 presents the average unmapped holdings investments of mutual funds grouped by their self-declared investment style. Most notable is an absence of derivatives exposure by aggressive

growth and small cap value funds, but otherwise, there is almost no distinction to be made between the unmapped holdings allocation of funds in different styles.

Unmapped Holdings Summary Statistics

	Percentage Unmapped Holdings							
	0%	(0%, 1%]	(1%, 2%]	(2%, 5%]	(5%, 10%]	(10%, 20%]	(20%, 50%]	(50%, 100%)
Extent of Unmapped Assets								
Pct Funds (1 or more obs)	16.64%	2.96%	40.58%	62.69%	48.85%	27.05%	13.17%	1.79%
Pct Funds (Avg Wt)	1.43%	8.01%	15.01%	34.51%	22.92%	11.28%	6.33%	0.51%
Nbr Funds	28	157	294	676	449	221	124	10
Fund Characteristics by Unmapped Allocation								
Avg MV (mil)	6.94	5.48	3.32	1.43	2.18	4.58	16.02	54.07
Avg Issues Held	66.1	195.2	227.9	145.4	138.9	140.0	209.0	257.7
Avg Unmapped MV (mil)	0.000	0.035	0.052	0.047	0.152	0.657	4.519	34.604
Avg Unmapped Issues	0.0	4.1	5.0	4.7	7.9	15.3	49.0	124.7
Allocation								
Cash	0.00%	0.39%	1.05%	2.26%	4.06%	6.22%	6.97%	23.37%
Derivatives	0.00%	0.00%	0.02%	0.02%	0.14%	0.28%	0.99%	0.53%
Bonds	0.00%	0.08%	0.12%	0.33%	1.03%	2.35%	8.57%	19.75%
Foreign	0.00%	0.01%	0.01%	0.04%	0.13%	0.62%	1.08%	1.76%
Currency	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.08%
Warrant/Pfd	0.00%	0.00%	0.03%	0.03%	0.12%	0.17%	0.62%	0.64%
No ID	0.00%	0.16%	0.34%	0.63%	1.49%	4.73%	9.99%	17.87%

Table 1: This table presents summary statistics for funds with variousallocation weights to unmapped holdings. The first lines labeled "Extent of Unmapped Assets" presents the percentage of funds at each unmapped holdings allocation level. The.second group of data labeled "Fund Characteristics by Unmapped Allocation" presents market value and holdings characteristics for funds at each level of unmapped holdings. The third group of data, labeled "Allocation" presents the average allocation of funds to different types of unmapped assets.

U.S. Equity Mutual Fund Unmapped Holdings: Allocation Summary

	All Funds	AG	SCG	G	GI	SCV
Total Funds	2583	40	78	96	132	22
Avg. Assets (bil)	$13.995	$35.781	$26.025	$69.090	$62.461	$16.761
% Cash	3.21%	3.28%	3.81%	2.45%	2.74%	3.52%
% Derivatives	0.08%	0.00%	0.03%	0.03%	0.04%	0.00%
% Bonds	2.72%	0.89%	2.41%	3.44%	4.75%	2.22%
% Foreign	0.80%	0.49%	0.99%	0.87%	1.07%	0.44%
% Currency	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
% Warrants/Pref	0.14%	0.15%	0.12%	0.19%	0.16%	0.06%
% No ID	5.12%	4.67%	5.42%	6.98%	4.72%	1.15%

Table 2: This table presents the average allocation that mutual funds in unmapped assets. Unmapped assets were identified using name, maturity/expiration date, coupon, and any other identifiable characteristics available. A subset of these funds provided no information other than asset name and thus could not be distinguished between bonds, derivatives, or foreign stocks, such assets are classified as 'no ID' in the last line of the table.

This analysis spans mutual fund holdings report dates from 9/30/2003 through 6/30/2006. Table 1 reports that 1,931 out of 1,959 funds average a non-zero exposure to unmapped holdings, representing 98.6% of the sample population. As the sample is restricted to a greater allocation in unmapped holdings, the percentage expectedly declines. Even so, a substantial number of funds hold an allocation of 10% or more of unmapped assets in their portfolio. Differences between funds are not well distinguished across fund characteristics and styles.

3 Methodology

The research methods in this paper cover two primary components: fund characterization by their unmapped holdings commonalities, and performance predictability with the commonalities controlled.

A desirable characterization of funds by their unmapped holdings should demonstrate the following characteristics: first, it should be empirically tractable by allowing simple measures to distinguish different groups of funds; second, it should be extendible into past data when unmapped holdings are not observable; third, it should be consistent with portfolio assets; fourth, it should demonstrate economically significant differences between groups of funds.

To meet these conditions, I apply three statistics that capture the role of unmapped assets in mutual fund portfolios: the correlation between mapped and unmapped assets, the percentage market value invested in unmapped holdings, and the change to portfolio standard deviation attributable to unmapped holdings. In section 3.1, I use a simple two stock portfolio framework where no trading occurs between reporting dates to define the implied correlation between mapped and unmapped assets. Section 3.2 uses the implied correlation of section 3.1, the percentage invested in unmapped holdings, and the change to portfolio standard deviation to classify funds by their unmapped holdings exposure.

A complexity in this analysis is the influence of unobserved actions upon the statistical measures. Since fund managers can trade throughout each measurement period, and because returns of unmapped holdings are not observed, the effects of both unmapped holdings and the unobserved actions of fund managers are intertwined. The results of section 3.1 are generalized to a framework that allows for unobserved trading in the appendix. The appendix derivation defines explicitly the dynamics of both effects. The primary implication of the appendix results is that the effects of unobserved actions are most amplified when the allocation to unmapped holdings are small.

Clearly, implications due to unmapped holdings are strongest among funds with sizeable investment in unmapped holdings. When there is a small allocation to unmapped holdings, distinction between unmapped holdings and unobserved actions is non-trivial.

By controlling for unmapped holdings characteristics, one can infer different conclusions about the predictability of mutual fund performance than has been previously claimed. Since the performance of unmapped holdings cannot be observed between reporting periods, we define an endogenous measure of unmapped holdings performance in section 3.3. The section concludes by describing how the test for predictable mutual fund performance is implemented.

3.1 A Two Stock Portfolio

Unmapped portfolio holdings can be most generally considered as a component in a two stock portfolio. Assume that a portfolio is constructed at time t, and that it cannot be traded again until time $t+1$.

Mapped holdings observed at time t generate subsequent returns that can be observed at time $t+1$ which will be denoted as $r_{M,t+1}$. Unmapped holdings are observed at time t and at time $t+1$, but prices are costly to observe and hence cannot be directly measured for risk or returns. Measurable portfolio holdings (mapped holdings) will be denoted as component M and the immeasurable portfolio set (unmapped holdings) as component U. The expected return of a portfolio is then defined as the weighted average expected return of the two components:

$$E(r_p) = w_M E(r_M) + w_U E(r_U)$$

The variables w_M and w_U represent the percentage invested in the measurable and immeasurable portfolio components respectively.

The expected return of the portfolio will be denoted as $E(r_p) = \mu_p$, the expected return of the measurable component as $E(r_M) = \mu_M$, and the expected return of the immeasurable component as $E(r_U) = \mu_U$. Likewise, denote the variance of the portfolio as σ_p^2, the variance of the measurable component as σ_M^2 and the variance of the immeasurable component as σ_U^2.

One may take advantage of an indirect measure of the return on unmapped holdings, the return gap, because it represents an alternative observable measurement of portfolio return and volatility. The return gap is defined in Kacsperczyk et. al. as the difference between the gross portfolio return and the return of mapped mutual fund holdings. The return gap is represented in the following relation:

$$r_g = r_p - r_M$$

The return gap r_g represents a long-short portfolio constructed from a long position in the actual mutual fund portfolio and a short position in a hypothetical portfolio that is constructed from mapped portfolio holdings. In this simplified one period framework, only the returns of unmapped holdings can influence the return gap. Since $w_M = 1 - w_U$, the expected value of the return gap $E(r_g)$ is equal to $w_U(\mu_U - \mu_M)$ and the variance of the return gap σ_g^2 is equal to $w_U^2(\sigma_U^2 + \sigma_M^2 - 2\sigma_{MU})$ where σ_{MU} represents the covariance between the measurable and immeasurable portfolio components. One can observe that both the return gap and its variance are directly proportional to

weight of a fund's investment in unmapped holdings.

Since unmapped holdings are not observable, one may derive from observed data implied values of σ_U^2 and σ_{MU} which are obtained in proposition 3.1:

Proposition 3.1 *The variance and covariance of unmapped holdings:*

1. *The variance of unmapped holdings or unobserved actions (σ_U^2) is equal to:*

$$\sigma_U^2 = \frac{\sigma_g}{w_U^2} - \sigma_M^2 + 2\sigma_{UM} = \frac{w_U\sigma_p^2 + w_M\sigma_g^2 - w_M w_U \sigma_M^2}{w_U^2}$$

2. *The covariance of the unmapped holdings (σ_{UM}) is equal to:*

$$\sigma_{UM} = \frac{\sigma_p^2 - \sigma_g^2 + (w_U - w_M)\sigma_M^2}{2w_U}$$

Proof *These results are obtained by first solving part 2 for σ_{UM}, and then using σ_{UM} to obtain σ_U^2 in part 1.*

Part 2: One derives σ_{UM} by recognizing that the difference $(\sigma_p^2 - \sigma_g^2)$ produces a term where σ_U^2 cancels. All that remains is an equation with only one unknown: σ_{UM}. Simple algebraic manipulation then produces the result.

Part 1: Obtained by direct substitution of σ_{UM} into the variance equation for σ_g^2 and rearranging the result. ∎

The advantage of the equations in proposition 3.1 is that both σ_U^2 and σ_{UM} can be computed using directly observable values: σ_p^2, σ_g^2, σ_M^2, w_U, and w_M. In the absence of unobserved actions, proposition 3.1 suggests that one can infer the volatility and covariance of unmapped holdings. In the appendix, this result is generalized to show that these relations can also be interpreted under a portfolio with any combination of unmapped holdings and unobserved actions. The main difference between proposition 3.1 and the generalized result in the appendix is that the implied variance of unmapped holdings (σ_U^2) and the implied covariance between mapped and unmapped holdings (σ_{UM}) have the same form as proposition 3.1, but also contain additional error due

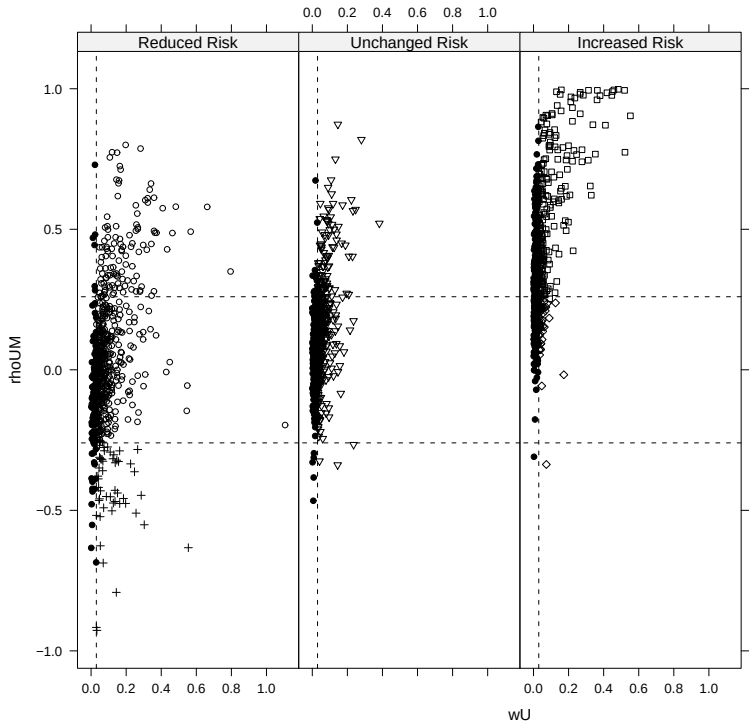

Figure 1: **Unmapped Holdings by Characteristics:** The above graphs illustrate differences between funds according to their unmapped holdings characteristics. Funds are distinguished by how unmapped holdings influence portfolio risk, and then the respective groups plotted to show the correlation between mapped and unmapped holdings (ρ_{UM}) as a function of allocation to unmapped holdings (w_U). In each plot, graph symbols represent subgroups of funds that are constructed using these three measures ($\Delta\sigma_p, \rho_{UM}, w_U$).

to unobservable trading throughout the period. The empirical separation of these two effects is non-trivial.

3.2 Mutual Fund Classification by Unmapped Holdings

The results of proposition 3.1 provide one statistical measure that can be used to classify mutual funds by their unmapped holdings: the correlation between mapped and unmapped assets ($\rho_{UM} = \frac{\sigma_{UM}}{\sigma_U \sigma_M}$). In addition to that measure, one may also

distinguish funds by how unmapped holdings alter portfolio risk ($\Delta\sigma_p$) and how much portfolio weight a fund invests in unmapped holdings (w_U). In this section, funds are characterized using these three measures and the validity of this characterization is demonstrated.

The effect of unmapped holdings upon portfolio risk ($\Delta\sigma_p$) is obtained by calculating the daily return volatility of every mutual fund portfolio (σ_p) and the daily return volatility of every fund's mapped holdings (σ_M) in three month measurement windows subsequent to each holdings report date. I define risk dummy variables as *risk decreasing* ($\Delta\sigma_p < 0$) if σ_p was less than σ_M at least 70% of the time, *risk increasing* ($\Delta\sigma_p > 0$) if σ_p was less than σ_M less than 30% of the time, and *no change* ($\Delta\sigma_p \approx 0$) for all remaining funds.

Figure 1 plots funds by these three characteristics to illustrate how they separate groups of funds. By construction, funds with the greatest allocation to unmapped assets are most heavily represented among funds that show strongly altered risk (either increasing or decreasing). Funds with the greatest positive correlation between mapped and unmapped assets appear among funds where unmapped holdings either increase or leave unaltered portfolio risk. Funds with lower correlation between mapped and unmapped assets appear among funds where unmapped assets generally decrease portfolio risk. A substantial number of funds appear to have a strong negative correlation between unmapped holdings and mapped holdings ($\rho_{UM} < -0.25$) and a substantial allocation to unmapped holdings ($w_U > 3\%$), which results in a net decrease to portfolio risk ($\Delta\sigma_p$).

Portfolios with a small allocation to unmapped holdings will generate characteristics

that are less attributable to unmapped holdings. Funds with a small w_U are more sensitive both to unobserved actions by fund managers and to the random effects of any single asset. For this reason, funds with less than a 3% investment in unmapped holdings are considered as a single group of funds in this analysis, and this group is generally excluded from inferences.

Using the three statistics $(\Delta\sigma_p, \rho_{UM}, w_U)$ and figure 1, one may define five groups of funds:

- Group I *(Diversification)*: $\Delta\sigma_p < 0, \rho_{UM} \geq -0.25, w_U \geq 3\%$
- Group II *(Aggressive Diversification)*: $\Delta\sigma_p < 0, \rho_{UM} < -0.25, w_U \geq 3\%$
- Group III *(Replication)*: $\Delta\sigma_p \approx 0, w_U \geq 3\%$
- Group IV *(Allocation)*: $\Delta\sigma_p > 0, \rho_{UM} \geq 0.25, w_U \geq 3\%$
- Group V *(Minimization)*: $\Delta\sigma_p > 0, \rho_{UM} < 0.25, w_U \geq 3\%$
- Group VI *(Minimization)*: $w_U < 3\%$

To validate this segmentation of funds, table 3 presents the allocation characteristics of each of these groups. Characterization by risk effects $(\Delta\sigma_p)$ shows very different bond allocation between risk reducing fund groups I and II $(\Delta\sigma_p < 0)$, and either the risk unchanged group III $(\Delta\sigma_p \approx 0)$ or risk increasing groups IV and V $(\Delta\sigma_p > 0)$. Characterization by the correlation between mapped and unmapped assets (ρ_{UM}) shows strong differences between groups I and II and between groups IV and V. Group I's holdings show a substantially higer allocation to non-identifyable assets (generally international stocks and/or bonds) and positive derivative exposure while group II holdings show negative derivative exposure. Group IV funds have higher deriva-

tive exposure (generally futures contracts), while group V funds hold the least cash. Categorization by percentage allocation to unmapped assets (w_U) distinguishes funds that hold substantial exposure to unmapped holdings from funds where the effects of unmapped holdings are less distinguishable from unobserved actions of managers, thus removing group VI from the remaining groups.

A more intuitive description of each of the fund groups can be drawn from table 3. Group I *(Diversification)* funds generally diversify their portfolios through sizeable bond and stock investments. They tend to use currency contracts which likely hedge their foreign currency exposure. Group II *(Aggressive Diversification)* funds also diversify their portfolio risk, but do so using more sophisticated derivatives techniques to accompany their bond and cash holdings. Covered call strategies or market exposure change through derivative holdings are components of funds within this group. Group III *(Replication)* funds do not show any substantial management of unmapped assets, nor do they take as large positions in unmapped assets (as seen in figure 1). Funds in this group do not appear to use their unmapped holdings for any meaningful portfolio strategy. Group IV *(Allocation)* funds show the largest 'market value' exposure to derivative contracts. Derivatives held in this group of funds are generally index futures contracts (not shown). Funds in this group have minimal bond exposure. Group IV funds appear to use futures contracts to incorporate index returns into the more liquid cash holdings of the portfolio, thereby permitting the funds portfolios to hold greater allocations to cash. Groups V *(Minimization)* funds have the smallest use of derivatives and the smallest allocation to cash of all of the fund groups. Group V funds appear to minimize their exposure to unmapped holdings, and thus are grouped together with

Group VI funds that have less than 3% of their assets allocated to unmapped holdings.

Groups IV's larger portfolio risk (σ_p) to asset risk (σ_M) may be explained in two ways:

- Funds in this group (including index funds) tend to use cash and futures as their unmapped holdings. Any difference in value between cash and futures positions may bias to greater futures exposure. Such a bias would induce a slight amplification in fund returns.
- Daily pricing by funds in this group (including index funds) may be based upon futures or exchange for physical pricing, instead of asset market pricing due to the potential influence of stale prices. Because σ_M represents asset market volatility, this would implicitly lower than the portfolio volatility reported by funds that do so. Index funds are well suited to such pricing conventions.

Regardless of the mechanism, risk attributable to unmapped holdings is distinct between funds in Group IV and funds in other groups. Furthermore, the larger proportion of index funds within this group is substantial.

To validate the economic significance of differences between these groups, return correlations were analyzed between groups average returns and between funds within each group. This analysis is contained in appendix B, and shows that after controlling for the common risk factors of the market, book-to-market, capitalization, and momentum, group residuals have a very low correlation across groups while the correlation of fund residuals between funds within each group are much higher.

U.S. Equity Mutual Funds Unmapped Holdings Allocation Summary
By Classified Subgroup

	Group I Div.	Group II Agg. Div.	Group III Repl.	Group IV Alloc.	Group V Min.	Group VI Min.
Total Funds	482	54	357	209	46	796
Avg. Assets (bil.)	22.65	6.19	13.14	14.23	6.04	11.97
% Cash	4.28%	8.92%	4.02%	3.59%	2.61%	1.12%
% Derivatives	0.06%	-0.11%	0.04%	0.30%	0.01%	0.05%
% Derivative Issues	0.21%	1.64%	0.16%	0.16%	0.14%	0.15%
% Bonds	4.95%	8.21%	1.47%	0.70%	0.71%	0.26%
% Foreign	0.87%	0.34%	0.95%	0.75%	0.19%	0.19%
% Currency	0.03%	0.00%	0.01%	0.00%	0.00%	0.00%
% Warrants/Pref	0.51%	0.34%	0.16%	0.08%	0.11%	0.07%
% No ID	9.12%	2.61%	3.34%	3.99%	2.96%	1.08%

Table 3: This table presents the unmapped holdings allocation statistics for the fund subgroups defined by the parameters $(\Delta\sigma_p, \rho_{UM}, w_U)$. All values are presented as percentage of market value except where noted. Unmapped assets were identified using name, maturity/expiration date, coupon, and any other identifiable characteristics available. A subset of these funds provided no information other than asset name and thus could not be distinguished between bonds, derivatives, or foreign stocks, such assets are classified as 'no ID' in the last line of the table.

Difference in Mean T-Statistics

Cash

	Group I	Group II	Group III	Group IV	Group V	Group VI
Group I	NA	-1.16	6.45***	2.56**	6.68***	15.81***
Group II	1.16	NA	3.60***	2.44**	4.39***	6.61***
Group III	-6.45***	-3.60***	NA	-1.91*	2.44**	15.69***
Group IV	-2.56**	-2.44**	1.91*	NA	3.28***	8.28***
Group V	-6.68***	-4.39***	-2.44**	-3.28***	NA	4.60***
Group VI	-15.81***	-6.61***	-15.69***	-8.28***	-4.60***	NA

Derivatives

	Group I	Group II	Group III	Group IV	Group V	Group VI
Group I	NA	0.69	-1.08	-4.18***	-0.65	1.41
Group II	-0.69	NA	-0.93	-3.62***	-0.86	-0.51
Group III	1.08	0.93	NA	-3.92***	0.16	2.10**
Group IV	4.18***	3.62***	3.92***	NA	3.92***	4.35***
Group V	0.65	0.86	-0.16	-3.92***	NA	1.34
Group VI	-1.41	0.51	-2.10**	-4.35***	-1.34	NA

Bonds

	Group I	Group II	Group III	Group IV	Group V	Group VI
Group I	NA	-2.02**	5.98***	2.17**	8.22***	9.56***
Group II	2.02**	NA	3.64***	2.75***	4.25***	4.36***
Group III	-5.98***	-3.64***	NA	-2.95***	3.50***	5.48***
Group IV	-2.17**	-2.75***	2.95***	NA	5.08***	5.93***
Group V	-8.22***	-4.25***	-3.50***	-5.08***	NA	0.62
Group VI	-9.56***	-4.36***	-5.48***	-5.93***	-0.62	NA

Foreign Assets

	Group I	Group II	Group III	Group IV	Group V	Group VI
Group I	NA	0.69	2.88***	3.80***	7.85***	8.07***
Group II	-0.69	NA	0.65	1.07	2.18**	2.15**
Group III	-2.88***	-0.65	NA	0.90	4.13***	4.19***
Group IV	-3.80***	-1.07	-0.90	NA	2.98***	3.00***
Group V	-7.85***	-2.18**	-4.13***	-2.98***	NA	-0.36
Group VI	-8.07***	-2.15**	-4.19***	-3.00***	0.36	NA

Warrants/Preferreds

	Group I	Group II	Group III	Group IV	Group V	Group VI
Group I	NA	0.38	2.66***	3.70***	2.41**	3.79***
Group II	-0.38	NA	1.62	2.40**	1.63	2.44**
Group III	-2.66***	-1.62	NA	2.49**	0.39	2.93***
Group IV	-3.70***	-2.40**	-2.49**	NA	-0.73	0.06
Group V	-2.41**	-1.63	-0.39	0.73	NA	0.77
Group VI	-3.79***	-2.44**	-2.93***	-0.06	-0.77	NA

Not Identified

	Group I	Group II	Group III	Group IV	Group V	Group VI
Group I	NA	3.10***	2.91***	-2.30**	0.03	9.49***
Group II	-3.10***	NA	-0.80	-4.17***	-0.91	3.08***
Group III	-2.91***	0.80	NA	-3.99***	-0.68	5.90***
Group IV	2.30**	4.17***	3.99***	NA	0.95	6.94***
Group V	-0.03	0.91	0.68	-0.95	NA	1.73*
Group VI	-9.49***	-3.08***	-5.90***	-6.94***	-1.73*	NA

Table 4: This table shows the t-statistics of differences in mean allocation percentages of funds grouped by unmapped holdings characteristics ($w_{unmap}, \Delta\sigma_p, \rho_{unmap,map}$). Groups I and II represent the diversification groups, groups III and IV represent allocation groups, and groups V and VI represent minimization groups.

3.3 Predictability of Mutual Fund Returns and Endogenous Benchmarks

To control for unmapped asset exposure in predictability tests, one can construct group endogenous factors following the technique of Hunter et al. as follows:

1. Construct the average group return (r_g)
2. Remove known risk factor exposures: market, book-to-market, capitalization, and momentum.
3. Define endogenous group risk factor y_g as the group residuals plus its intercept.

In this paper, predictability of mutual fund returns is tested by constructing a trading strategy based upon past mutual fund performance. The three measures evaluated here are the return gap, the 4-factor model alpha, and the 4-factor plus endogenous benchmark alpha.

Section 4 will first show that actively managed mutual funds and index funds are distinct when classified by the use of unmapped holdings. The performance implications for both are described. Afterwards, the predictability of mutual fund performance is tested when controlling for unmapped holdings as described previously.

4 Results

Classification by unmapped holdings produces the following implications:

- Unmapped holdings distinguish funds by risk and return characteristics and span each of the commonly known investment styles.

- Unmapped holdings distinguish actively managed mutual funds from index funds. This difference is particularly meaningful when comparing the performance of the two. Actively managed mutual funds on average perform just as well as index funds when the risk-adjusted effects of unmapped holdings are considered in the comparison.
- The return gap is an inferior predictor of fund performance.
- Unmapped holdings as a risk factor improves predictability of mutual fund performance, but only among funds where common equity factors do not fit as well.

The following sections will discuss the evidence in support of these implications.

4.1 Unmapped Holdings Management

The classes of funds described in section 3.2 identified substantial differences in how funds use unmapped holdings. Statistical differences among subgroups of funds can imply different purposes for managing unmapped holdings.

Table 6 presents risk and return characteristics between the different groups of funds. The first column of table 6, labeled 'All Portfolios', presents return gaps ($r_p - r_M$), volatility differences between portfolio returns and mapped holdings returns ($\sigma_p - \sigma_M$), and differences in Sharpe ratios between the portfolio and a hypothetical mapped holdings portfolio.

In all portfolios, one observes a mean daily return gap of -0.75%. This is consistent with the conclusions of Wermers[23] that unmapped holdings tend to reduce portfolio returns by an average of 0.70% annually. Unmapped holdings also reduce portfolio

volatility by an average of 6 basis points per year. Unmapped holdings reduce the sharpe ratio for all mutual funds on average by 23 basis points per unit of standard deviation per year.

An interesting pattern appears when these averages are segmented into the different groups of funds that manage unmapped holdings distinctly (see Table 6). Diversification groups I and II demonstrate the most substantial loss in annual performance. These funds suffered an average performance reduction of 122 and 150 basis points respectively. Likewise, these same groups also benefitted with the greatest reduction in portfolio volatility, with a 101 a nd 276 basis point reduction, respectively. Diversification fund groups consequently have the best return to risk ratio over any other groups of funds.

Allocation group IV benefits from the smallest decline in performance due to unmapped holdings, earning a loss of only 21 basis points per year. However, they also incur the greatest increase in portfolio volatility at 147 basis points per year. Consequently, funds in this group suffer the greatest loss in their return-to-risk ratio with a decline of 52 basis points a year per unit of standard deviation.

The remaining three groups, group III (Replication) and groups V and VI (Minimization), have average to slightly better than average performance losses and mid-range increases in portfolio risk due to unmapped holdings. On average, funds in these groups tend to also have mid-range decreases in their reward to risk ratio. Minimization groups V and VI incur roughly average performance losses due to unmapped holdings.

One interpretation of these results is that management of unmapped holdings can create product differentiation within management asset classes and styles. Funds can

tailor their returns to different investor types through their management of unmapped holdings. Since many investors invest in funds based upon past performance (Chevalier and Ellison 1997[6], Berk and Green 2004[2]), those investors would tend to favor funds in the allocation group IV. Alternatively, mean-variance type investors who consider return-to-risk or Sharpe ratio type measures would tend to favor funds in diversification groups I and II. This inference is supported by the dispersion of unmapped holdings groups across mutual fund investment styles. table 5 shows that unmapped holdings groups effectively span each of the common investment styles of mutual funds.

Unmapped Holdings Group Span of Style

Group	AG	G	GI	SCG	SCV
I	11.4%	18.8%	38.7%	26.6%	29.4%
II	5.7%	0.0%	2.8%	0.0%	11.8%
III	40.0%	20.0%	17.9%	20.3%	11.8%
IV	0.0%	16.3%	2.8%	14.1%	11.8%
V	5.7%	5.0%	0.9%	4.7%	5.9%
VI	37.1%	40.0%	36.8%	34.4%	29.4%

Table 5: This table presents the distribution of unmapped holdings groups across conventional mutual fund investment styles. Numbers are represented as a percentage of funds in each style.

Contributions to Portfolio Risk and Returns

	All Portfolios	Group I Div.	Group II Agg. Div.	Group III Repl.	Group IV Alloc.	Group V Min.	Group VI Min.
N	1944	482	54	357	209	46	796
$r_p - r_M$	-0.748% (-167.49***)	-1.215% (-151.50***)	-1.497% (-33.17***)	-0.421% (-40.61***)	-0.209% (-12.01***)	-0.516% (-11.10***)	-0.703% (-115.50***)
$\sigma_p - \sigma_M$	-0.061% (-29.15***)	-1.008% (-242.96***)	-2.755% (-211.48***)	0.014% (2.87***)	1.466% (227.22***)	1.004% (69.90***)	0.220% (66.70***)
$\Delta \frac{r}{\sigma}$	-0.239% (-6.46***)	-0.074% (-0.80)	0.978% (2.54***)	-0.182% (-2.53***)	-0.520% (-4.93***)	-0.262% (-0.74)	-0.371% (-8.32***)

Table 6: This table presents the mean return gap ($r_p - r_M$), the change in portfolio volatility ($\sigma_p - \sigma_M$), and the change in the Sharpe ratio between the portfolio and mapped holdings as a hypothetical portfolio ($\Delta \frac{r}{\sigma}$). The first column represents all portfolios combined and subsequent columns correspond with individual groups of funds based upon their unmapped holdings characteristics. Returns are annualized in percent, as are standard deviations.

4.2 Active Management vs. Indexing

Unmapped holdings differences also demonstrate important distinctions between actively managed funds and index funds. In particular, index funds are generally uniform in their use of unmapped assets. Their unmapped holdings generally consist of cash and/or other cash-like assets together with long futures contract exposure. Funds with unmapped holdings of this form generally appear in either allocation group IV or minimization group III.

Table 7 confirms this by presenting the percentage of index funds within each unmapped holdings group. Index funds were distinguished from actively managed funds using fund names. Any fund with the word 'index' or S&P in its name was designated as an index fund. A visual inspection of remaining funds was done to identify any additional index funds. The table shows that nearly all index funds are distributed between allocation group IV and minimization group VI. If the funds in group VI were merged into their closest bordering fund group, then nearly 97% of all index funds appear within group IV (Allocation) while only 26.2% of actively managed funds do the same (not shown).

Active Management vs. Indexing: Unmapped Risk Effects

	Total	**Diversification**		**Replic**	**Alloc**	**Minimization**	
	Funds	**Group I**	**Group II**	**Group III**	**Group IV**	**Group V**	**Group VI**
Index Funds	119	0.84%	0.00%	2.52%	38.66%	0.00%	57.98%
Active Mgmt	2058	24.39%	2.53%	17.30%	15.01%	1.94%	38.82%

Table 7: This table presents the risk effects of unmapped holdings in indexfunds and actively managed funds. Strong increasing risk is defined as$\sigma_p > \sigma_M$ in over 70% of the holdings dates in the sample, weakincreasing risk shows $\sigma_p > \sigma_M$ between 50% and 70% of the dates;weak decreasing shows $\sigma_p > \sigma_M$ between 30% and 50% of the dates,and strong decreasing risk is represented by funds with $\sigma_p > \sigma_M$in less than 30% of the dates in our sample.

This fact is relevant because of the risk and return implications described in section 4.1. Table 8 presents risk and return statistics for actively managed funds and index funds separately. The table shows that the unmapped holdings portion of index funds earns a strongly higher average return than the unmapped holdings portion of actively managed funds. The unmapped holdings of index funds add an average of between 7 and 17 basis points to index funds per year. In contrast, actively managed funds lose an average of between 29 and 150 basis points per year due to their investment in unmapped holdings.

A comparison between allocation group IV index funds and actively managed funds suggest index funds are superior in both risk and return. Unmapped holdings in allocation group IV index funds increased returns by 7 basis points while actively managed funds in the same group decreased returns by an average of 29 basis points. Risk in index funds was increased by 75 basis points while risk in actively managed funds in the same group increased by 167 basis points. Unmapped holdings significantly reduced the Sharpe ration in both index funds and actively managed funds in this group, but index funds were affected to a much smaller degree.

A comparison between group IV index funds and actively managed funds in other groups is more problematic. For example, unmapped holdings induced returns of actively managed funds in the diversification groups I and II produced considerably lower returns than index funds, but they also substantially reduced portfolio volatility. As a consequence, the effect of unmapped holdings upon the Sharpe ratio of group I and II actively managed funds is superior to the comparable Sharpe ratio effect in index funds.

These results produce mixed inferences about the value of active management. Suppose one compares only actively managed funds and index funds that have common unmapped holdings characteristics (allocation group IV). Then the unmapped holdings management of index funds strongly outperforms comparable returns by actively managed funds. This outcome implies that active management destroys value.

Alternatively, perhaps allocation group IV is a poor representation of active managers. One may instead compare unmapped holdings effects in indexes (allocation group IV) against the comparable effects in actively managed funds where the greatest number of active funds exist (diversification group I). Index funds obtain better performance due to unmapped holdings than actively managed funds, but also induce greater portfolio volatility. The effect of unmapped holdings upon the sharpe ratio of actively managed funds appears slightly better than that of indexes. This outcome suggests that the largest group of active managers do not do worse than indexes in terms of risk adjusted returns. However, this outcome does not imply that active managers are informed investors, because the source of their risk adjusted gains may simply result from diversification and not information.

Only aggressive diversification group II funds strongly improve their Sharpe ratio through the use of unmapped holdings and consequently strongly outperform index funds.

The strongest implication of table 8 is that a direct comparison of returns between indexes and actively managed funds is invalid because it fails to account for risk effects due to different ways funds managed unmapped holdings. Furthermore, this problem is not corrected through standard equity factor model regressions because by definition

unmapped holdings are not U.S. equities and are likely not highly correlated with equity factors.

Active Funds: Contributions to Portfolio Risk and Returns

	Group I Div.	Group II Agg. Div.	Group III Repl.	Group IV Alloc.	Group V Min.	Group VI Min.
N	481	54	354	163	46	727
$r_p - r_M$	-1.217%	-1.497%	-0.427%	-0.289%	-0.516%	-0.777%
	(-151.46***)	(-33.17***)	(-40.76***)	(-13.06***)	(-11.10***)	(-117.92***)
$\sigma_p - \sigma_M$	-1.009%	-2.755%	0.013%	1.669%	1.004%	0.204%
	(-243.02***)	(-211.48***)	(2.59**)	(228.33***)	(69.90***)	(59.17***)
$\Delta\frac{r}{\sigma}$	-0.074%	0.978%	-0.183%	-0.607%	-0.262%	-0.397%
	(-0.81)	(2.54**)	(-2.53**)	(-4.58***)	(-0.74)	(-8.21***)

Index Funds: Contributions to Portfolio Risk and Returns

	Group I Div.	Group III Repl.	Group IV Alloc.	Group VI Min.
N	1	3	46	69
$r_p - r_M$	-0.217%	0.174%	0.071%	0.087%
		(15.67***)	(5.84***)	(11.11***)
$\sigma_p - \sigma_M$	-0.364%	0.169%	0.750%	0.386%
		(3.15***)	(54.68***)	(34.50***)
$\Delta\frac{r}{\sigma}$	0.080%	-0.006%	-0.215%	-0.103%
		(-0.06)	(-2.48**)	(-1.56)

Table 8: This table presents the mean return gap ($r_p r_M$), the change in portfolio volatility ($\sigma_p \sigma_M$), and the change in the Sharpe ratio between the portfolio and mapped holdings as a hypothetical portfolio ($\Delta\frac{r}{\sigma}$). Panel A summarizes data for non-index funds, and panel B summarizes data for only index funds. The first column represents all portfolios combined and subsequent columns correspond with individual groups of funds based upon their unmapped holdings characteristics. Returns are annualized in percent, as are standard deviations.

4.3 The Return Gap

The return gap has recently been found to be a measure which can predict mutual fund returns. Kacperczyk et. al. demonstrate that the return gap is a strongly persistent positive predictor of mutual fund performance. The greatest strength in their findings are at the tails of the gap distributions. That is, top decile return gap funds predict

future top decile return performance and bottom decile return gap funds predict future bottom decile return performance. Kacperczyk et. al. attribute this persistence to the unobserved actions of fund managers, without considering the potential influence of unmapped holdings upon their results.

Table 9 produces evidence that shows that the return gap does have a statistically significant effect upon equity risk factors. The table presents risk coefficient estimates of the return gap regressed upon the Carhart 4-factor model. It shows statistically significant risk factor exposure on each of the factors, which suggests that funds with large positions in unmapped holdings also substantially alter the equity-related risk of their portfolio. Unmapped holdings may also introduce additional risk factor exposure that is not measureable by U.S. equity markets. Unmodeled risk factor effects would tend to push funds with such exposures towards the tails of a return gap sort, or precisely to the tails where the Kacperczyk results are strongest. Return gap predictability of performance may be considerably weakened if performance differences merely reflect differences in risk.

One may also observe evidence of unmapped holdings in the volatility of the return gap. Since return gap volatility is directly proportional to unmapped holdings, funds with large positions in unmapped holdings should also have the greatest return gap volatility immediately after holdings are reported. Table 10 shows the standard deviation of the return gap subsequent to holdings report dates. Table 10 shows that the initial 10 day standard deviation for funds with the lowest or highest return gap is high relative to the initial volatility of funds with less extreme return gap ranks. Over the subsequent 20 days there is an additional increase in volatility which would correspond

Gap Risk Coefficient Statistics by Group

	I	II	III	IV	V	VI
Group:	Div.	Agg. Div.	Replic.	Alloc.	Min.	Min.
β_{RMRF}	-0.074	-0.213	0.004	0.102	0.054	0.017
	(-29.54***)	(-12.46***)	(2.71***)	(12.05***)	(13.00***)	(15.72***)
β_{HML}	0.006	0.058	0.051	0.051	0.089	0.032
	(2.00**)	(2.81***)	(14.76***)	(8.57***)	(4.00***)	(19.38***)
β_{SMB}	-0.012	-0.028	0.031	0.060	0.055	0.024
	(-8.79***)	(-2.68***)	(18.60***)	(14.10***)	(7.89***)	(23.89***)
β_{UMD}	0.014	0.011	0.058	0.075	0.116	0.048
	(7.04***)	(1.01***)	(22.14***)	(21.98***)	(7.47***)	(34.38***)

Table 9: This table presents the average statistics for funds with unmappedholdings. Subgroups were assigned based upon how unmapped holdings alter portfolio risk, how correlated they are with mapped holdings, and how muchweight funds invest in them. Subgroup names were assigned after a careful review of the types of unmapped assets that are contained within the funds ineach subgroup. The criteria to identify each subgroup is provided in the text.

to volatility due to unobserved actions. There is evidence supporting the existence of both unmapped holdings and unobserved actions within extreme return gap portfolios, but because of the presence of unmapped holdings, it is not clear if predictability of the return gap can be attributed to informed trading by fund managers or simply to risk factor exposure that isn't measured because its existence is due to unmapped holdings.

The test for predictable performance is performed by calculating each measure using daily fund returns over a 3 month period from time $t-1$ to t. At time t, the measures are ranked and a long-short portfolio of top and bottom quintile funds is constructed based upon each ranked measure. Although one cannot "short sell" a mutual fund, the performance of this hypothetical portfolio directly demonstrates the success or failure of each of the measures ability to predict performance. The actual performance of hypothetical long-short portfolios is therefore measured from time t to time $t+1$. At time $t+1$, the three measures are recalculated and the long-short

Standard Deviation by Gap Rank and Time Elapsed

Gap Rank	10 days	30 days	60 days	90 days	180 days	300 days
1	2.97%	4.51%	3.42%	3.02%	3.07%	3.07%
2	0.70%	0.78%	0.83%	0.96%	1.18%	1.18%
3	0.67%	0.71%	0.83%	0.89%	0.95%	0.95%
4	0.64%	0.61%	0.63%	0.71%	0.77%	0.77%
5	0.51%	0.55%	0.64%	0.67%	0.69%	0.69%
6	0.53%	0.53%	0.58%	0.56%	0.62%	0.62%
7	0.66%	0.65%	0.65%	0.71%	0.79%	0.79%
8	0.49%	0.63%	0.74%	0.79%	1.03%	1.03%
9	1.08%	0.96%	1.08%	1.12%	1.44%	1.44%
10	1.66%	1.81%	1.84%	1.91%	2.64%	2.64%

Table 10: **Standard Deviation by Gap Rank and Time Elapsed** This table shows the daily standard deviation of the return gap, ranked by average return gap decile.

portfolio is reconstituted exactly as was done at time t.

4.4 Testing Predictability of the Return Gap with Endogenous Factors

A test for predictability of mutual fund performance is this paper's final test. Having established the relevance and economic significance of unmapped holdings, it is useful to test for predicatbility mutual fund performance when unmapped holdings are controlled. In this paper, predictability due to the return gap, the 4 factor regression model, and the 4 plus endogenous factor model is compared. These factors are ranked and a long-short portfolio of top and bottom quintile mutual funds is constructed. The performance results of these portfolios are detailed in tables 11, 12, and 13.

Table 11 presents the absolute performance results of this test. In all funds and in nearly every subgroup of funds, one observes the greatest performance is obtained when the endogenous factor is incorporated in the ranking of past performance. Performance

gains are obtained through a blend of long and short exposure, as can be seen by the lack of statistical significance in both the long and the short portfolio segments.

Tables 12 and 13 produce the same performance results as table 11, except the long-short portfolio return series is regressed on factor models to eliminate known risk exposures. Table 12 presents results when the long-short portfolio returns are regressed on the Carhart 4-factor model, and table 13 presents results when the long-short portfolio returns are regressed on the 4-factor model plus the endogenous unmapped holdings factor.

According to either the 4 factor model risk adjustment (table 12) or the 4 factor model plus endogenous factor risk adjustment (table 13), it is hard to distinguish superior performance between a ranking on either th past 4-factor alpha and past 4-factor plus endogenous alpha. In almost all cases, these two rankings outperform a ranking on the past average return gap.

These results provide strong evidence that the average return gap is an inferior predictor of mutual fund performance. Inclusion of an endogenous factor to control for unmapped holdings in the estimation of a portfolio's performance generally improves ex-post absolute returns and in many cases outperforms comparable predictions when the factor is omitted. Inclusion of the endogenous factor in the prediction also generally strengthens the t-statistics of ex-post excess-performance.

Ex-Post Portfolio Returns (Annualized)
4/1/2004 through 8/31/2006

Rank	Long-Short		Long Only		Short Only	
Criteria	Return	T-Stat	Return	T-Stat	Return	T-Stat
all funds						
Avg. Gap	0.35%	0.29	6.43%	0.77	6.07%	0.78
4-Factor	1.59%	1.22	6.84%	0.84	5.25%	0.65
4+e Factor	1.63%	1.55	6.95%	0.86	5.31%	0.66
Group I: Diversification						
Avg. Gap	1.30%	0.96	8.08%	1.14	6.78%	1.07
4-Factor	2.51%	1.93*	8.15%	1.17	5.64%	0.82
4+e Factor	2.33%	1.88*	7.82%	1.13	5.49%	0.8
Group II: Aggressive Diversification						
Avg. Gap	3.79%	0.91	7.30%	1.31	3.51%	0.52
4-Factor	1.04%	0.27	6.08%	1.12	5.04%	0.73
4+e Factor	4.10%	0.92	5.88%	1.13	1.78%	0.23
Group III: Replication						
Avg. Gap	-0.24%	-0.24	5.67%	0.67	5.91%	0.69
4-Factor	-0.48%	-0.32	5.65%	0.67	6.13%	0.75
4+e Factor	-0.24%	-0.18	5.84%	0.69	6.07%	0.74
Group IV: Allocation						
Avg. Gap	1.09%	0.24	9.73%	0.85	8.64%	0.91
4-Factor	4.12%	1.2	11.67%	1.18	7.55%	0.71
4+e Factor	4.26%	1.35	11.47%	1.16	7.21%	0.68
Group V: Minimization						
Avg. Gap	-3.71%	-0.84	3.03%	0.35	6.75%	0.71
4-Factor	0.33%	0.07	4.87%	0.54	4.55%	0.5
4+e Factor	1.26%	0.28	6.05%	0.67	4.79%	0.53
Group VI: Minimization						
Avg. Gap	-0.55%	-0.55	5.09%	0.61	5.64%	0.68
4-Factor	2.66%	1.68*	6.49%	0.77	3.84%	0.46
4+e Factor	2.78%	1.77*	6.34%	0.75	3.56%	0.42

Table 11: This table presents the ex-post performance of a portfolio that is constructed from quintile ranks in 3 month estimation windows using the average return gap, the 4-factor alpha, or the 4-factor + endogenous alpha. Investments are made based upon this ranking with a long position in top quintile funds and a short postion in bottom quintile funds. Returns are measured and rebalanced daily, and new portfolio weights are selected each calendar quarter.

Risk Adjusted Performance Results (annualized)
4/1/2004 through 8/31/2006
4-Factor Model Risk Control

Rank	Long-Short			Long Only			Short Only		
Criteria	Intercept	Tstat	R^2	Intercept	Tstat	R^2	Intercept	Tstat	R^2
All Funds									
Avg. Gap	0.27%	0.24	0.23	-0.11%	-0.11	0.99	-0.38%	-0.41	0.99
4-Factor	1.79%	1.38	0.06	0.48%	0.43	0.98	-1.30%	-1.45	0.99
4+e Factor	1.70%	1.64*	0.06	0.55%	0.55	0.99	-1.16%	-1.31	0.99
Group I: Diversification									
Avg. Gap	0.52%	0.48	0.38	0.76%	0.83	0.98	0.24%	0.30	0.98
4-Factor	2.06%	1.68*	0.14	0.92%	0.98	0.98	-1.14%	-1.21	0.98
4+e Factor	2.08%	1.73*	0.09	0.73%	0.80	0.98	-1.34%	-1.44	0.98
Group II: Aggressive Diversification									
Avg. Gap	5.50%	1.33	0.05	1.19%	0.65	0.89	-4.31%	-1.11	0.68
4-Factor	0.66%	0.19	0.20	-0.04%	-0.02	0.82	-0.70%	-0.29	0.88
4+e Factor	6.14%	1.60	0.28	0.34%	0.15	0.82	-5.80%	-1.79	0.83
Group III: Replication									
Avg. Gap	0.09%	0.09	0.06	-0.59%	-0.53	0.98	-0.67%	-0.56	0.98
4-Factor	-0.70%	-0.47	0.05	-0.86%	-0.66	0.98	-0.16%	-0.14	0.98
4+e Factor	-0.35%	-0.27	0.05	-0.51%	-0.42	0.98	-0.16%	-0.14	0.98
Group IV: Allocation									
Avg. Gap	1.43%	0.35	0.21	1.79%	0.69	0.95	0.36%	0.17	0.95
4-Factor	5.86%	1.75*	0.08	5.03%	2.31	0.95	-0.83%	-0.37	0.96
4+e Factor	5.77%	1.85*	0.06	4.74%	2.28	0.96	-1.03%	-0.47	0.96
Group V: Minimization									
Avg. Gap	-0.53%	-0.13	0.13	-0.46%	-0.13	0.84	0.07%	0.02	0.89
4-Factor	1.01%	0.22	0.03	-0.23%	-0.07	0.87	-1.24%	-0.39	0.88
4+e Factor	1.83%	0.40	0.03	0.68%	0.20	0.87	-1.14%	-0.36	0.88
Group VI: Minimization									
Avg. Gap	-0.02%	-0.02	0.04	-0.73%	-0.69	0.98	-0.71%	-0.69	0.99
4-Factor	2.48%	1.57	0.03	0.39%	0.30	0.98	-2.09%	-1.68	0.98
4+e Factor	2.28%	1.46	0.05	0.10%	0.08	0.98	-2.18%	-1.71	0.98

Table 12: This table presents risk controlled performance results for a portfolio constructed from ranked return gap, ranked 4-factor, and ranked 4+e factor models. Rankings are based upon daily returns over three monthmeasurement windows. A long position is invested in top quintile mutual fundsand a short position is invested in bottom quintile mutual funds. Risk factors are controlled using a 4-factor model regression.

Risk Adjusted Performance Results (annualized)
4/1/2004 through 8/31/2006
4 Plus Endogenous Factor Model Risk Control

Rank	Long-Short			Long Only			Short Only		
Criteria	Intercept	Tstat	R^2	Intercept	Tstat	R^2	Intercept	Tstat	R^2
All Funds (Pooled)									
Avg. Gap	0.60%	0.55	0.26	0.31%	0.34	0.99	-0.29%	-0.33	0.99
4-Factor	1.38%	1.08	0.09	0.57%	0.54	0.98	-0.81%	-0.98	0.99
4+e Factor	1.33%	1.30	0.11	0.59%	0.64	0.99	-0.74%	-0.92	0.99
Group I: Diversification									
Avg. Gap	0.51%	0.47	0.38	0.63%	0.70	0.98	0.12%	0.15	0.98
4-Factor	2.01%	1.64*	0.14	0.77%	0.83	0.98	-1.24%	-1.32	0.98
4+e Factor	2.05%	1.71*	0.09	0.58%	0.64	0.98	-1.47%	-1.58	0.98
Group II: Aggressive Diversification									
Avg. Gap	5.61%	1.35	0.05	1.07%	0.59	0.90	-4.54%	-1.18	0.69
4-Factor	0.60%	0.17	0.20	-0.20%	-0.09	0.83	-0.80%	-0.33	0.88
4+e Factor	6.21%	1.62*	0.28	0.21%	0.09	0.82	-6.01%	-1.87	0.83
Group III: Replication									
Avg. Gap	0.11%	0.11	0.07	-0.64%	-0.59	0.98	-0.74%	-0.64	0.98
4-Factor	-0.70%	-0.47	0.05	-0.90%	-0.69	0.98	-0.19%	-0.17	0.98
4+e Factor	-0.37%	-0.28	0.05	-0.56%	-0.47	0.98	-0.20%	-0.17	0.98
Group IV: Allocation									
Avg. Gap	1.42%	0.35	0.21	1.90%	0.74	0.95	0.48%	0.22	0.95
4-Factor	5.54%	1.66*	0.09	4.97%	2.28**	0.95	-0.57%	-0.26	0.96
4+e Factor	5.44%	1.76*	0.07	4.67%	2.25**	0.96	-0.77%	-0.35	0.96
Group V: Minimization									
Avg. Gap	0.28%	0.07	0.14	0.67%	0.19	0.84	0.39%	0.12	0.90
4-Factor	0.30%	0.07	0.04	0.14%	0.04	0.87	-0.16%	-0.05	0.89
4+e Factor	1.32%	0.29	0.03	1.25%	0.37	0.87	-0.07%	-0.02	0.89
Group VI: Minimization									
Avg. Gap	-0.02%	-0.02	0.04	-0.81%	-0.80	0.99	-0.79%	-0.81	0.99
4-Factor	2.44%	1.55	0.04	0.26%	0.21	0.98	-2.19%	-1.81*	0.98
4+e Factor	2.28%	1.46	0.05	-0.01%	-0.01	0.98	-2.29%	-1.87*	0.98

Table 13: This table presents risk controlled performance results for a portfolio constructed from ranked return gap, ranked 4-factor, and ranked 4+e factor models. Rankings are based upon daily returns over three monthmeasurement windows. A long position is invested in top quintile mutual fundsand a short position is invested in bottom quintile mutual funds. Risk factors are controlled using a 4-factor model regression plus the endogenous factor.

5 Conclusion

This paper has explored a newly available dataset containing rich information about unmapped holdings in mutual fund portfolios. Due to the voluntary reporting of this data, these findings represent a conservative measure of potential additional assets that may be held.

Using a simple two stock portfolio, this paper derives from portfolio returns the implied properties of unmapped holdings including return, volatility, and correlation with mapped holdings. These measures are used to identify and decompose how different mutual funds invest unmapped holdings. Mutual funds are classified into groups based upon their unmapped holdings characteristics, and their holdings were reviewed to validate that the groups are classified appropriately. Three primary groups of funds can be distinguished by their use of unmapped holdings: Diversifiers, Minimizers, and Allocators. These groups can be distinguished by three parameters: weight of investment in unmapped holdings (w_U), correlation of unmapped holdings with unmapped holdings (ρ_{UM}), and the influence of unmapped holdings upon portfolio risk ($\Delta\sigma_p$)

Management of unmapped holdings varies across commonly known investment styles. Their use appears to be one way funds within each of the styles can appeal to different investor criteria. A classification of funds into unmapped holdings groups also exposes a key difference between index funds and actively managed funds. Index funds tend to be allocators (group IV funds) while actively managed funds tend to be more spread across groups with the majority being either diversifiers (groups I and II) or minimizers (group VI). Return comparisons between indexes and actively man-

aged funds must consider Sharpe ratios effects of unmapped holdings instead of simple return comparisions or regressions upon equity risk factors.

Mutual fund returns are better predicted when unmapped holdings are controlled. The return gap is inferior in its prediction of returns. This inferiority is largely due to its implicit correspondence to investment in unmapped holdings. A test of mutual fund predictability using the return gap, 4-factor model regressions, and 4-factor plus endogenous benchmark regressions show that a control for unmapped holdings generally produces the strongest return predictions, stronger t-statistics, and stronger absolute returns. A caveat to the strength of the predictability results is that there is only a limited time series of data available for analysis.

6 Appendix A: Unmapped Holdings and Unobserved Actions: Intertemporal Trades

Proposition 3.1 relies upon the assumption of a static investment portfolio that is held between holdings report dates. In this section, those results are generalized and confirm similar findings when allowing for unobserved trading by fund managers.

Assume a single investment period that begins at the effective date that holdings are observed, and ends upon the subsequent holdings date. Assume that within this period, there are N securities available and D possible dates in which the fund can trade. Furthermore, there are $M \leq N$ total assets that are mapped and observable throughout the trading period, and $N-M$ total assets that are reported at the holdings date, but it is costly to monitor them (unobservable). For notational clarity, denote

a variable representing a vector or matrix in **boldprint**. Scalars and other variables will be denoted in regular text. One may define the vectors $\mathbf{w_d}$ and $\mathbf{r_d}$ to represent portfolio weights after trading date d and asset returns from trading date d to $d+1$:

$$\mathbf{w_d} = \begin{pmatrix} w_{1,d} \\ . \\ . \\ w_{M,d} \\ w_{M+1,d} \\ . \\ . \\ w_{N,d} \end{pmatrix}, \mathbf{w_{Md}} = \begin{pmatrix} w_{1,d} \\ . \\ . \\ w_{M,d} \\ 0 \\ . \\ . \\ 0 \end{pmatrix}, \mathbf{w_{Ud}} = \begin{pmatrix} 0 \\ . \\ . \\ 0 \\ w_{M+1,d} \\ . \\ . \\ w_{N,d} \end{pmatrix}, \mathbf{r_d} = \begin{pmatrix} r_{1,d} \\ . \\ . \\ r_{M,d} \\ r_{M+1,d} \\ . \\ . \\ r_{N,d} \end{pmatrix}$$

Here $\mathbf{w_d} = \mathbf{w_{Md}} + \mathbf{w_{Ud}}$ representing portfolio weights in mapped and unmapped holdings, respectively.

To distinguish unmapped holdings and unobserved actions separately through the trading period, recognize that $\mathbf{w_0}$, $\mathbf{w_{M0}}$, and $\mathbf{w_{U0}}$ represent beginning of period (observable) portfolio weights. Define $\mathbf{\Delta_d} = \mathbf{w_d} - \mathbf{w_0}$ which represents unobserved intra-period trading.

To derive these results, use sampling theory on portfolio level returns. Define portfolio returns over the holdings period:

$$\mathbf{R_p} = \begin{pmatrix} \mathbf{w_1'r_1} \\ . \\ . \\ \mathbf{w_D'r_D} \end{pmatrix}$$

Similarly, define the components of portfolio returns as follows:

$$\mathbf{R_\Delta} = \begin{pmatrix} \mathbf{\Delta_1'r_1} \\ . \\ . \\ \mathbf{\Delta_D'r_D} \end{pmatrix}, \mathbf{R_{w_0}} = \begin{pmatrix} \mathbf{w_0'r_1} \\ . \\ . \\ \mathbf{w_0'r_D} \end{pmatrix},$$

$$\mathbf{R_{w_{M0}}} = \begin{pmatrix} (\mathbf{w_{M0}1_N})^{-1}\mathbf{w'_{M0}r_1} \\ \cdot \\ \cdot \\ (\mathbf{w_{M0}1_N})^{-1}\mathbf{w'_{M0}r_D} \end{pmatrix}, \mathbf{R_{w_{U0}}} = \begin{pmatrix} (\mathbf{w_{U0}1_N})^{-1}\mathbf{w'_{U0}r_1} \\ \cdot \\ \cdot \\ (\mathbf{w_{U0}1_N})^{-1}\mathbf{w'_{U0}r_D} \end{pmatrix}$$

Note that $(\mathbf{w_{M0}1_N})^{-1}$ and $(\mathbf{w_{U0}1_N})^{-1}$ are normalizations. From this definition, one can identify the sample properties of portfolio returns.

Lemma 6.1 *The sample properties of portfolio returns are as follows:*

1. *Portfolio Return:* $\mathbf{R_p} = \mathbf{R_\Delta} + (\mathbf{w'_{M0}}1_D)\mathbf{R_{w_{M0}}} + (\mathbf{w'_{U0}}1_D)\mathbf{R_{w_{U0}}}$
2. *Sample Mean:* $\bar{r}_p = \frac{1}{D}\mathbf{R'_p 1_D} = \frac{1}{D}\left(\mathbf{R'_\Delta 1_D} + (\mathbf{w'_{M0}}1_D)\mathbf{R'_{w_{M0}}1_D} + (\mathbf{w'_{U0}}1_D)\mathbf{R'_{w_{U0}}1_D}\right)$
3. *Sample Variance:*

$$s_p^2 = \frac{1}{D}E(R_pR'_p - E(R_p)E(R'_p)) = (\mathbf{w'_{U0}1_N})^2 s_U^2 + (\mathbf{w'_{M0}1_N})^2 s_M^2 + 2\mathbf{w'_{U0}1_N w'_{M0}1_N}s_{UM} + s_\Delta^2 + 2s_{\Delta,w_0}^2$$

$$s_U^2 = \frac{1}{D}E(R_{w_U0}R'_{w_U0} - E(R_{w_U0})E(R'_{w_U0}))$$

$$s_M^2 = \frac{1}{D}E(R_{w_M0}R'_{w_M0} - E(R_{w_M0})E(R'_{w_M0}))$$

$$s_{UM} = \frac{1}{D}E(R_{w_U0}R'_{w_M0} - E(R_{w_U0})E(R'_{w_M0}))$$

$$s_\Delta^2 = \frac{1}{D}E(R_\Delta R'_\Delta - E(R_\Delta)E(R'_\Delta))$$

$$s_{\Delta,w_0} = \frac{1}{D}E(R_\Delta R'_{w_0} - E(R_\Delta)E(R'_{w_0}))$$

Proof *Results follow from sampling theory of functions of random variables.* ∎

Lemma 6.1 confirms that one can segment portfolio returns into three components: mapped holdings, unmapped holdings, and unobserved actions. An alternative measure of portfolio returns is the return gap which is identified below in lemma 6.2.

Lemma 6.2 *The sample properties of the return gap are as follows:*

1. *Portfolio Return Gap:* $\mathbf{R_g} = \mathbf{R_p} - \mathbf{R_{w_{M0}}} = \mathbf{R_\Delta} + (\mathbf{w'_{U0}}1_D)(\mathbf{R_{w_{U0}}} - \mathbf{R_{w_{M0}}})$
2. *Sample Mean:* $\bar{r}_g = \frac{1}{D}\mathbf{R'_g 1_D} = \frac{1}{D}\mathbf{R'_\Delta 1_D} + (\mathbf{w'_{U0}}1_D)(\mathbf{R'_{w_{U0}}1_D} - \mathbf{R'_{w_{M0}}1_D})$
3. *Sample Variance:*

$$s_g^2 = \frac{1}{D}E(R_gR'_g - E(R_g)E(R'_g)) = (\mathbf{w'_{U0}1_N})^2(s_U^2 + s_M^2 - 2s_{UM}) + s_\Delta^2 - 2s_{\Delta,w_0}$$

Proof *Results follow from sampling theory of functions of random variables.* ∎

The return gap represents a long-short portfolio which consists of a long position in unmapped holdings, a short position in mapped holdings, and a long-short position in unobserved actions. As previously derived in section 3.1, one can take advantage of differences between sample statistics of total portfolio returns and the return gap to derive the variance of unmapped holdings and the covariance between unmapped holdings and mapped holdings. To simplify notation, let $w_0 = \mathbf{w_0'1_N}$, $w_M = \mathbf{w'_{M0}1_N}$, and $w_U = \mathbf{w'_{U0}1_N}$.

Proposition 6.3 *Unmapped holdings have the following implied sample properties:*

1. *Implied Unmapped Holdings Return:*

$$\mathbf{\hat{R}_{wU0}} = w_U^{-1}(\mathbf{R_p} - (w_M \mathbf{R_{wM0}})) = \mathbf{R_{wU0}} + \xi$$

$$\xi = \mathbf{\hat{R}_{wU0}} - \mathbf{R_{wU0}} = w_U^{-1}\mathbf{R_\Delta}$$

2. *Implied Unmapped Holdings Variance:*

$$\hat{s}_U^2 = \frac{w_U s_p^2 + w_M s_g^2 - w_U w_M s_M^2}{w_U^2} = s_U^2 + \Upsilon$$

$$\Upsilon = \hat{s}_U^2 - s_U^2 = \frac{s_\Delta^2 - 2s_{\Delta,w_0} + 2w_M s_{\Delta,M}}{w_U^2}$$

3. *Implied Covariance between Mapped and Unmapped Holdings:*

$$\hat{s}_{UM} = \frac{s_p^2 - s_g^2 - s_M^2(w_M - w_U)}{2w_U} + \vartheta$$

$$\vartheta = \hat{s}_{UM} - s_{UM} = \frac{s_{\Delta,M}}{w_U}$$

Proof *The proof is the virtually identical to the proof in proposition 3.1, but with the added unobserved trading components.* ∎

The results of proposition 6.3 define explicitly the combined effects of unmapped holdings and unobserved actions on implied holdings, implied variance, and implied co-

variance of unmapped holdings. In section 3.2, these properties are used to distinguish mutual funds by their use of unmapped holdings.

7 Appendix B: Group Correlation Analysis

Correlations between group average returns indicate how much of each group's performance is common with the performance of other groups. Tables 14 and 15 present the between group correlations of average group returns after removing factor exposure for market, book-to-market, capitalization, and momentum exposure.

Table 14 presents the correlations between groups with common changes in risk. One finds low correlations between subsets of funds, suggesting that the groups have distinct common effects. The different columns in the table represent the across group correlation at each calendar quarter in the sample. Two pairs of groups share the highest average correlation across all dates. The cash management and portfolio extension groups shows an average correlation of 0.66. The index or levered group has a similar correlation of 0.65 with cash management funds.

Across-group correlations among remaining groups are considerably smaller. Risk reducing groups of funds have average correlations of 0.46 between allocation or balanced type funds and market neutral or option funds, 0.28 between allocation or balanced funds and cash management funds, and 0.23 between market neutral or option funds and cash management funds. The across-group correlations for funds where unmapped holdings increased risk showed average correlations of 0.20 between the index or levered funds and market neutral or derivative strategy funds, and 0.49 between the

market neutral or derivative stragegy funds and cash management funds.

Table 15 presents across group correlations that compare groups of funds with similar portfolio weights and correlations in their unmapped holdings. There is a very strong correlation between cash management groups between those with no change in portfolio risk and a slight increase in portfolio risk. The average correlation between these groups was 0.87. For all remaining across-group correlations, one observes considerably smaller correlations. Correlations between cash management funds that increased risk and cash management funds that decreased risk averaged 0.35. The correlation between cash management funds that decreased in risk and made no change to risk averaged 0.42. Low correlations between suggest that distinct risk factors influence the different groups of funds.

The weakest correlation between groups in table 15 is between increasing risk and decreasing risk funds that hold larger positions in unmapped assets. For example, market neutral and derivative strategy funds that increased portfolio risk averaged a correlation of about 0.1 to 0.2 throughout the measurement period against market neutral and derivative strategy funds that decreased portfolio risk. Likewise, allocation or balanced funds averaged a negative correlation of between 0.1 or 0.2 compared to index or levered funds. Both tables 14 and 15 reflect evidence that the fund subgroups represent distinct risk factors that are beyond what is captured in classical regression models. Likewise, the correlations in these tables give a strong indication of the degree that the fund subgroups share similarities or differences. Finally, these tables confirm that the greatest differences in correlation between subgroups correspond with the groups of funds that have the greatest allocation to unmapped holdings.

Pairwise correlations between funds within each group allows an assessment of how well the endogenous benchmarks explain common risk in those funds. Fund returns are regressed upon both the Carhart 4-factor model and an extended Carhart 4-factor model that includes an endogenous benchmark y_g that corresponds to each group. Table 16 shows the percentage of mutual funds with significant pairwise residual correlations under both regression models. Table 16 shows that significant residual correlations in all groups decline by roughly one-third. In every group the percentage of funds with significant residual correlations in the basic 4-factor model is between 25% and 80%, most commonly averaging close to 35%. When the endogenous benchmark y_g is included in the regression, funds with significant residual correlations decline to around 20%. Funds in the index and levered group show the greatest residual correlations after including the endogenous factor with around 40% of these funds persisting to show significant residual correlation, but the correlation is still an improvement from the 4-factor model alone where around 50% of funds have a significant residual correlation.

A final test of group representation by the endogenous benchmark can be constructed from a simple regression test where each fund in the group is regressed on the Carhart 4-factor model plus the endogenous factor. The percentage of funds with a significant factor loading on the endogenous factor represents a measure of how well the factor explains funds in the group. Likewise, the change in R^2 between a 4-factor model regression and the extended 4-factor model with the endogenous factor included provides a measure of economic significance. The distribution of significant regression estimates using the 4-factor model, the 4-factor plus endogenous model, and the endogenous benchmark (non-orthogonalized) alone are shown in table 17. Among all

groups, one observes a consistent distribution of alpha estimates, regardless of if the 4-factor model was used or the 4-factor plus endogenous benchmark model was used.

Column 3 of table 17 shows the percentage of regressions where the endogenous benchmark was significant. The endogenous benchmark is significant in about 50% of the fund regressions, and positive in roughly 90% of all regressions. The greatest significance of the endogenous factor appears within the index or leveraged fund group. In this group 75.8% of funds had statistically significant exposure to the endogenous benchmark. This is also the group where the 4-factor model has the best fit as demonstrated by the 4-factor R^2 of 96.8%.

A regression on endogenous benchmark alone (r_g), represented in column 4 of table 17 shows that the endogenous benchmark (not orthogonolized) obtains an R^2 that is not as strong as the R^2 of the 4 factor model, but that it does perform nearly as well.

There is a small increase the R^2 every time the endogenous factor is included in the regression. The greatest increase in R^2 appears among funds that reduced portfolio volatility in the market neutral or option funds group. Of all groups of funds in the sample, this is also the group with the lowest average R^2 when using the 4-factor model regression. Thus funds with market neutral or options strategies that reduce portfolio risk are among the most poorly fit funds by the 4-factor regression model. The average R^2 for these funds averaged was 80.3% using the 4-factor regression model. The R^2 using the 4 plus endogenous model was 82.7%. Most of the remaining groups had an increase in R^2 of about 1.1%. The group with the least improvement in R^2 was cash management funds where their unmapped assets decreased portfolio risk. Among these funds, the 4-factor R^2 was 92.8% and increased to 93.4% when the endogenous

factor was included. This represents and increase of only 0.6%. The increased R^2 in these funds show a small but economically significant effect of unmapped holdings upon portfolio returns.

Correlation Pairs	Correlations (across groups) Period											
	07/03 09/03	10/03 12/03	01/04 03/04	04/04 06/04	07/04 09/04	10/04 12/04	01/05 03/05	04/05 06/05	07/05 09/05	10/05 12/05	01/06 03/06	04/06 - 06/06
Risk Reducing												
Alloc/Bal	0.58	0.28	0.62	0.37	0.35	0.5	0.49	0.35	0.44	0.5	0.48	0.59
& Mkt Ntrl/Opts	0.000	0.014	0.000	0.002	0.002	0.000	0.000	0.003	0.000	0.000	0.000	0.000
Alloc/Bal	0.56	0.54	0.3	0.35	0.31	0.58	0.04	0.36	-0.03	0.27	-0.05	0.15
& Cash Mgmt	0.000	0.000	0.008	0.003	0.007	0.000	0.384	0.002	0.591	0.015	0.656	0.128
Mkt Ntrl/Options	0.46	0.08	0.44	0.63	0.47	0.36	0.16	0.07	-0.16	0.1	-0.05	0.19
& Cash Mgmt	0.001	0.272	0.000	0.000	0.000	0.002	0.103	0.287	0.892	0.225	0.662	0.071
No Risk Change												
Port Replic	0.78	0.67	0.48	0.58	0.65	0.52	0.61	0.78	0.55	0.77	0.72	0.77
& Cash Mgmt	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Risk Increasing												
Index/Leverage	0.30	0.13	0.02	0.02	0.09	0.22	0.45	0.60	-0.14	0.29	0.09	0.34
& Mkt Ntrl/Lvg/Mcap	0.026	0.146	0.449	0.426	0.241	0.038	0.000	0.000	0.873	0.010	0.250	0.003
Index/Leverage	0.62	0.6	0.78	0.76	0.62	0.61	0.67	0.73	0.52	0.56	0.64	0.73
& Cash Mgmt	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Mkt Ntrl/Lvg/Mcap	0.71	0.31	0.35	0.17	0.45	0.52	0.71	0.68	0.31	0.64	0.48	0.6
& Cash Mgmt	0.000	0.007	0.003	0.097	0.000	0.000	0.000	0.000	0.007	0.000	0.000	0.000

Table 14: This table presents the correlation of average group returns after orthogonalization against the Carhart 4-factor regression model (y_g in the document). Correlation pairs are grouped by risk posture in the top half of the table, and then by similar allocations in the bottom half of the table. Throughout the table, the first line represents the correlation estimate, and the second line is the p-value.

Correlation Pairs	Correlations (across groups) Period											
	07/03 09/03	10/03 12/03	01/04 03/04	04/04 06/04	07/04 09/04	10/04 12/04	01/05 03/05	04/05 06/05	07/05 09/05	10/05 12/05	01/06 03/06	04/06 - 06/06
Cash Management												
Incr. Risk	0.89	0.9	0.81	0.86	0.92	0.89	0.89	0.93	0.84	0.89	0.81	0.86
& No Chg Risk	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Incr. Risk	0.27	0.45	0.29	0.48	0.45	0.26	0.25	0.25	0.42	0.23	0.48	0.38
& Decr Risk	0.043	0.000	0.012	0.000	0.000	0.019	0.028	0.024	0.000	0.037	0.000	0.001
No Chg Risk	0.43	0.5	0.52	0.62	0.47	0.41	0.36	0.3	0.32	0.21	0.47	0.41
& Decr Risk	0.002	0.000	0.000	0.000	0.000	0.000	0.002	0.009	0.004	0.051	0.000	0.000
Mkt Ntrl/Lvg/Options												
Incr. Risk	0.32	0.27	0.31	0.41	0.21	0.19	0.1	0.18	0.27	0.41	0.25	0.15
& Decr Risk	0.020	0.016	0.007	0.000	0.050	0.065	0.232	0.081	0.016	0.000	0.024	0.117
Alloc/Bal vs. Index/Leverage												
Incr. Risk	-0.06	-0.06	-0.46	-0.49	-0.18	-0.27	-0.14	-0.28	-0.31	-0.13	-0.35	-0.21
& Decr Risk	0.637	0.684	1.000	1.000	0.921	0.985	0.852	0.987	0.994	0.849	0.998	0.948

Table 15: This table presents the correlation of average group returns after orthogonalization against the Carhart 4-factor regression model. Correlation pairs are grouped by similar allocations. Throughout the table, the first line represents the correlation estimate, and the second line is the p-value.

Within Group Correlations
(Percent of Funds with Significant (T-Stat ¿ 1.96) Residual Correlations)

Correlation Pairs	Model	Date 10/03 -12/03	01/04 -03/04	04/04 -06/04	07/04 -09/04	10/04 -12/04	01/05 -03/05	04/05 -06/05	07/05 -09/05	10/05 -12/05	01/06 -03/06	04/06 -06/06	07/06 -09/06
Reduce:Alloc/Bal	4 Factor	0.25	0.3	0.35	0.37	0.27	0.28	0.23	0.27	0.28	0.25	0.24	0.31
	4 Factor + e	0.17	0.16	0.19	0.19	0.19	0.2	0.17	0.18	0.18	0.17	0.16	0.19
Reduc:MNtrl/Opt	4 Factor	0.34	0.26	0.34	0.32	0.25	0.31	0.32	0.24	0.21	0.22	0.22	0.21
	4 Factor + e	0.19	0.2	0.22	0.26	0.19	0.17	0.17	0.14	0.15	0.15	0.15	0.16
Reduc:Cash Mgt	4 Factor	0.26	0.26	0.29	0.35	0.3	0.29	0.29	0.27	0.3	0.24	0.3	0.35
	4 Factor + e	0.21	0.19	0.2	0.25	0.22	0.19	0.18	0.19	0.24	0.19	0.18	0.18
No Chg:Port Replic	4 Factor	0.38	0.33	0.42	0.37	0.31	0.28	0.25	0.33	0.27	0.35	0.26	0.3
	4 Factor + e	0.22	0.16	0.23	0.29	0.2	0.18	0.19	0.2	0.16	0.17	0.19	0.18
No Chg:Cash Mgt	4 Factor	0.44	0.36	0.32	0.4	0.4	0.32	0.32	0.35	0.27	0.31	0.27	0.34
	4 Factor + e	0.19	0.17	0.19	0.2	0.17	0.17	0.17	0.17	0.16	0.17	0.16	0.18
Incr:Index/Lvg	4 Factor	0.54	0.62	0.63	0.51	0.5	0.48	0.46	0.5	0.48	0.51	0.6	0.54
	4 Factor + e	0.49	0.38	0.41	0.37	0.41	0.42	0.38	0.43	0.4	0.41	0.5	0.35
Incr:Mntrl/Lvg/Mcap	4 Factor	0.78	0.39	0.62	0.35	0.33	0.36	0.29	0.37	0.3	0.28	0.23	0.29
	4 Factor + e	0.24	0.23	0.2	0.25	0.27	0.24	0.17	0.19	0.21	0.19	0.18	0.21
Incr:Cash Mgt	4 Factor	0.35	0.36	0.37	0.47	0.35	0.32	0.31	0.38	0.3	0.31	0.3	0.35
	4 Factor + e	0.21	0.23	0.25	0.25	0.23	0.2	0.2	0.19	0.19	0.19	0.19	0.2

Table 16: This table presents the percentage of funds with significant residual correlations after regressing with the 4-factor model and with the 4-factor regression plus an endogenous benchmark. Regressions are constructed using daily data over 3 month non-overlapping periods.

Factor Estimation Results

	4-F Model α (%)	4 F + Endg α (%)	Endg F	Endg α (%)		4-F Model α (%)	4 F + Endg α (%)	Endg F	Endg α (%)
Risk Decreased					**Risk Increased**				
Allocation/Balanced					**Index/Leverage**				
Pos Sig	3.5%	2.7%	49.7%	4.0%	Pos Sig	1.7%	2.9%	75.8%	6.5%
Pos Not Sig	50.8%	45.9%	39.6%	44.9%	Pos Not Sig	39.7%	33.5%	19.9%	46.9%
Neg Not Sig	43.6%	47.8%	10.4%	46.6%	Neg Not Sig	51.2%	57.8%	3.9%	46.0%
Neg Sig	2.0%	3.7%	0.3%	4.6%	Neg Sig	7.5%	5.9%	0.4%	0.7%
R^2	89.9%	91.0%		86.4%	R^2	96.8%	97.8%		92.5%
Market Neutral/Options					**Market Neutral/Leverage/Microcap**				
Pos Sig	2.4%	2.3%	53.2%	2.9%	Pos Sig	2.3%	1.8%	58.3%	4.8%
Pos Not Sig	45.4%	48.8%	34.4%	49.8%	Pos Not Sig	55.9%	51.6%	34.2%	44.9%
Neg Not Sig	48.5%	45.6%	12.3%	43.8%	Neg Not Sig	38.0%	41.9%	6.9%	47.8%
Neg Sig	3.7%	3.4%	0.2%	3.6%	Neg Sig	3.9%	4.8%	0.5%	2.5%
R^2	80.3%	82.7%		75.4%	R^2	90.4%	91.6%		86.4%
Cash Management					**Cash Management**				
Pos Sig	1.7%	1.9%	51.4%	4.8%	Pos Sig	1.1%	2.9%	59.1%	4.6%
Pos Not Sig	48.1%	48.0%	39.2%	44.9%	Pos Not Sig	45.7%	47.1%	26.3%	45.9%
Neg Not Sig	46.3%	45.9%	9.0%	46.8%	Neg Not Sig	48.7%	46.7%	11.7%	46.3%
Neg Sig	3.9%	4.3%	0.4%	3.5%	Neg Sig	4.5%	3.3%	3.0%	3.2%
R^2	92.8%	93.4%		88.3%	R^2	93.4%	94.5%		89.6%
Risk Unchanged									
Portfolio extension									
Pos Sig	2.3%	2.4%	54.8%	3.7%					
Pos Not Sig	49.9%	46.0%	35.0%	42.8%					
Neg Not Sig	45.8%	49.9%	9.8%	50.0%					
Neg Sig	2.1%	1.7%	0.3%	3.5%					
R^2	93.2%	94.3%		89.9%					
Cash Management									
Pos Sig	2.8%	2.8%	56.9%	4.3%					
Pos Not Sig	47.8%	46.3%	28.9%	46.0%					
Neg Not Sig	46.4%	47.7%	12.0%	45.9%					
Neg Sig	3.0%	3.2%	2.3%	3.8%					
R^2	91.9%	93.1%		88.4%					

Table 17: This table presents the distribution of alpha estimates and the average R^2 estimates for each unmapped holdings subgroup. The first column of data represents the distribution of fund α estimates using the Carhart 4-factor model regression, the second column represents α estimates using the 4-factor model with an additional endogenous factor, and the third column represents the coefficient estimates of the endogenous factor. The fourth column represents regression results using only the non-orthogonolized endogenous factor.

One observes in these data that unmapped holdings have a statistically significant effect upon portfolio risk, and that by grouping funds by their unmapped holdings characteristics, one can identify different groups of funds with very distinct portfolio effects of their unmapped holdings. By constructing an endogenous benchmark for each of these groups, one observes that the endogenous benchmark successfully captures variation specific to these groups that is not explained by standard 4-factor regression analysis. The presence of unmapped holdings in a U.S. domestic equity portfolio introduces unmodeled risk factors into portfolio returns, and endogenous benchmarks may be used to help control for these risks.