Holdings-based Fund Performance Measures: Estimation and Inference[1]

Wayne E. Ferson
University of Southern California and NBER

Junbo L. Wang
Louisiana State University

December 12, 2017

Abstract

This paper introduces a predictive panel regression framework for holdings-based fund performance measures. The approach facilitates the correction of lagged stochastic regressor bias and allows the application of other statistical tools for panel regressions. We find an "Average Alpha Effect" in three of the four classical measures we study. This Effect is a cross-sectional relation between a fund's average portfolio weights and the stocks' average alphas in the benchmark model. The Average Alpha Effect is large in simulated strategies and in data on active US equity funds. It dominates the cross-sectional variation in the classical measures and is stronger in more recent data. The Effect is not related to active management and does not help predict stock returns, but it is related to the tendency of a fund to follow a buy-and-hold strategy. Removing the Average Alpha Effect changes the before-cost performance of the typical active fund from positive to negative.

[1] Junbo Wang may be contacted at junbowang@lsu.edu. Wayne Ferson may be contacted at ferson@usc.edu. This paper is developed from and supersedes Wang's PhD thesis at the University of Southern California, entitled "Can weight-based measures distinguish between informed and uninformed fund managers." We are grateful to Wayne Chang, Chris Jones, David Solomon, Scott Joslin, Yan Liu and Jinchi Lv for comments. This paper has also benefitted from workshops at Boston College, the Universite' Laval, Florida International University, the Universities of Southern California, Massachusetts at Amherst, Washington, York, and from presentations at the 2016 Northern Finance Association meetings, the 2017 Research Affiliates Advisory Board, the 2017 Society for Financial Econometrics meetings, the 2017 NBER-NSF time-series conference and the 2017 Southern Finance meetings.

Holdings-based Fund Performance Measures: Estimation and Inference

December 12, 2017

Abstract

This paper introduces a predictive panel regression framework for holdings-based fund performance measures. The approach facilitates the correction of lagged stochastic regressor bias and allows the application of other statistical tools for panel regressions. We find an "Average Alpha Effect" in three of the four classical measures we study. This Effect is a cross-sectional relation between a fund's average portfolio weights and the stocks' average alphas in the benchmark model. The Average Alpha Effect is large in simulated strategies and in data on active US equity funds. It dominates the cross-sectional variation in the classical measures and is stronger in more recent data. The Effect is not related to active management and does not help predict stock returns, but it is related to the tendency of a fund to follow a buy-and-hold strategy. Removing the Average Alpha Effect changes the before-cost performance of the typical active fund from positive to negative.

Introduction

With the advent of widely available data on mutual funds' holdings, measures of fund performance based on holdings have become more widely used. However, no coherent statistical framework for the various measures has been offered, and the literature provides little information about their statistical properties.[2]

We introduce a predictive panel regression framework for holdings-based performance measures, where future stock returns are regressed on a fund's lagged portfolio weights in the stocks. An informed manager's portfolio weights should predict the future stock returns. Depending on the specification, the slope coefficient on the lagged weights is proportional to the portfolio change measure of Grinblatt and Titman (GT, 1993), the Characteristic Selectivity measure of Daniel, Grinblatt, Titman and Wermers (DGTW, 1997), the Conditional Weight Measure of Ferson and Khang (CWM, 2002) or the stochastic discount factor (SDF) measure of Ferson and Mo (FM, 2016).

Treating these "classical" measures as special cases of the panel regression makes statistical tools developed for panels available for holdings-based measures. This perspective reveals a lagged stochastic regressor bias, similar to that described by Stambaugh (1999) and Hjalmarsson (2008). We evaluate alternative approaches for addressing the bias using simulations, and find two bias correction methods that work well.

Our panel regressions are run at the fund level, using all of the stocks held by a fund. Given that the number of observations in the panel is roughly the number of stocks multiplied by the

[2] Ferson and Khang (2002) use simulations to examine conditional weight-based performance measures. Jiang, Yao and Tong (2007) use simulations to study weight-based timing measures. Kothari and Warner (2001) examine the power of the holdings-based measures of Grinblatt, Titman and Wermers (1997). None of these studies uses a panel regression framework. We examine the statistical properties of the various measures using this framework and provide a number of new insights about their properties.

number of time periods in a fund's life, control variables can be easily handled. We also discuss the role of stock and time fixed effects.

Introducing stock fixed effects in the panel regressions isolates an "Average Alpha Effect" in three of the four classical measures (it does not appear in the CWM). The Average Alpha Effect is a cross-sectional relation between funds' time-series average portfolio weights in the stocks, and the stocks' average alphas in the benchmark model. One may view the average alpha effect as a bias to be removed from the measures, or as a valid component of performance. We show how to isolate the Average Alpha Effect empirically and find that the Effect is large and significant. We therefore present a detailed analysis of the Effect in an attempt to understand its interpretation. The evidence shows that the Effect is related to passive strategies like buy-and-hold, and our analysis tilts us in favor of viewing the Effect as a bias, but we think that the evidence also admits the alternative view.

The Average Alpha Effects appears as a large component of the three classical holdings-based performance measures for active equity mutual funds. The Effects are also large in simulated buy-and-hold fund strategies and they are larger in more recent than in older data. This is interesting in view of studies that find passive fund management and "closet indexing" to be more prevalent in more recent data (e.g. Cremers and Petajisto (2009), Kim, 2016). We find the Effect to be stronger in funds with a stronger tendency towards a buy-and-hold strategy. The Average Alpha Effects are also large in simulated momentum strategies, and they appear in passive index funds. The aggregate holdings of US active equity funds embeds an Average Alpha Effect that ranges from 0.3% to 1.8% per year across the three classical measures that have the Effect.

A striking result is that much of the cross-sectional variation when funds are sorted on their classical performance measures, is driven by their Average Alpha Effects. We also examine the Average Alpha Effects in cross-sections of funds formed in relation to several well-known proxies for active management. We find no clear relation to active management. We investigate the ability of the holdings of funds, sorted by their Average Alpha Effects, to predict future stock returns. We find no evidence of predictive information for future one-month to two-year abnormal stock returns.

We examine the relation of the Average Alpha Effects to characteristics of funds that investors might have preferences over. We run regressions of funds' Average Alpha Effects on fund size, dividend yield, age, turnover, expense ratio, the average market capitalization of the fund's holdings and an indicator for an aggressive growth style. We also include a measure of the tendency of a fund toward momentum trading similar to Grinblatt, Titman and Wermers (1995). We find that funds with a greater tendency toward buy-and-hold tend to have larger Average Alpha Effects. The Effects are also associated with higher dividend yields, but only weakly correlated with the other characteristics.

We run regressions of funds' new money flows on their Average Alpha Effects and fund characteristics. When the Effects are in the regressions alone, flows seem to respond to the Effects, but when we control for the other characteristics, the Average Alpha Effects become insignificant in the flow performance regressions.

If one makes the assumption that the Average Alpha Effect is a passive component of performance, or if it is a bias that should be removed from the measures, our analysis changes the overall inference about active mutual fund performance. Previous studies using holdings-based measures typically find positive performance in actively managed mutual funds on a before-cost basis. The performance is of a magnitude similar to funds' expense ratios (see the review and evidence in Wermers, 2000). At the same time, the abnormal after-cost returns for investors are close to zero. This conforms to a view of the mutual fund industry, advocated by Berk and Green (2004), where fund managers with skills at active management have positive before-cost performance, but face increasing costs with scale and leave no abnormal returns for investors after costs. When we remove the Average Alpha Effect from the classical holdings-based performance measures, the remaining before-cost performance averaged by fund groups and for the median fund, is negative under each of the measures. This evidence suggests that there may not be positive performance before costs, attributed to skilled active management, in the typical mutual fund.

Our finding that the Average Alpha Effect is a dominant component of measured performance may have been foreshaddowed to some extent by earlier work. Grinblatt, Titman and Wermers (1995) find that growth style mutual funds in particular tend to follow momentum strategies, and that much of their performance can be attributed to momentum. Momentum carries a positive average alpha in the CAPM and other models, so momentum trading can be a source of an Average Alpha Effect. However, we find strong Average Alpha Effects even in the DGTW (1997) model, which includes a factor to control for momentum. A fund's tendency toward buy-and-hold is more strongly related to its Average Alpha Effect than its tendency towards momentum trading.

The rest of the paper is organized as follows. Section 2 introduces the holdings-based measures of performance that we study and describes the Average Alpha Effect. Section 3 presents the predictive panel regression approach and shows how the classical holdings-based measures are special cases. Incorporating fixed effects in the regressions, we isolate the Average Alpha Effects. We discuss the lagged stochastic regressor bias, and address the bias in our context using results from Hjalmarsson (2008, 2010) and a differenced instrumental variables approach similar to Anderson and Hsiao (1981) and Wang (2015). Section 4 describes the data. Section 5 presents empirical results for active US equity mutual funds. Section 6 concludes the paper. An Appendix presents our simulation methods, simulation results and other ancillary results.

2. Holdings-Based Performance Measures

Denote the portfolio weights of a fund with N stocks at time t as $\mathbf{w}_t = [w_t^1, \cdots, w_t^N]'$, and denote the next-period stock returns in excess of a Treasury bill, as $\mathbf{r}_{t+1} = [r_{t+1}^1, \cdots, r_{t+1}^N]'$. Holdings-based performance measures are versions of $cov(\mathbf{w}_t{}'\mathbf{r}_{t+1}) = \sum_{i=1}^{N} cov(w_t^i, r_{t+1}^i)$, the sum of the covariances between the current weights and the future stock returns. Grinblatt and Titman (1993) show that in a model with normally distributed returns, an agent with nonincreasing absolute risk aversion and an informative signal about future stock returns, will display a positive holdings-based measure. It is important to sum the covariances across the assets, because a manager might overweight some stocks and underweight others to implement an informed portfolio strategy, so the covariances between subsets of the stocks' returns and their weights might not be positive even when a fund has information.

From the definition of covariance we can write:

$$cov(\mathbf{w}_t{}'\mathbf{r}_{t+1}) = E(\mathbf{w}_t{}'(\mathbf{r}_{t+1} - E(\mathbf{r}_{t+1}))) = E((\mathbf{w}_t - E(\mathbf{w}_t))'\mathbf{r}_{t+1}). \tag{1}$$

Thus, holdings-based measures can be computed by de-meaning the porftolio weights or the returns, or both. Versions of all three approaches appear in the literature. In place of the expected weight we typically find a benchmark weight, and in place of the expected return we find a benchmark return.

Holdings-based measures are estimated as versions of (1/T) $\Sigma_i \Sigma_t w^i_t r^i_{t+1}$, where either the weights w^i_t or the returns r^i_{t+1} are demeaned with a benchmark, or both are demeaned. The measures that we study in this paper are:

$$GT = \frac{1}{T-\tau}\sum_{t=\tau+1}^{T}((\mathbf{w}_t - \mathbf{w}_{t-\tau})'\mathbf{r}_{t+1}). \tag{2a}$$

$$DGTW = \frac{1}{T}\sum_{t=1}^{T}(\mathbf{w}_t'(\mathbf{r}_{t+1} - \mathbf{r}_{t+1}^{D,t})), \tag{2b}$$

$$\text{CWM} = \frac{1}{T}\sum_{t=1}^{T}(\mathbf{w}_t - \mathbf{w}_{bt})'(\mathbf{r}_{t+1} - E(\mathbf{r}_{t+1}|\mathbf{Z}_t)), \tag{2c}$$

$$\text{FM} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{w}_t'\mathbf{r}_{t+1}(a-\mathbf{b}'\mathbf{r}_{Bt+1}), \tag{2d}$$

Equation (2a) is the portfolio change measure of Grinblatt and Titman (1993). This is an example of demeaning the weights. The weight of the fund τ periods before the current period proxies for the expected weight. In this measure a manager records performance when the current portfolio $\mathbf{w}_t'\mathbf{r}_{t+1}$, achieves higher average hypothetical returns than the past portfolio weights, $\mathbf{w}_{t-\tau}$, would earn on the same returns.[3]

Grinblatt and Titman (1993) discuss the choice of the lag, τ. If τ is too small, the past weights might still contain information about the future stock returns, leading to an underestimation of the information in the current weights. If τ is too large, the portfolio's risk might change between the two dates and the measure, because it involves no risk adjustment, might be biased. We adopt the same criteria as Grinblatt and Titman (1993). We use $\tau = 4$ with quarterly data and $\tau = 12$ with monthly data.

[3] The returns are hypothetical because the weights are based on, say quarterly, snapshots of the actual fund holdings and the measured stock returns ignore all trading costs and other fund costs.

Equation (2b) is the Characteristic Selectivity measure of Daniel, Grinblatt, Titman and Wermers (DGTW, 1997). This is an example of demeaning the stock returns, using the benchmark return vector $\mathbf{r}_{t+1}^{D,t} = [r_{t+1}^{D_1,t}, \cdots, r_{t+1}^{D_N,t}]'$, matched to each stock's size, book/market ratio and past momentum. Intuitively, a manager is informed if the hypothetical portfolio of stocks can beat the DGTW benchmark portfolio returns.[4]

Equation (2c) is the Ferson and Khang (2002) Conditional Weight-based Measure. This is an example of demeaning both the stock returns and the portfolio weights. The benchmark weight vector, $\mathbf{w}_{bt}$, is the actual weight from τ periods ago, updated with a buy-and-hold strategy: $w^i_{bt} = w^i_{t-\tau} \Pi_{j=1,\ldots\tau} [R^i_{t-\tau+j} / \Sigma_i w^i_{t-\tau+j-1} R^i_{t-\tau+j})]$, where R^i_t is the gross (one plus the rate of) return of stock *i* at the subscripted date. The assumption is that, under the null hypothesis of no ability, the manager is expected to use a buy-and-hold strategy. The stock returns are demeaned using $E(\mathbf{r}_{t+1}|Z_t)$, the conditional mean returns given standard lagged instruments, Z_t. The conditional expected returns are estimated using regressions of returns on the lagged instruments. We use the same lagged instruments as in Ferson and Khang in our illustrations.

The intuition for the CWM is that a truly informed manager should depart from a buy and hold strategy when she can predict returns, over and above their predictability using public information. A fund delivers performance in the CWM when the portfolio's hypothetical unexpected return (based on the public information) exceeds that of the buy-and-hold benchmark. The hypothesis that the CWM is zero assumes semi-strong form efficient markets in the sense of

[4] The DGTW measure is one term in a decomposition of the GT measure. The other terms refer to factor timing and average style exposure (see DGTW, 1997). The DGTW benchmark return for each stock is constructed as follows. First, stocks are ranked by firm size and divided into five size groups, with each group having the same number of stocks. Within each size group the stocks are ranked by their market-to-book values, and divided into five market-to-book groups. Finally, in each of the 25 groups, the stocks are sorted by their average returns during the past months (*t*-2 to *t*-12) before the current month *t*, and split into five groups according to their past average returns. This produces 125 stock groups, each containing the same number of stocks. The value-weighted returns of the stocks in each of the 125 groups become the DGTW benchmark returns. Each stock is assigned one of the 125 benchmarks based on the closest match to its size, book/market and past returns.

Fama (1970), giving managers no credit for the mechanical use of the public information in Z_t.

Equation (2d) is the Ferson and Mo (FM, 2016) measure. This measure is equivalent to the SDF alpha, $\alpha_p = E(m_{t+1}\ r_{pt+1})$, where $r_{pt+1} = \mathbf{w}_t`\mathbf{r}_{t+1}$ is the fund's hypothetical portfolio excess returns and m_{t+1} is the stochastic discount factor. Ferson and Mo assume a linear factor model for the SDF:

$$m_{t+1} = (a-\mathbf{b}'\mathbf{r}_{Bt+1}), \tag{3}$$

where $\mathbf{r}_{Bt+1}$ is a vector of benchmark excess returns. Thus, the FM measure replaces r^i_{t+1} with the risk-adjusted excess stock returns, $r^i_{t+1}\,(a-\mathbf{b}'\mathbf{r}_{Bt+1})$, and is an example of demeaning the stock returns. A fund delivers abnormal performance in the FM measure by over-weighting stocks with subsequently high risk-adjusted returns and under-weighting those with low risk-adjusted returns. Ferson and Mo (2016) consider different choices for the benchmark returns, $\mathbf{r}_{Bt+1}$, including the Carhart (1997) four factor model that we use here.

It is important to keep in mind that all of the holdings-based performance measures are on a *before-cost* basis. They are designed merely to capture the information in a fund's portfolio weights about the future stock returns. They do not reflect the returns to investors, who must bear funds' turnover-related trading costs, expense ratios and the impact of funds' trading between reporting dates.

2.1 The Average Alpha Effect

While the theoretical holdings-based measures are stated as covariances, they have been implemented as versions of of $(1/T)\ \Sigma_i\ \Sigma_t\ w^i_t\ r^i_{t+1}$, where the r^i_{t+1} and w^i_t variables are not necessarily mean zero. This embeds an *Average Alpha Effect* in the measures. To illustrate, if the predictive covariance is E{**w**'[**r**-E(**r**)]} and the returns are demeaned, but the proxy for E(**r**) used in the estimation, Ê(**r**), is such that $\boldsymbol{\alpha} \equiv$ E[**r**- Ê(**r**)] ≠ 0, then E{**w**'[**r**- Ê(**r**)]} = Cov{**w**'**r**} + E(**w**)'$\boldsymbol{\alpha}$. The second term is the Average Alpha Effect. We find that the Average Alpha Effect looms large in three of the four classical measures (GT, DGTW, and FM).

In the DGTW measure the stocks' alphas are given by $\boldsymbol{\alpha}$ = E(**r** - $\mathbf{r}^D$), where $\mathbf{r}^D$ is the N vector of stocks' benchmark returns, and the expected value of (2d) is E(**w**)'$\boldsymbol{\alpha}$ + Cov(**w**'(**r** - $\mathbf{r}^D$)). The Average Alpha Effect can be nonzero when the alphas of the stocks in the DGTW model are not zero, which we find to be the case.

In the FM measure the SDF is m = (a - **b**'$\mathbf{r}_B$) and the N-vector of stocks' alphas in the model is $\boldsymbol{\alpha}$ = E(**r***m) and E(FM) = E(**w**'**r***m) = E(**w**)'$\boldsymbol{\alpha}$ + Cov(**w**'**r***m). The first term on the right hand side is the Average Alpha Effect and the second term is the pure predictive covariance.

The Average Alpha Effect does not appear in the CWM, because in that model the adjusted return is an N-vector of regression residuals, with mean zero both in the population and when fit in the sample.

The GT measure illustrates demeaning the weights, where ($\mathbf{w}_t$ - $\mathbf{w}_{t-\tau}$) replaces (**w**-E(**w**)) in Equation (1). The Average Alpha Effect is E{($\mathbf{w}_t$ - $\mathbf{w}_{t-\tau}$)'$\boldsymbol{\alpha}$}, which will be small when the mean weight change is small, but positive if the weights drift toward higher-alpha stocks, which we find is the typical case.

2.2 Intperpreting the Average Alpha Effect

The Average Alpha Effect, Σ_i E(w^i)α_i , captures the cross-sectional relation between the mean weights in the stocks held by the fund and the stocks' average alphas. The predictive covariance, without the Average Alpha Effect, is the focus of the original Grinblatt and Titman (1993) theory that motivates holdings-based measures.[5] The Average Alpha Effect captures a differrent cross-sectional relation. As such, its interpretation may be debated.

One view is that the Average Alpha Effect is simply a bias that should be removed from the measures. If the weights are properly demeaned, then the Average Alpha Effect is zero. If the stocks alphas are zero, the Average Alpha Effect is zero. One can view nonzero alphas in the factor model simply as a misspecification of the factor model. We find that passive index funds have nonzero average alpha effects. The aggregate holdings of funds embeds Average Alpha Effects. We also find using simulations, that clearly uninformed strategies like buy-and-hold, can have large Average Alpha Effects. These facts seem to support the view that the Average Alpha Effect is a bias.

A different view is that the average alpha effect is a valid component of performance. Measuring the expected product of the adjusted returns or weights through $(1/T)\Sigma_i\Sigma_t w^i_t\, r^i_{t+1}$ lends the classical measures a natural and appealing interpretion as an average excess return relative to that of a benchmark. It is natural to interpret an average return in excess of the benchmark as performance. The Average Alpha Effect can be a large component of the return difference. A positive Average Alpha Effect says that the fund on average, overweights the high average alpha stocks and underweights the low alpha stocks. Imagine a hypothetical fund with constant weights over time. The alpha of its hypothetical return (formed from its holdings and the stock returns) in a factor model regression will equal its Average Alpha Effect.

---

[5] In a precursor to the holdings-based measures, Grinblatt and Titman (1989) described a "positive period weighting" measure. They assumed that the benchmark was unconditionally minimum variance efficient, so that the average alphas of all the stocks would be zero. If this measure is implemented on a benchmark that is not efficient, it would also generate average alpha effects.

Users of the performance measures might take the benchmark model as given, or not know about the cross-section of stocks' average alphas relative to the model. When the benchmark model must be estimated in real time using past data, there may be a lot of uncertainty about the average alphas. Some "Quant" portfolio strategies are based in part on overweighting the larger average alphas relative to benchmark models. One can argue that funds which can use the cross-section of average alphas to beat the benchmark have a skill that should be measured and even rewarded.

Nonzero alphas are a fact of life for factor pricing models of stock returns, and we have to confront this fact in practice. Stocks' average alphas relative to the DGTW factor model are not zero in the data. Cremers, Petajisto and Zitzewitz (2012) show that passsive indexes like the Russell 2000 have nonzero alphas in the standard factor models used in fund performance evaluation. We find that the average alpha effect drives the cross-sectional variation in the classical performance measures. We show how to use panel regression methods to isolate the Average Alpha Effect from the classical measures. The Effect may be considered as a component of performance or not, depending on the preferences of the user of the performance measure. The inferences about performance will likely be changed if the Effect is removed from the classical measures, and we find that the overall inference about before cost performance is changed if the Average Alpha Effect is removed.

3. The Predictive Panel Regression Approach

Previous studies have employed panel regression methods in mutual fund research. Typically, the left-hand side variables are measured at the fund level. For example, flow-performance studies going back to Ippolito (1992) and Sirri and Tufano (1998) regress the new money flows to mutual funds on past measures of fund performance and various fund-level control variables. Studies put measures of fund performance on the left hand side to investigate its relation to things like fund size (Ferreira, Keswanit, Miguel and Ramos, 2013), fund and industry size (Pastor, Stambaugh and Taylor, (2015), Magkotsios, 2017) or other variables. In contrast, stock returns are on the left hand side of our regressions, and a particular fund's holdings of the stocks are on the right hand side. We estimate a separate panel regression for each fund, obtaining a fund-specific performance measure.

We start with the simplest case. For a given fund, assuming that the portfolio contains *N* stocks and exists for *T* periods, the regression is:

$$r_{t+1}^{i} = \beta w_{t}^{i} + \varepsilon_{t+1}^{i}.,\ i=1,\ldots,N;\ t=1,\ldots T. \tag{4}$$

In this regression, r_{t+1}^{i} is the excess return of stock i at time $t+1$ and w_{t}^{i} is the weight of the stock held in the fund at time *t*. The slope coefficient β captures the ability of the fund's weights to predict future excess stock returns. The estimated β should be positive when the manager is informed. The OLS slope coefficent estimator in the panel regression (4) is:

$$\hat{\beta} = (1/T)\Sigma_i \Sigma_t (r^i_{t+1} w^i_t) / (1/T)\Sigma_i \Sigma_t (w^i_t w^i_{t'}) \tag{5}$$

Because the coefficient β is not stock-specific, the panel regression "automatically" sums across the stocks in a fund's portfolio, precisely as specified by the holdings-based measures.

The numerator of the coefficient in Equation (5) includes all of the holdings-based measures as special cases:

$$(1/T)\ \Sigma_i\ \Sigma_t\ w^i_t\ r^i_{t+1}, \tag{6}$$

when either the weights or the returns or both are demeaned with a benchmark. In particular, using the weight change, $w_{it} - w_{it-\tau}$, in place of w^i_t we obtain the GT measure. Replacing the stock returns with the DGTW benchmark-adjusted returns, we obtain the DGTW measure. Replacing w_t with $(w_t - w_{bt})$ and r^i_{t+1} with $[r^i_{t+1} - E(r^i_{t+1}|Z_t)]$, we obtain the CWM. Finally, replacing r^i_t with $r^i_t\ (a-\mathbf{b}'\mathbf{r}_{Bt+1})$ we obtain the FM measure.

3.1 Introducing Fixed Effects

Equation (4) is unrealistic for stock returns, because under the null hypothesis that β=0, if the residuals are mean zero it implies that all the stocks have the same expected return equal to zero. We therefore introduce individual stock fixed effects, and the panel regression model becomes:

$$r^i_{t+1} = a^i + \beta\, w^i_t + \varepsilon^i_{t+1}. \tag{7}$$

The model is estimated by introducing *N* stock dummy variables, which take the value of one (for each date) if the return belongs to stock *i* and zero otherwise. The coefficient of the dummy variable for stock *i* is a^i, the fixed effect of stock *i*. Under the null hypothesis ($\beta = 0$), the fixed effect is the expected return of the stock.

With stock fixed effects in the model, the Frisch–Waugh–Lovell theorem shows that the OLS slope coefficient estimator of the regression (7) is the same as that obtained by subtracting the time-series means from each of the left and right-hand side variables, and running the regression on the demeaned variables with no intercept. This is the "Within" group estimator. Thus, the numerator of the OLS slope estimator of (7) is equal to:

$$\hat{\beta}_{w,num} = \Sigma_i\, (1/T)\, \Sigma_t\, (r^i_{t+1} - \overline{r^i})(w^i_t - \overline{w}^i), \tag{8}$$

where $\overline{r^i} = (1/T)\, \Sigma_t\, r^i_{t+1}$ and $\overline{w^i} = (1/T)\, \Sigma_t\, w^i_t$. The Within estimator captures the average across the stocks, of the time-series predictive covariances between the weights and the future returns. The numerator of the within estimator is an example of a holdings-based measure that demeans both the returns and the weights, although it is equivalent to de-meaning only the returns or only the weights, according to (8).

3.2 Interpreting Fixed Effects

With stock fixed effects in the model, the slope coefficient does not reflect differences in the average returns across the stocks, only the predictive time-series covariances between the weights and the future returns. We can relate the estimator with the stock dummies to the estimator without dummies using:

$$(1/T)\Sigma_i\, \Sigma_t\, r^i_{t+1}\, w^i_t = (1/T)\Sigma_i\, \Sigma_t\, (r^i_{t+1} - \overline{r^i})(w^i_t - \overline{w}^i) + (1/T)\Sigma_i\, \Sigma_t\, \overline{r^i}\, \overline{w}^i\, . \tag{9}$$

Stock dummies remove the term $(1/T)\, \Sigma_i\, \Sigma_t\, \overline{r^i}\, \overline{w}^i = \Sigma_i\, \overline{r^i}\, \overline{w}^i$ from the OLS slope estimator's numerator. If the time-series sample means of the benchmark-adjusted returns or weights used in the regression are zero for *all* stocks, then the far right hand term of Equation (9) is zero, the estimator with fixed effects is equivalent to the OLS estimator of Equation (5) and there is no

Average Alpha Effect.

In practice, the benchmark-adjusted returns and weights in the classical holdings based performance measures are not mean zero for each stock. We show in the Appendix that in this case, the difference between the numerator of the OLS estimator in (5) and that of the slope coefficient in the model with fixed effects is a good proxy for the Average Alpha Effect.

One might consider including time dummies in the panel regression. By the Frisch–Waugh–Lovell theorem the OLS slope coefficient in the regression with time dummies is the same as the one obtained by subtracting the cross-sectional mean values from both the left and right-hand side variables, and running the regression with the demeaned variables and no intercept. The cross-sectionally demeaned returns are the returns net of the return for an equally-weighted portfolio of the stocks. Since the weights in a fund's portfolio sum to 1.0, the cross-sectional mean of the weights is (1/N). Thus, the numerator of the basic panel regression (4) is equal to the numerator of the regression with time dummies, less the time-series average return of an equally weighted portfolio of the stock returns. The impact of time fixed effects on the slope β in this problem is exactly the same as the impact of including a common intercept in the panel regression.[6]

It is common in panel regressions to include stock-specific control variables, x_{it} as additional explanatory variables. Given that the panel regressions estimating the holdings-based performance measures have approximately TN observations, there will be enough degrees of freedeom for several control variables in actual practice. (The average fund holds N=114 stocks

[6] Hajmarlsson (2008) proposes a panel regression with a common intercept: $r^i_{t+1} = a + \beta w^i_t + \varepsilon^i_{t+1}$. With a common intercept in the regression, the Frisch–Waugh–Lovell theorem shows that slope coefficient estimator is the same as that obtained by subtracting the overall time-series and cross-sectional means from the left and right-hand side variables, and running the regression on the demeaned variables with no intercept. The numerator of the pooled OLS estimator is equal to the numerator of the OLS estimator of (4), minus the average over time of the return on an equally-weighted portfolio. While the model with a common intercept offers little economic insight, it has been suggested as a solution to finite sample bias by Hjalmarsson (2008), so we examine it below in that context.

and there are T=128 quarters in a sample for 1980-2012.) Consider a regression including the term $\boldsymbol{a}'\boldsymbol{x}_{it}$, where $\boldsymbol{a}$ is an L-vector of coefficients. The coefficients are the same for each stock. Using the Frisch–Waugh–Lovell theorem, it is easy to show that including $\boldsymbol{x}_{it}$ removes the information in the weights about the stock-specific control variables and finds any performance from the remaining information in the weights.

3.3 Standard Errors

The basic panel regression, with fixed effects or other considerations, is essentially the model examined by Petersen (2009), except that the predictor variables here are endogenous, predetermined and persistent. Petersen focusses on panel standard errors for OLS slope coefficient estimators under various models for the error terms and various clustering strategies. This machinery may now be applied to holdings-based performance measures. The literature on holdings-based measures has not emphasized its standard error estimators. DGTW use the time-series variance of the CS measures for each period, $CS_t = (\mathbf{w}_t'(\mathbf{r}_{t+1} - \mathbf{r}_{t+1}^{D,t}))$, to compute standard errors. Ferson and Khang (2002) use time-series GMM-derived standard errors for their CWM. Seeing the measures as the result of a panel estimation allows future research to compute standard errors for holdings-based measures using results from panel econometrics.

We defer a detailed study of the accurracy of various standard error estimators to future work, and use simple panel methods to compute the standard errors in this study. Since stock returns are correlated at a point in time, but have little time-series correlation, we cluster the standard errors by time. When we estimate the panel using recursively demeaned or differenced estimation as described below, we include autocovariance terms as in Newey and West (1987) to accommodate the induced serial dependence.

3.4 Lagged Stochastic Regressor Bias

Our approach is to estimate holdings-based performance through predictive panel regressions. Predictive panel regressions are affected by a bias due to lagged stochastic regressors, studied by Hjalmarsson (2008, 2010). This bias, also known as the "Stambaugh bias," is related to the persistence of the portfolio weights and the correlation of their future innovations with the stock returns. The Stambaugh bias is examined in time-series regressions for stock market returns by Stambaugh (1999), Pastor and Stambaugh (2009), Amihud and Hurvich (2004) and Amihud $et.al.$ (2008, 2010). We examine three alternative methods to address the bias. These include a parametric bias correction (Hjalmarsson, 2008), a recursively demeaned instrumental variables approach, (Hjalmarsson, 2010) and a differenced instrumental variables approach (Anderson and Hsiao (1981), Wang, 2015).

To capture the fact that the fund's portfolio weights are highly autocorrelated, the parametric bias correction assumes that they follow an $AR(1)$ model:

$$w_{t+1}^{i} = \gamma^{i} + \rho w_{t}^{i} + v_{t+1}^{i}. \tag{10}$$

Equations (7) and (10) form a panel "predictive system" (e.g., Pastor and Stambaugh, 2009).

The lagged stochastic regressor bias in the panel regression with fixed effects may be understood as follows. Consider the numerator of the least squares dummy variable estimator, written with only the weights demeaned:

$$\hat{\beta}_{w,num} = (1/T)\Sigma_i \Sigma_t \ r^i_{t+1} (w^i_t - (1/T)\Sigma_t w^i_t), \tag{11}$$

Substituting from Equation (7) for the returns yields:

$$E(\hat{\beta}_{w,num} - \beta_{num}) = -E[(1/T)\Sigma_i \Sigma_t \varepsilon^i_{t+1} (1/T)\Sigma_{\tau\geq t+1} w^i_\tau]. \tag{12}$$

Equation (12) shows that, on the assumption that the shocks to the stock returns are uncorrelated with lagged weights, the lagged stochastic regressor bias arises if ε^i_{t+1} is correlated with $\Sigma_{\tau\geq t+1} w^i_\tau$. The implicit demeaning of the variables, because of the intercepts in the regresssion, introduces the future weights. Under a buy-and-hold portfolio strategy, for example, we would expect a positive correlation between ε^i_{t+1} and w^i_{t+1}, as a positive return shock increases the portfolio weight, and therefore we expect a negative bias. The larger is the serial correlation in the weights, the larger is the expected bias as the shocks at time *t+1* accumulate in the future weights, w^i_τ, $\tau\geq t+1$.

The Appendix describes several methods for correcting the lagged stochastic regressor bias and provides simulation analyses of their efficacy. Two methods perform well in samples of the sizes in this paper. One is a recursive demeaning approach from Hjalmarsson (Haj, 2010) and the other is a differenced instrumental variables approach from Anderson and Hsiao (1981) and Wang (Diff IV, 2015). We use these methods in our subsequent analyses.

3.5 Scaling

The panel slope coefficients are proportional to the various holdings-based performance measures. Thus, the coefficients will be zero under the null hypothesis that the performance is zero. In order to test the null hypothesis of zero performance, the unbiased panel slope coefficients may be used directly. However, the holdings-based measures are the numerators of the slope coefficients, and these are set in economically meaningful units. For example, the FM measure is a certainty equivalent excess return for an agent with the SDF used in the model, economically equivalent to an extra risk-free return earned by the fund. Thus, we scale the panel slope coefficients to the scale of the original measures. This is a simple matter of multiplying the left-hand side variables by the variance of the weights, such as $(1/T)\ \Sigma_i\ \Sigma_t\ w^i_t{}^2$ in the base case of Equation (4). For the Within estimator we use the de-meaned weights, and the appropriate weight variances for the other cases. We carry this scaling convention to our simulations and evaluate the sampling properties of the scaled measures.

4. Data

We obtain quarterly holdings data for mutual funds from Thomson-Reuters. The holdings data cover 1980 to 2012. To avoid some standard biases, we employ several screening methods to filter the data. We exclude data before 1984 in most of our analyses, since Fama and French (2010) show that there is a selection bias in mutual fund data before 1984. Since most of the research papers using holdings data focus on US equity funds, we exclude other types of mutual funds. Evans (2010) discusses an incubation bias in fund performance measures, and following his suggestions, we exclude observations before the reported date of fund organization or when a fund first has total net asset value (TNA) of less than 15 million dollars. There are in total 3596 equity funds in this sample. We also collect the holdings for a sample of 201 index funds for 1994-2012, identified by either the CRSP index fund flag or by searching for the string “index” in the funds’ name.

In addition to the holdings data, we use monthly prices and returns of individual stocks from the CRSP monthly stock file. We obtain the delisting returns to deal with cases where firms go out of the market. The sample contains stocks with at least one month of returns from 1970 to 2012. The DGTW benchmark returns are collected from Russ Wermers' website, and are then combined with the monthly stock file[7]. For some of our analyses we use daily stock returns and mutual fund returns from CRSP, available starting in 1999 for the mutual funds. We also use daily data on interest rates from the Federal Reserve Data base.

We use holdings and stock prices to construct portfolio quarterly and monthly weights for the mutual funds. We first describe the quarterly weights. Let the holdings of the stocks (measured as the number of shares held) and stock prices at the t'th quarter be $\mathbf{h}_t$ and $\mathbf{p}_t$, where $\mathbf{h}_t = [h_{1t}, \cdots, h_{Nt}]'$, and $\mathbf{p}_t = [p_{1t}, \cdots, p_{Nt}]'$. The weights of the stocks in the fund portfolio are $w_t^i = \dfrac{h_{it}\, p_{it}}{\boldsymbol{h}_t' \boldsymbol{p}_t}$.

To construct monthly weights we assume that for the months between consecutive reporting dates the funds keep holding the same number of shares. The monthly weights at month t are constructed as the product of a fund's preceding reported holdings and the month t prices of the held stocks. The with-dividend stock returns are used, implicitly assuming that a fund reinvests the dividends in the same stock. Weights constructed in this way are used by DGTW (1997), Kacperczyk, Sialm and Zheng (2008), Busse and Tong (2012), Amihud and Goyenko (2013) and others. We explore the impact of using quarterly versus monthly weights on our main results in a robustness section and find little change.

Summary statistics for the data are shown in Table 1. In panel A, for each fund in the

---

7 We only select stocks with non missing values of the returns and DGTW benchmark returns. According to Wermers (2004), the stocks selected for the DGTW benchmarks have at least two years of data on book values, returns and market capitalization. We exclude stocks with less than 48 months of data.

sample we compute the time-series average of its total net assets (TNA) in millions of dollars, the average number of stocks held, the return gap from Kacperczyk, Sialm and Zheng (2008) and the active weight as in Doshi et al. (2013). We compute the sample standard deviations of the reported fund returns, σ, and the R-squares of factor model regressions in the Carhart (1997) four-factor model. We also compute the sample autocorrelations of the funds' portfolio weights, ρ_j, averaged across the holdings, at various lags j, j=1,…5. The means, std, maximum and minimum are taken across the funds in the sample. Finally, for each fund we compute the correlation between the errors of the portfolio weight autoregression and the panel regression of future stock returns on the weights, averaged across the holdings. The average correlation is denoted as Error Corr in the table, and the covariance for fund *i* serves as Cov(ε_i,v_i) in the bias-adjusted estimator of Hjalmarsson (2008). The sample period is from 1984 to 2012, and the number of funds is 3596. We exclude the sample of index funds from these summary statistics.

The average fund has total assets of $684 million and holds 114 stocks. As shown by Kacperczyk, Sialm and Zheng (2008), the return gap (the gross of expense-ratio reported returns minus the hypothetical returns formed from fund weights and stock returns), is small on average. However, the maximum and minimum values for the return gap are vary substantially across funds. The average first order autocorrelation of the funds' weights is 0.90, so the weights are persistent time-series. This value is slighly below the value where the simulation evidence in Ferson, Sarkissian and Simin (2003a) indicates that spurious regression becomes a concern. However, the autocorrelation is high enough to make the lagged stochastic regressor bias a concern.

As a reality check for the AR(1) model assumption on the weight process we estimate the coefficients ρ_j in the following panel regressions: $w^i_{t+1} = \gamma^i + \rho_j w^i_{t+1-j} + v^i_{t+1}$, where *j* runs from 1 to 5. If the weights follow an $AR(1)$ process, we should observe that $\rho_j = \rho_1^j$. Panel A of Table 1 suggests this is a good approximation. The values of ρ_1^j, are close to the values of ρ_j (for example, at the mean values, $\rho_1^2 = 0.815$ and $\rho_2 = 0.813$, while $\rho_1^5 = 0.599$ and $\rho_5 = 0.586$). In Panel C of Table 1, we compute the autocorrelation of the first differences in the weights. The first order autocorrelations are small.

The last row of Panel A in Table 1 presents statistics for Error Corr, the sample correlations of two errors (ε_t^i and v_t^i), averaged across the stocks in the predictive regression system for a given fund. While the average value is small, at -0.004, the values can be either positive or negative. The maximum and minimum values are 0.24 and -0.19, respectively, which suggests that the Stambaugh bias can change signs for different funds.

5. Empirical Results for Mutual Funds

We begin our analysis by revisting some of the main results in DGTW (1997) for the GT and DGTW measures, focusing on the results of pulling the Average Alpha Effects out of the measures. Table 2 summarizes, in panel A, the sample period from 1980-1994, ending on the same date as the sample in DGTW (1997). The subscript *H* indicates a model with fixed effects, estimated using the Haj10 bias-adjusted method. Avg α is the difference between the classical estimator and the Haj10 estimator -- our proxy for the Average Alpha Effect. (In the Internet Appendix Table A.3 we report the similar results using the DiffIV estimator.) T-ratios are on the second line, calculated as in DGTW as the time series standard errors of the monthly values of the equally-weighted averages of the measures for the funds in each group in that month. (For the adjusted measures, 30 Newey West lags are used to control for autocorrelation.)[8]

The results in Panel A of Table 2 closely match those reported in DGTW (1997). The GT measure for All Active funds is about two percent per year, with a large t-ratio. The DGTW measures are smaller but positive, ranging from 38 basis points to 1.15% per year across the active fund groups. The GT measures are strongly significant, but the t-ratios for the DGTW measures are all less than 2.0.

The Haj10 estimates of performance under the GT measure in Panel A are 63-93 basis points per year smaller than the classical GT measures, revealing a substantial positive average alpha effect. The t-ratios of the Haj10 measures remain significant; larger than 3.7. Thus, the conclusion of significantly positive before-cost performance in the earlier sample period holds up under the GT_H measure.

Grinblatt and Titman (1993) note that the GT measure might show performance if funds follow momentum strategies. DGTW (1997) point out that while the DGTW measures should not have exposure to the size, book/market and momentum factors, in practice they might because the benchmarking may be imperfect. Following DGTW (1997), we report Carhart adjusted measures, which are the intercepts from regressing the monthly performance measures on the four Carhart (1997) factors.

With results very similar to those reported by DGTW (1997), the classical GT measures in Panel A indicate smaller performance after Carhart adjustment, now between 0.44 and 1.9% per year, although the Aggressive Growth funds' 1.9% is still quite significant. Carhart adjustment reduces both the GT and the GT_H measures. The Average Alpha Effects are positive, between 18

[8] The monthly measures are estimated as follows, taking the DGTW measure as an example. The classical measure is $DGTW_t = \mathbf{w}_{t-1}'(\mathbf{r}_t - \mathbf{r}_t^D)$, where $\mathbf{r}_t^D$ is the N-vector of DGTW adjusted returns. The monthly Haj10 estimator is $(\mathbf{w}_{t-1} - (1/(t-\tau)) \sum_{s=1,\ldots,t-\tau} \mathbf{w}_s)'(\mathbf{r}_t - \mathbf{r}_t^D - (1/(T-t-\tau+1))\sum_{s=t+\tau,\ldots,T} (\mathbf{r}_s - \mathbf{r}_s^D))$ and the monthly Average Alpha Effect is the difference between the two. We use a lag of τ=12 months.

and 50 basis points per year in the GT measures. The DGTW measure is affected less by the Carhart adjustment, posting slightly higher DGTW values and lower $DGTW_H$ values after adjustment. The Average Alpha Effect in DGTW is positive, between 26 and 56 basis points per year after Carhart adjustment.

Panel B of Table 2 examines the 1980-2012 period. Recent studies (e.g. Chen and Ferson (2017), Jones and Mo, 2017) find that mutual fund performance deteriorates in more recent data, and all of the measures reflect that pattern. The raw performance is smaller than in the earlier subperiod by 60 basis points to 1.5% per year under the GT measure. Removing the Average Alpha Effects, which are similar to the values found in Panel A for the GT measures, the performance is smaller still. The Aggressive Growth funds still show the strongest performance, with 2.1% per year under GT_H, with a t-ratio of 2.3 in the recent sample.

The DGTW measures of performance remain smaller than the GT measures in the full sample. The $DGTW_H$ measures are negative for each of the fund groups, and the Average Alpha Effects are positive and larger than in the earlier period, ranging from 0.51% to 1.35% per year across the groups. (Using DiffIV it is 0.30 to 0.66% per year). Thus, the Average Alpha Effects are positive and large in the more recent data under both measures.

We include a row for index funds in Panel B of Table 2, covering 201 finds during the 1994-2012 period. The index fund performance is close to zero under GT and GT_H. The performance is slightly larger after Carhart adjustment, and the Average Alpha Effect in the GT measure is small after Carhart adjustment. The DGTW measure attaches a large Average Alpha Effect to the index funds, 1.35% per year, and this is not changed by the Carhart adjustment.

Overall, while the Average Alpha Effect in the classical measures would likely not change the conclusions of the original DGTW (1997) paper for data up to 1994, in more recent data the

Average Alpha Effect is more important. Removing it flips the signs of the performance under the DGTW measure from positive to negative. The performance net of the Average Alpha Effect is negative for each of the fund groups, in three of the four cases presented in Panel B of Table 2. The Average Alpha Effect has become larger in more recent data. It is also large for the index funds under the DGTW measure after 1994. Thus, in the more recent sample the overall inference about before-cost performance of the mutual funds is more pessimistic and is strongly influenced by the Average Alpha Effect.

5.1 Average Alpha Effects in Simulated Strategies

To help better interpet the Average Alpha Effects we simulate strategies for hypothetical mutual funds. The first strategy is a buy-and-hold strategy, motivated by the results for index funds in Table 2, where large positive Average Alpha Effects are found under the DGTW measure. A positive Average Alpha Effect is expected under a buy-and-hold strategy, as the weights "automatically" drift toward the higher average return stocks. In our sample of funds, the correlations between the estimated average alphas of individual stocks and their sample mean returns vary from 0.71 to 0.84 across the various models, so when the weights drift toward higher average returns they likely drift toward larger average alphas.

The simulations also examine a stylized momentum strategy. Results in Table 2 are consistent with the claim of Grinblatt and Titman (1993) that the GT measure can record positive performance if a fund uses momentum, and the positive GT performance is removed after adjustment using the Carhart factors, which include a momentum factor. The Average Alpha Effect in the GT measure is smaller after adjustment with the Carhart factors, suggesting a relation between momentum and the Average Alpha Effect.

Jointly simulating stock returns and portfolio weights for the hypothetical strategies introduces some complications. These include missing data patterns and realistic restrictions on funds' portfolio weights. The Appendix describes how we handle these complications.

5.2 Results for Simulated Strategies

Using the bootstrapped returns and the simulated strategy weights we estimate the various performance measures and calculate the standard errors and T-statistics in each simulation trial. We simulate monthly weights and returns. By simulating the returns and strategies 1000 times, we have 1000 point estimates, standard errors, and T-ratios for each case.

The Average Measures in the first rows of Table 3 are the averages across the 1000 simulated performance measure estimates. The expected values of the GT, DGTW and FM measures are positive for both the buy-and-hold and momentum strategies. Consistent with the claim of Grinblatt and Titman (1993) the GT measure does record positive performance under the momentum strategy, averaging 0.42% per year. Under buy-and-hold, the GT performance is even larger, at 1.04% per year. Even though the DGTW measure includes a momentum adjustment, the performance according to the classical measure is more than 7½ percent per year under the momentum strategy and 9.4% for buy-and-hold. The FM measures are also larger than 9% per year under buy-and-hold.

The average Diff IV and Haj10 estimators in Table 3 are less than ten basis points per year for most of the measures. Thus, the performance of these hypothetical strategies as captured by the classical measures is driven by their Average Alpha effects.

It is striking that the Average Alpha Effects are so large, and even larger under buy-and-hold than under momentum. We see much smaller effects in the mutual fund data. In the mutual fund data we examine averages; either portfolios of funds each month (Table 2) or the estimates for individual funds averaged within groups (tables 4 and 5 below). One reason these averages show smaller effects is that the Average Alpha Effects can be negatively skewed in the cross-section of individual funds. We calculate the skewness to be -9.62 in the GT measure, -3.51 in DGTW and -6.6 in FM. Extreme negative skewness reduces the average values taken across funds.

With 1000 T-ratios, we can study the distributional properties of the T-statistics for the performance measures. We rank the 1000 simulated T-ratios and select the 25-th, 50-th, 950-th and 975-th values and report them in Table 3. The distributions of the T-ratios show that all of the classical performance measures would find significant performance in a large fraction of the cases (nearly 100% for DGTW), and this is driven by the positive average alpha effects under buy-and-hold or momentum. When the average alpha effect is removed using the Diff IV or Haj10 estimators, the distributions of the T-ratios appear slightly more peaked than a normal, with slighly thinner tails. However, the critical values of the T-statistics are pretty close to the corresponding values for the standard normal distribution. Although the classical measures are strongly influenced by the average alpha effect, once it is removed the distributions of the simulated T-statistics are close to standard normal.[9]

Overall, Table 3 shows that simulated buy-and-hold or momentum strategies can cause large Average Alpha Effects in the classical measures. Under the DiffIV or Haj10 measures, neither strategy shows abnormal performance, so all of the performance for the simulated

[9] We test whether the distributions of 1000 T-ratios for each performance measure are normal following the Jarque-Bera (JB 1986) test and Lilliefors (1967) test . The JB test compares the skewness and kurtosis of the distributions with those of a normal distribution. The Lilliefors test examines the difference between the CDF of the sample distribution and that of a normal distribution. The JB and Lilliefors tests show that, for most of the measures after the aveage alpha effect is removed, the distributions of the T-statistics are not significantly different from those of a normal.

strategies comes from the Average Alpha Effect. To the extent that the Average Alpha Effect is driven by buy-and-hold strategies in actual mutual funds, this tilts us toward the bias interpretation of the effect.

5.3 The Cross-sectional Variation in Fund Performance Measures

So far, the analysis has focused on broad groups of funds or hypothetical strategies, and we saw that the classical measures and the measures net of the Average Alpha Effects can lead to different inferences. Now we turn to cross-sections of individual funds. Correlations taken across funds (Internet Appendix, Table A.1) reveal that the classical measures and their Haj10 estimates have modest correlations, ranging from 0.08 (FM) to 0.56 (GT). Thus, inferences about the performance of individual funds will likely be different with the different measures.

We first sort funds into five quintiles on the basis of the classical performance measures, shown as the first column of Table 4. The classical performance measures are estimated by pooled OLS on monthly data. We then estimate the model with fixed effects for each fund using the Haj10 estimator and average the estimates across the funds in each quintile. In the Internet Appendix Table 5A we use the differenced IV approach and the results are similar. We discuss quarterly estimates in a robustness section. Panel A of Table 4 shows that the average style exposure change component of the GT measure is small, and since this component is absent from the other measures, it further justifies our use of the difference between a classical measure and the fixed effects estimate as the Average Alpha Effect.[10]

The bottom rows of each panel of Table 4 show results for a group of 201 index funds, for 1994-2012. The DGTW measures characterize the index funds as having poor selectivity

[10] To calculate the average style change effect in the GT measure, we estimate the betas for each stock using daily stock returns with the carhart four factors as the benchmark, and compute $(\Sigma_i \overline{\Delta w^i} \beta_i)' \overline{r_B}$.

performance in $DGTW_H$ (-0.09% to -0.10% per year) masked by a positive average alpha effect of 0.11%. The sum of the two parts in the classical DGTW is near zero. The index funds produce negative Average Alpha Effects in the GT and FM measures.

The approach in Table 4 weights the underlying fund-period observations differently from Table 2. In Table 2 a portfolio of funds is formed each month, and the monthly time-series of the portfolio measures are averaged over time. Each month gets the same weight in the average, independent of how many funds are present that month. In Table 4 we estimate a single value of the measure for each fund using its time-series data, and then average across the funds in each quintile. This gives each fund equal weight. Because there are many more funds later in the sample period, Table 4 puts greater weight on the more recent data than Table 2. Thus, we expect that some of the differences observed between panels A and B of Table 2, because of more recent data in Panel B, should be found in Table 4.

The index funds illustrate how the differences in weighting influence the results. In Table 4 we calculate a measure for each fund over its existing sample period and then take the cross-sectional average. This is not likely to be a good proxy for a buy-and-hold portfolio, such as examined in Table 3. If instead we calculate a cross-sectional average index fund at each month, and take its average measure over time as in Table 2, the index funds show large positive Average Alpha Effects in the GT and especially, the DGTW measures. The same calculation for the FM and FM_H measures produce 0.3% and -2.10% per year, and thus large positive Average Alpha Effects for index funds. Many index funds started after 2000, and their Average Alpha Effects are negative and large during the financial crisis, as many stocks had negative alpha estimates. Like the skewness affect described above, if we average the measures across funds as in Table 2, we get small or even negative Average Alpha Effects. But if we average a portfolio of the measures over time, the financial crisis only impacts a short interval during the 20 year time series, and the Average Alpha Effects are larger for the index funds.

In Table 4 the performance under the GT measure is lower on average, and the Average Alpha Effects in the GT measures are smaller than in Table 2. Also, consistent with the greater weight on more recent data, the performance in Table 4 under the classical DGTW measure is worse than in Table 2, and the Average Alpha Effect is positive. The spread across the quintiles shows that the Average Alpha Effects have negatively-skewed distributions, as previously described, which has the effect of making the average effects smaller.

The third quintile in Table 4 contains the median fund, and finds that the Average Alpha Effect is positive under every classical measure. This is consistent with previous holdings-based studied that find positive before-cost performance for the typical fund (e.g., Wermers, 2000). At the same time, studies find that the after cost performance of the typical mutual fund is zero or negative (e.g. Chen and Ferson, 2017), which is consistent with the view of the money management industry taken by Berk and Green (2004). However, Table 4 shows that the median fund records negative performance under every performance measure when the Average Alpha Effect is removed. For this to be consistent with the Berk and Green (2004) view of the fund management industry, the Average Alpha Effects should be associated with skilled fund management.

Table 4 presents the average differences between the performance measures for the top and bottom quintiles (HML), summarizing the cross-sectional variation. Sorts on the classical measures produce HMLs of 0.64% per month for GT, 0.55% for DGTW and 2.31% for FM. The *H* measures show smaller HML spreads, ranging from 0.09 to 0.34% per month.

The GT and GT_H estimates show similar patterns across the GT quintiles, until we find the lowest performance group with a large, negative average alpha effect. When the average alpha effect is removed in GT_H, the lower quintile performance changes from -0.29% per month to only -0.08% per month. Thus, the poor performers under GT experience large negative average alpha effects which capture most of their poor performance. This means that they increase their holdings on average in the lower average alpha stocks.

The Average Alpha Effect in the DGTW measure has a large amount of cross-sectional variation, ranging from -0.19% to +0.25% per month across the DGTW-sorted quintiles, and it accounts for much of the differences in the classical DGTW quintiles. The good performers under DGTW have positive Average Alpha Effects, while the bad performers have negative Average Alpha Effects. The sample correlation across individual funds between their classical DGTW measure and its Average Alpha Effect is 0.69. Sorting funds on their classsical DGTW measures is a lot like sorting them on the their Average Alpha Effects.

Panel C of Table 4 presents results for the conditional weight measure, CWM. The classical measure is $(1/T)\Sigma_i\ \Sigma_t\,((r^i_{t+1} - r^{i,C}_{t+1})(w^i_t - w_{ibt})$, with the conditional mean of the stock returns as benchmark return: $r^{i,C}_{t+1} = \delta^{i\,'} Z_t$., which is estimated by a regression in the first step. There is no Average Alpha Effect in the original CWM, but there is a lagged stochastic regressor bias due to the regression that predicts the stock returns. We therefore present only an unbiased version of the CWM, estimated using the Haj10 approach. The performance under CWM_H is negative for all but the best performing quintiles, broadly consistent with the results of the other *H* measures.

Panel D of Table 4 presents results for the FM measure. The results show very large values of the FM measures at the high and low quintiles, and also large average alpha effects that vary between -1.44% and +.80% per month across the quintiles. Once again, the average alpha effect largely drives the variation in the measured performance across the FM quintiles. At the fund level, the correlation between FM and its Average Alpha Effect is 0.94. Removing the average alpha effects, the FM_H measures produce with an HML difference of only 0.07% per month. Sorting funds on their FM measures is a lot like sorting them on the their Average Alpha Effects.

Overall, the analyses in Table 4 show that variation in the cross-section of the classical holdings-based performance measures is driven to a large extent by variation in the Average Alpha Effects. When the Average Alpha Effects are removed, the remaining performance is negative for the median fund under all of the measures. These results present a puzzle. If active fund managers are to have positive performance before costs (Berk and Green, 2004) then the Average Alpha Effects should be associated with skilled fund management. However, our simulations show that large Average Alpha Effects can arise in a buy-and-hold strategy. This motivates further analysis of the relation between the Average Effects and active fund management.

5.4 Cross-sections Related to Active Managment

In this section we ask if grouping funds by proxies for active fund management reveals differrences in their Average Alpha Effects. If the Average Alpha Effects are related to active management it tilts us toward the view that the Average Alpha Effect is a valid component of performance. Previous studies suggest several proxies for active fund management, including higher return gap (Kacperczyk, Sialm and Zheng, 2008) lower factor model regression R-square (Amihud and Goyenko, 2013), larger active weight (Doshi, Elkamhi and Simutin, 2015) and higher fund return volatility (Jordan and Riley, 2016).

In Table 5 we sort funds by the different active managment measures, calculate the classical and Haj10 measures and average them across the funds in each group. (The Internet Appendix Table 6A shows the similar results using *DiffIV*.) The sample period for return-gap and active weight is from 1980 to 2012. For R-squared and fund return volatility we start the analyses in 1999 when daily fund return data become available. The estimated measures and their T-statistics are shown in the table. To compute the T-statistics we cluster by time and apply a Newey-West (1987) estimator for panels allowing for nonzero autocovariance up to 30 months.

The bottom row of Table 5 reports the differences between the high and low quintile performance. With four performance measures and four measures of active management, there are 16 HMLs in the bottom rows of Table 5. Using the classical measures, seven of the HMLs sport t-ratios larger than 2.0. This evidence is broadly consistent with the previous studies finding that active management is associated with better performance under the classical measures.

Sorting on return gap in Panel A of Table 5 shows that the differences between the original measures and the Haj10 estimates, approximating the average alpha effects, are fairly similar across the return gap quintiles. The remaining three panels of Table 5 record similar results for the other measures of active management. For the FM sorted by R-squares there might be significant variation, but there is otherwise relatively little variation in the Average Alpha Effects across the quintiles of funds, sorted by R-square, active weight or return volatility. The average alpha effects are large but they vary little across the quintiles, so the quintile spreads for the classical and Haj10 measures are similar. In the Internet Appendix Table 6B we present a similar exercise using a few more sorting criteria. These include the portfolio-weighted size of the fund, the rank gap, a backwards return gap and fund turnover. Other than a weak relation to turnover (consistent with Pastor, Stambaugh and Taylor, 2017) we find little relatiion between the Average Alpha Effects and the measures of active management.

5.5 Predictive Inability

We compare the ability of the various holdings-based performance measures and their embedded Average Alpha Effects to detect funds whose holdings emphasize stocks that will subsequently outperform. For these predictive exercises we estimate the measures for each month using the "up-to-t" version of the Haj10 estimator described in the Appendix, so that the performance measure uses only data available at the forecast date. We average each monthly measure over the past 12 months to reduce noise. We then rank the funds on the averaged measures each month and group them into quintiles. We form equal weighted portfolios of the stocks held by each quintile of funds, and examine the following one-month returns of the stocks before any costs. Rolling this procedure forward in time, we record the subsequent Carhart (1997) four-factor alphas for the five monthly portfolio returns. We also examine the one year ahead and two-year ahead performance of the portfolios, the high-low decile spreads and their t-ratios. We find no evidence that any of the measures can predict the future stocks' abnormal performance.

5.6 Fund Characteristics Associated with Average Alphas

The evidence so far suggests that Average Alpha Effects are not related to active fund management. When the Average Alpha Effect is removed, the remaining performance of the average mutual fund is negative, even before costs. This seems inconsistent with the Berk and Green (2004) view of performance in the fund management industry. This also seems to deepen the puzzle of active management, as raised by Gruber (1996) and others. Why would investors pay for active management that delivers negative excess returns even before fund costs? One potential answer is fund clientelles, as suggested by Glode (2011), Ferson and Lin (2014) and Chretien and Kammoun (2017).

Subsets of investors may disagree with the benchmark risk adjustments used in the measures and reward funds with particular characteristics. Funds that cater to such preferences may be viable even with negative performance measured relative to benchmark risks. If the Average Alpha Effects are associated with fund characteristics that investors have a preference for, the the classical measures by including the Average Alpha Effects, could better reflect investors' preferences for funds. We examine the relation between Average Alpha Effects and various fund characteristics that investors may have preferences for.

Table 6 presents panel regressions of the Average Alpha Effects in the classical performance measures on lagged fund characteristics. The simulation results of Table 3 show that buy and hold and momentum strategies can have large Average Alpha Effects, so we include as controls measures of funds' tendencies toward buy-and-hold and momentum strategies. TBH is our measure of the tendency towards buy-and-hold:

$$TBH_t = 1 - (1/24)\Sigma_{s=0,\ldots,11} |\mathbf{w}_{t-s} - \mathbf{w}^{bh}_{t-s}|, \tag{12}$$

where $\mathbf{w}^{bh}_t$ is what the fund's weights would have been at time *t*, had the fund had the same share holdings as 12 months before. The monthly measure is averaged over the past year to reduce noise. It takes a value between zero and one, and is equal to one if the fund exactly follows the buy-and-hold strategy.

Fund investors might have a preference for momentum or contrarian trading styles, and the simulations of Table 3 show that a momentum strategy produces a positive Average Alpha Effect. We include a lagged momentum measure similar to Grinblatt, Titman and Wermers (1995):

$$LM_t = \Sigma_i (\mathbf{w}_t - \mathbf{w}^{bh}{}_t)'(\Sigma_{s=t-12,\ldots,t-2}\, \mathbf{r}_s). \qquad (13)$$

This measures a relation between a fund's deviation from buy-and-hold weights and the average returns on the stocks over the previous 11 months, after skipping one month. A positive LM means that the fund pursues a momentum trading strategy and a negative value indicates a "contrarian" strategy.

Some investors might have preferences for funds with high dividend yields (e.g. Harris, Hartzmark and Solomon, 2017). The variable Div Yield is a fund-portfolio-weighted average of the dividends per share of each stock held during the past 12 months divided by its price per share. Some investors may prefer a tilt towards large or small-capitalization stocks that is not perfectly captured by the factor loadings in the model. HLDSize is the portfolio weighted market capitalizations of the stocks held by the fund. Investors might have preferences for fund characteristics associated with expense ratios, annual turnover, age or fund size. We include these characteristics in the regressions. We also include the dummy variable *Aggressive*, which turns on if the fund is classified as an Aggressive Growth fund. All of these fund characteristics are lagged one month relative to the performance measures used to construct the Average Alpha Effects. The Average Alpha Effects are estimated monthly as the difference between a classical measure, estimated as in Table 2, and its up-to-t version of the Haj10 estimate.

Panel A of Table 6 covers the 1980-1994 period and panel B the full 1980-2012 sample. The table shows, as the simulations foreshadowed, a positive association between Average Alpha

Effects and a fund's tendency toward buy-and-hold as captured in TBH, for all of the measures.

To some surprise, the panel regressions indicate that the Average Alpha Effects are not significantly correlated with a tendency for momentum trading as captured by LM, controlling for the other variables, excepting in the GT measure during the full sample, where the coefficient is negative. In the DGTW measure the coefficient on LM is negative but insignificant.

In the GT measure the Average Alpha Effect refers to average weight changes, and the Average Alpha Effects are thus smaller than in the FM and DGTW measures. The Average Alpha Effect in the GT measure presents no significant relation to TBH or any of the other variables, excepting Div Yield. (We examine the robustness of these results to the use of Fama MacBeth (1973) regressions in the Internet Appendix and find similar results.)

The Average Alpha Effect is positively associated with the dividend yields of the stocks held by a fund in all of the measures, and is significantly related to yields for all of the measures in the full sample period in Panel B. A positive relation between the Average Alpha Effect and Dividend Yield could arise if the dividend yields of the stocks are correlated with their average alphas in the DGTW and Carhart factor models. This would be implied by the Brennan (1970) after-tax CAPM, which predicts a positive premium associated with a stock's dividend yield. We find correlations across the stocks in our sample, between their average alphas and their dividend yields, of 0.074 in the DGTW model and 0.258 in the Carhart model. If high yield funds are doing dividend capture trades, and if ex-dividend price declines are less than the amount of the dividend (Elton and Gruber, 1970) then high yield funds could have larger average alphas before trading costs.

As a further check on the Div Yld we estimate a dividend yield adjusted excess return for each stock and run Table 6 again using the adjusted returns. The adjusted returns contain an adjustment for a premium associated with the dividend yield. These are formed by running Fama and

MacBeth (1973) style regressions of stock returns on rolling estimates of their market betas and their dividend yields. The cross-sectional regression coefficient each month, multiplied by the stock's dividend yield, is subtracted from the adjusted excess return.

Using the dividend yield adjusted excess returns in place of the original excess returns we find that the relation between the Average Alpha Effect and dividend yield remains marginally significant over the full sample period, with a coefficient about 2/3 of the previous size, but it flips sign in the DGTW model. In the earlier subperiod Div Yld remains strongly significant for the Average Alpha Effect in the GT measure but becomes insignificant in FM and flips sign in DGTW.

To summarize, Table 6 shows that funds with a greater tendency toward buy-and-hold strategies have larger Average Alpha Effects. Controlling for the TBH effect, there is no relation between funds' Average Alpha Effects and their tendency toward momentum strategies. We find no strong relations between Average Alpha Effects and the other fund characteristics, other than dividend yields. Differences in Average Alphas across funds are related to their dividend yields, so if investors have a preference for yield this would show up in the Average Alpha Effects in the classical measures. The dividend yield relation to Average Alpha is reduced but not eliminated by including a premium for dividend yield in the model.

5.7 Fund Flows and Average Alphas

Even if the Average Alpha Effect is a passive component of performance, that does not imply that fund investors would pay no attention to it. Studies by Del Guercio and Tkac (2002), Berk and van Binsbergen (2015) and Barber, Huang and Odean (2016) find that investor flows respond to relatively simple measures of abnormal fund performance. If fund investors reward larger Average Alpha Effect with greater fund flows, then fund managers would be induced to produce large Average Alpha Effects. Table 7 examines this conjecture. Annual new money flows are regressed on the Average Alpha Effects in the performance measures over the past year. The Average Alpha Effects are estimated as in Table 2 and are used to form ranked performance measures within each of five performance quintiles, as in Sirri and Tufano (1998). In panel A there are no controls for fund characteristics, and in panel B fund characteristics are included as control variables. The sample period is 1980-2012. T-ratios are calculated by clustering by time and using Newey-West (1987)-weighted autocovariances to 30 lags. The units of the Average Alpha Effects are annual percent.

Panel A of Table 7 finds a strong relation of flows over the next year to funds' Average Alpha Effects over the previous year, consistent with the idea that funds may be motivated by investor flows to generate large average alpha effects. While the results vary some across the models, the coefficients for the performance quintiles positive and significant coefficients for the high group and smaller, typically negative signs for the lower quintiles. The t-ratio on performance measured by the Average Alpha Effect is 3.68 for the FM measure and 7.25 for the DGTW measure in the top performance quintile.[11]

Panel B of Table 7 includes the fund characteristics from Table 6 as additional control variables in the flow performance relation. None of the performance rankings based on the Average Alpha Effects produce significant coefficients, indicating that the flow response to the Average Alpha Effects in Panel A is subsumed by the other fund characteristics. Style flow and turnover are strong flow predictors, and some of the other characteristics like fund Age, TNA,

---

[11] We also examine regressions where the Average Alpha Effect is measured over the previous three years, which produces similar but less statstically significant results. We also examine regressions where the classical measures are introduced with the Average Alpha Effects in the regression and find that the original measures subsume the Average Alpha Effects.

market capitalization and the BHL and LM measures are significant or marginally significant. We conclude that the response of flows to the Average Alpha Effects is subsumed by the other fund characteristics.

6.8 Robustness

Table 4 uses monthly stock returns and monthly weights constructed from the quarterly holdings data as described above. These features raise potential concerns about the robustness of the results.

The first concern is errors or microstructure effects in the closing prices used to compute the weights and the returns. At the beginning of each month the same price appears in the weight and in the denominator of the future return calculation, so an error in the price might induce a spurious negative relation between the lagged weight and the future return. In momentum studies (e.g. Grinblatt and Titman, 1993) it is common to skip a day between the formation period and the future return calculation interval to handle such concerns. In a similar spirit we replicate Table 5 using 29-day returns on the left hand side of the regressions, skipping a day relative to the price data in the weights.

Table 5B in the Internet Appendix presents results using the 29-day returns. The holdings-based measures are not larger than in the original Table 4, in constrast to what would be expected if errors in prices caused a negative bias in the original results. The DGTW measures are actually smaller using the 29-day returns, but otherwise the results are very similar.

A second potential concern about the fund level results is the implicit buy-and-hold assumption that is applied between the quarterly reporting dates. We saw in Table 3 that a simulated buy-and-hold strategy can generate a positive average alpha effect. The concern here is that the large average alpha effects of the actual funds are partly an artifact of the buy-and-hold assumption.

Table 5C in the Internet Appendix replicates Table 4 using quarterly data for the returns and portfolio weights, this avoiding the buy-and-hold assumption. The results for the GT measure show slightly more extreme average alpha effects in both the high and low quintiles, as would be expected given the higher estimation error of the quarterly measures. However, the differences with the Table 5 results are only .01% or .02% per month. The average alpha effect under the GT measure for the W5000 fund is more negative in the quarterly data. It changes from -0.06% to -0.13% per month under GT. The W5000 average alpha effect is also slightly smaller at +0.06% per month (versus 0.11%) under DGTW in the quarterly data. The DGTW measures display slightly higher average alpha effects across the quintiles in the quarterly data (about 0.01% to 0.03% higher), in contrast to what would be expected if there was a positive bias in the monthly results from the buy-and-hold assumption between quarterly reporting dates.

Tables 6A and 6B in the Internet Appendix present furhter robustness checks on the analysis of Table 5 where funds are grouped according to proxies for active management. We sort on portfolio-weighted size, rank gap, backwards return gap and turnover.

7. Conclusions

We introduce a predictive panel regression framework for the estimation of holdings-based measures of portfolio performance. Previous measures appear as special cases of a panel regression. Estimating the measures from a panel regression makes the tools of panel econometrics available to holdings-based measures. This perspective reveals that the classical holdings-based performance measures can suffer from a panel version of a lagged stochastic regressor bias similar to Stambaugh (1999). We examine several bias-adjusted measures with simulation and find two that work well.

Our analysis reveals an Average Alpha Effect in the classical performance measures and shows how to isolate it using a regression with stock fixed effects. The Average Alpha Effect is a cross-sectional relation between the time-averaged portfolio weights held by a fund and the average stock alphas in the model. It arises because the average alphas of the stocks are not zero. Funds with positive Average Alpha Effects put more weight, on average, on the higher alpha stocks.

The Average Alpha Effect is interesting because it dominates the variation in the classical performance measures across funds. It appears larger in more recent fund data, where previous studies find more passive management and closet indexing. The Effect is stronger in funds with a stronger tendency towards a buy-and-hold strategy and for funds with higher dividend yields. Average Alpha Effects are large in simulated buy-and-hold and momentum strategies, and also appear in passive index funds.

We find no strong relation between the Average Alpha Effect and well-known proxies for funds' active management, including factor model R-squares (Amihud and Goyenko, 2013), fund volatility (Jordan and Riley, 2016), the active weight measure of Doshi, Elkamhi and Simutin (2015) the return gap of Kacperczyk, Sialm and Zheng, (2008) or other measures. We find no information about future stock returns in the holdings of funds are sorted by their Average Alpha Effects.

Our analysis leads to a different inference about mutual fund performance in holdings-based measures. Previous studies find positive performance on a before-cost basis, of a magnitude similar to funds' expense ratios, while the after-cost abnormal returns for investors is close to zero. This conforms to a view of the mutual fund industry, advocated by Berk and Green (2004), where fund managers with skills at active management have significant before-cost performance, but leave no abnormal returns for investors after costs. Assuming that the Average Alpha Effect is either a passive component of performance or a bias, it is interesting to remove it from the classical performance measures. We find that when it is removed the remaining before-cost performance of the average mutual fund is negative. This casts doubt on the presence of skilled active management in the average mutual fund, even on a before-cost basis.

Appendix

A.1 Extracting the Average Alpha Effect

The difference between a classical holdings-based performance estimator and an estimate of a fixed-effects model can serve as an empirical measure of the Average Alpha Effect. Suppose that we estimate the regression (4) when the true model has stock fixed effects as in (7). The OLS estimator in (5) can then be written as:

$$\hat{\beta} = \beta + \Sigma_i\Sigma_t\, a^i w^i_t / \Sigma_i\Sigma_t\, {w^i_t}^2 + \Sigma_i\Sigma_t\, w^i_t\, \varepsilon^i_t / \Sigma_i\Sigma_t\, {w^i_t}^2, \qquad \text{(A.1)}$$

where β is the true value of the slope in (7) and the ε^i_t are the residuals of (7). Hjalmarsson's (2010) evidence and Panel B of our Table A.1 below suggest that the sample mean of the right-most term involving the residuals is close to zero. The fixed effects produce the middle term on the right hand side of (A.1). This term is evaluated using a factor model regression for the stocks on the vector of benchmark excess returns $\mathbf{r}_B$:

$$r^i_{t+1} = \alpha_i + \boldsymbol{\beta}_i{}'\mathbf{r}_{Bt+1} + u_{it+1}. \qquad \text{(A.2)}$$

This regression is not very restrictive, allowing for a stock-specific intercepts or alphas.[12] Note that the DGTW benchmark is a special case of the term $\boldsymbol{\beta}_i{}'\mathbf{r}_{Bt+1}$, where there are 125 benchmarks and the elements in $\boldsymbol{\beta}_i$ are either zero or 1.0. We make the standard regression assumption, $E(u_{it+1}) = 0$ and $E(\mathbf{r}_{Bt+1}\, u_{it+1}) = \mathbf{0}$.

---

[12] Regression (A.2) is explicitly an unconditional regression, and its specification does not in principal preclude the existence of a conditional model with a time-varying alpha. The interpretation is similar in this case.

Substituting $a^i = \alpha_i + \boldsymbol{\beta}_i{}' \overline{\mathbf{r}_B} - \beta \overline{w}^i$. The expected values of the first two of the three terms in (A.1) that result are the average alpha effect, $\Sigma_i\ \alpha_i \overline{w}^i$, and the average style effect, $(\Sigma_i \overline{w}^i \boldsymbol{\beta}_i)' \overline{\mathbf{r}_B}$. Denote those as [AA+AS], the decomposition implies:

$$E\hat{\beta} = [AA+AS] + \beta\,[1 - (1/T)\Sigma_i\Sigma_t \overline{w}^i\, w_{it} /(1/T)\Sigma_i\Sigma_t\, w_{it}^2] \tag{A.3}$$

$$= [AA+AS] + \beta\,[1 - (1/T)^2\Sigma_i\{\Sigma_t\, w_{it}^2 + \Sigma_\tau\Sigma_{t\neq\tau}\, w_{it}\, w_{i\tau}\}/(1/T)\Sigma_i\Sigma_t\, w_{it}^2],$$

$$= [AA+AS] + \beta\,[1 - (1/T)(1+2\rho/(1-\rho)],$$

where the last line uses an AR(1) approximation for the weights. If the final term in (A.3) is close to β it justifies using the original estimator of (4) minus an unbiased estimator for the model with fixed effects to proxy for [AA+AS]. As the AS part is small in the GT measure, and is equal to zero as in the other measures, this is approximately the Average Alpha Effect. As T gets large the approximation is exact, but it is likely to be close in realistic finite samples. If ρ = 0.9 and T = 200, the term multiplying β on the right-hand side of (A.3) is about 0.905.

A.2 Corrections for Bias

Hjalmarsson (2008) proposes a parametric correction for the Stambagh bias in a panel regression. He assumes a local-to-unity structure for the autoregressive parameter for the portfolio weights, $\rho = 1 - c/T$. He also makes the strong assumption that the errors in (7) are independent across stocks. (We relax this assumption in our simulations.) He uses sequential limit theory (first, fix N and let T go to infinity, then let N go to infinity) to obtain several results.

First, perhaps surprisingly, when there are no fixed effects in the model there is no Stambaugh bias and the pooled OLS estimator of Equation (4) is consistent.

Hjalmarsson (2008) proposes a bias-corrected estimator for the fixed effects case. Letting $R^i_{t+1} = r^i_{t+1} - \overline{r}^i$ and $W^i_t = w^i_t - \overline{w}^i$, the corrected estimator is:

$$\hat{\beta}_c = \Sigma_i \Sigma_t (R^i_{t+1} W^i_t - NT\, Cov(\varepsilon,\nu)\theta(c)) / \Sigma_i \Sigma_t (W^i_t W^i_{t'}), \tag{A.4}$$

where $Cov(\varepsilon,\nu) = (1/N)\, \Sigma_i\, Cov(\varepsilon_{it},\nu_{it})$ is consistently estimated from the OLS residuals of the panel predictive regression system. Hjalmarsson finds that the estimator is relatively insensitive to how the residuals are estimated. The crucial parameter is c in $\theta(c) = -(e^c - c - 1)/c^2$. He proposes to estimate the parameter c as $T(1-\rho_{pool})$, where ρ_{pool} is the pooled estimator of ρ with no intercept in the regression. Equation (A.4) is the first bias corrected estimator that we evaluate.

Hjalmarsson (2010) notes that with a common intercept in the model, the pooled OLS estimator is consistent and asymptotically normal. We evaluate this estimator with simulations. He also proposes a forward recursively-demeaned estimator for the fixed effects case. Let $w^{bd}_{it} = w^i_t - [1/(t-\tau)] \sum_{s=1}^{t-\tau} w^i_s$, $w^{fd}_{it} = w^i_t - [1/(T-t+\tau)] \sum_{s=t+\tau}^{T} w^i_s$, and $r^{fd}_{it} = r^i_t - [1/(T-t+\tau)] \sum_{s=t+\tau}^{T} r^i_s$ (We use τ=12 months as the gap between the current and mean value) be the backward and forward demeaned weights, and the forward demeaned returns. The estimator is:

$$\hat{\beta}_H = \Sigma_i \Sigma_t (r^{fd}_{it+1} w^{bd}_{it}) / \Sigma_i \Sigma_t (w^{bd}_{it} w^{fd}_{it}). \tag{A.5}$$

Note that the backward demaning weight, w^{bd}_{it} , is used as an instrument in (A.5). The intuition is that the forward demeaned returns are independent of the lagged weights. The recursive demeaning induces a moving average term in the covariance matrix of the errors of the demeaned model, which should be reflected in the panel equivalent of HAC standard errors. We use this approach.

In our empirical work we investigate the ability of the various performance measures to identify funds that can predict future stock returns. The future demeaning in the $\hat{\beta}_H$ measure cannot be used for this purpose. We therefore modifiy the approach so that the measure used to predict stock returns at month *t+1* uses only data for time-*t* and before. The "up-to-t" version replaces w^d_{it} with w_{it1} - $[1/(t1-\tau)]\sum_{s=1}^{t1-\tau} w^i_s$ and $r^d_{it} = r_{it1+1} - [1/(t-t1-\tau)]\sum_{s=t1+\tau+1}^{t} r^i_s$ for $t1 < t - \tau$. We use τ=12 months as the gap between the current and mean value used to demean in all of our recursively demeaned estimators. When t1=t-13 the future demeaned return is just r_t, and the weight used to predict the return for time t+1 is based on holding data at time t-13. For t1=t-24 we demean using a one-year future mean return or weight, using data up to time t only. The weight used to predict the return for time t+1 is based on holding data at time t-24.

When applied to the CWM, the estimators need to accommodate the estimation of the regression coefficients, **δ**, of the returns on the lagged information $\mathbf{Z}_t$. For example, we modify the Diff IV approach for the panel version of the CWM, replacing $(r^i_{t+1} - r^i_{t+1+\tau})$ in (21) with $((r^i_{t+1} - \boldsymbol{\delta}_t'\mathbf{Z}_t) - (r^i_{t+1+\tau} - \boldsymbol{\delta}_{t+\tau}'\mathbf{Z}_{t+\tau}))$, where the estimator of δ is $\boldsymbol{\delta}_t = (\mathbf{Z}_t^{T-1}\, \mathbf{Z}_t^{T-1})^{-1}\mathbf{Z}_t^{T-1'}\mathbf{r}_{t+1}^{T}$, with $r_{t+1}^T = (r_{t+\tau+1}, \ldots, r_T)$ and $Z_t^{T-1} = (Z_{t+\tau}, \ldots, Z_{T-1})$. We cancel the common intercepts from the difference between the two regression estimates. In the FM estimator we have to estimate the parameters of the SDF, [a;**b**]. We use monthly data on the benchmark factors to estimate these parameters, and evaluate the resulting two-step estimators by simulation.

A.3 Difference Estimation

Take the difference between equation (7) and the same equation τ periods before, and the fixed effects cancel out:

$$r_{t+1}^i - r_{t+1-\tau}^i = \beta(w_t^i - w_{t-\tau}^i) + \varepsilon_{t+1}^i - \varepsilon_{t+1-\tau}^i. \tag{A.6}$$

The classical difference estimator is

$$\hat{\beta}^{diff} = \frac{\sum_t \sum_i ((w_t^i - w_{t-\tau}^i)(r_{t+1}^i - r_{t+1-\tau}^i))}{\sum_t \sum_i ((w_t^i - w_{t-\tau}^i)^2)}. \tag{A.7}$$

We show that the classical difference estimator in Equation (A.7) suffers a Stambaugh bias for $\tau \geq 2$. If we plug $r_{t+1}^i = \alpha^i + \beta w_t^i + \varepsilon_{t+1}^i$ into the numerator of the difference estimator, equation (A.7) is equal to β plus a term whose numerator is:

$$\sum_t \sum_i ((w_t^i - w_{t-\tau}^i)(\varepsilon_{t+1}^i - \varepsilon_{t+1-\tau}^i)). \tag{A.8}$$

Using an AR(1) assumption for the weights,

$$w_t^i = \gamma^i (1 + \sum_{j=1}^{\tau-1} \rho^j) + w_{t-\tau}^i \rho^\tau + \sum_{j=1}^{\tau} \rho^j v_{t-j}^i. \tag{A.9}$$

and equation (A.8) becomes:

$$(w_{t-\tau}^i (1-\rho^\tau) - \gamma^i (1 + \sum_{j=1}^{\tau-1} \rho^j) - \sum_{j=1}^{\tau} \rho^j v_{t-j}^i)(\varepsilon_{t+1}^i - \varepsilon_{t+1-\tau}^i). \tag{A.10}$$

Under the null hypothesis of no ability, the past weights and future regression residuals are uncorrelated. However, if the innovations in the weights $v_{t+1-\tau}^i$ and the return innovations $\varepsilon_{t+1-\tau}^i$ are contemporaneously correlated, as seems highly likely, and if ρ is nonzero, the expected value of $(-\sum_{j=1}^{\tau} \rho^j v_{t-j}^i)(\varepsilon_{t+1}^i - \varepsilon_{t+1-\tau}^i)$ is not zero when $\tau \geq 2$. Thus, there is a Stambaugh bias in the classical difference estimator when τ≥2. We evaluate this estimator for τ=1 in our simulations.

Finally, we evaluate a *differenced IV* estimator following Anderson and Hsiao (1981) and Wang (2015). This estimator uses the lagged weight difference as an instrumental variable, forward differences the returns and weights:

$$\hat{\beta}^{DiffIV} = \frac{\sum_t \sum_i ((w_t^i - w_{t-\tau}^i)(r_{t+1}^i - r_{t+1+\tau}^i))}{\sum_t \sum_i ((w_t^i - w_{t-\tau}^i)(w_t^i - w_{t+\tau}^i))}. \qquad \text{(A.11)}$$

We show that the differenced IV estimator of Equation (A.11) is consistent. Plug in $r_{t+1}^i = \alpha^i + \beta w_t^i + \varepsilon_{t+1}^i$, and the numerator of Equation (A.11) can be written as

$$\sum_t \sum_i (w_t^i - w_{t-\tau}^i)(\beta(w_t^i - w_{t+\tau}^i) + (\varepsilon_{t+1}^i - \varepsilon_{t+1+\tau}^i)). \qquad \text{(A.12)}$$

By assumption, $\varepsilon_{t+1}^i - \varepsilon_{t+1+\tau}^i$ is not correlated with $w_t^i - w_{t-\tau}^i$ for τ≥1. Therefore, equation (A.12) converges, for large T, to $\sum_i \beta E((w_t^i - w_{t-\tau}^i)(w_t^i - w_{t+\tau}^i))$. Using the AR(1) assumption for the weights we find:

$$E((w_t^i - w_{t-\tau}^i)(w_t^i - w_{t+\tau}^i)) = (1-\rho^\tau)^2 Var(w_t^i). \qquad \text{(A.13)}$$

The last equation depends on the stationarity assumption: $cov(w_t^i, w_t^i) = cov(w_{t-\tau}^i, w_{t-\tau}^i)$. With this result, A.4 converges to $\sum_i \beta(1-\rho^\tau)^2 Var(w_t^i)$. By similar logic, the denominator of equation (18) converges to $\sum_i (1-\rho^\tau)^2 Var(w_t^i)$. Therefore, when number of stocks N is large, the differenced IV estimator converges to β.

A.4 Evaluating Lagged Stochastic Regressor Bias Adjustments

We examine the effectiveness of the adjustments for Stambaugh bias using simulations. We bootstrap versions of the predictive system of equations (7) and (10). We report results for various "true" values of the panel regression slope, β. To calibrate the regression slopes we estimate the pooled panel model for β using the differenced IV method (which turns out to have little bias) and sort funds based on the estimated slopes. We select five funds near the 5%, 10%, 50%, 90% and 95% cutoff values, and use the holdings data of the five selected funds to calibrate the simulations. The differenced IV β is taken to be the true value in calibrating these simulations. For a given value of the slope β, the intercept for each stock in Equation (7) is chosen to match the sample average benchmark adjusted returns. Thus, the data generating process features heterogeneous fixed effects across the stocks. The parameter ρ for the simulated weights is also allowed to differ across the stocks in these experiments. For a given value of ρ the intercepts in Equation (10) are fixed to match the average values of the actual weights for each stock.

We boostrap the vector of residuals from equations (7) and (10) for all of the stocks together with the weight residual vector for a given month, selecting months randomly from the data, with replacement. We build up the simulated weight and return series recursively, using the calibrated values of β, ρ and the intercepts. This preserves the serial dependence of the weights, the dependence across the stock returns, and the dependence between the returns and weight innovations, which are important for the Stambaugh bias. The bias-adjusted estimator of Hjalmarsson (2008) assumes cross-sectional independence, so we wish to evaluate its performance in the face of dependence across stocks. The number of months for each fund in the simulations is the same as the number of months where the fund exists in the real data. When a stock held by the fund has a missing return, the weight is set to zero.

Panel A in Table A.1 presents the results of the simulations to address bias in the estimators.

The true values of the slopes are shown in the first column and the estimated values, averaged across 1000 simulation trials, are shown in the other columns. The regressions are scaled, where the dependent variable is the return multiplied by the variance of the weight, so as to deliver the numerator of the slope coefficient, measured as an excess return in percent per month. The first column shows that the median "true" measure in the simulations is small, at 0.06% per quarter, but the range of β values in the experiments cover -0.3% to almost 0.6% per quarter.

The second column of Panel A shows that the OLS estimator of Equation (4) is expected to be much larger than the slope if the model has fixed effects. When the true intercept is positive and the regression supresses the intercept, the estimated slope coefficient gets larger to compensate. The OLS estimator with a common intercept (column 3) is also larger than the true coefficient with fixed effects, but to a lesser extent. Including a single intercept is a step toward estimating fixed effects, but is not an adequate solution.

Column 4 of Table A.1, Panel A, shows results for the bias adjusted estimator from Hjalmarsson (2008). This estimator has a small upward bias. This estimator ignores the dependence across stocks in the regression residuals, which our simulations capture.

Columns 5 and 6 of Table A.1, Panel A, present results for the standard least squares dummy variable estimator ("Within") and the classical difference estimator for the model with fixed effects ("Diff"). These estimators suffer from the Stambaugh bias and the bias for the slope is negative as expected. According to Equation (10) this is because the average correlation between the stock return residuals and the future weights is positive. Thus, without finite sample bias adjustment the standard panel regression estimators with fixed effects would likely be too pessimistic about fund performance.

Columns 7 and 8 of Table A.1, Panel A, present results for the differenced IV estimator (Diff

IV) and the Hjalmarsson (2010) estimator (Haj10). The differenced IV estimator performs pretty well in terms of bias. The largest bias across the five experiments is only 0.03% per month. For Haj10 the largest bias is only 0.01% per month. Thus, the two estimators that perform the best at removing the Stambaugh bias are differenced or forward demeaned instrumental variables estimators. We therefore concentrate on using these approaches to estimate the models in the sequel.

Panel B in Table A.1 presents simulations which show that the original holdings-based measures are close to unbiased when there are no fixed effects in the data generating model. The simulation procedure is the same as before, replacing the stock returns with either the DGTW adjusted returns, the conditional mean adjusted returns, where the conditional mean is $\boldsymbol{\delta}'\mathbf{Z}_t$, or the FM benchmark adjusted return, $m_{t+1}r^i_{t+1}$. The simulations for the FM measure incorporate the variability from estimating the $\boldsymbol{\delta}$ in conditional model and the SDF parameters (a,**b**), by estimating them in each simulation trial.[13] The key step for the simulations in Panel B is to set the firm fixed effects equal to zero in Equation (7) in the simulated returns. It is clear that each of the four measures in Panel B are close to their corresponding true values when there are no fixed effects in the data. This is consistent with the findings of Hjalmarsson (2008, 2010)[14]. Of course, as Panel A shows, in the more realistic case where the data have fixed effects, it is necessary to use a bias adjusted estimator such as Haj10 or DiffIV.

---

[13] To incorporate estimation error in $\boldsymbol{\delta}$, in each trial, we simulate stock returns by summing conditional mean adjusted returns and $\boldsymbol{\delta}'\mathbf{Z}_t$, where $\boldsymbol{\delta}$ is the estimated paremter from the real data, and $\mathbf{Z}_t$ is the simulated conditioning information based on an AR(1) model also estimated using the real data. We then apply the method described in section 3.7 to construct difference IV or Haj10 measure. To incorporate the estimation error in the parameters a and **b** we estimate them in each simulation trial based on simulated factors that we draw with each vector of return residuals. The estimated values of a and **b** from the data serve as the true values to calibrate the simulations, which produces $m_{t+1,}$ the SDF constructed using the true a and **b** and the original data. We resample the residuals from regressing the SDF adjusted returns, $m_{t+1}r^i_{t+1,}$ on the weights. Using these coefficients and simulated residuals, we construct the SDF adjusted returns in each simulation trial. Final, we multiply the SDF adjusted return in each simulation trial by a factor $m^{\wedge}_{t+1} / m^{*}_{t+1,}$ where $m^{\wedge}_{t+1}$ is the SDF constructed using the estimated a and **b** with the simulated factors, and m^{*}_{t+1} is the SDF constructed using the true a and **b** with the simulated factors.

[14] The original CWM suffers a lag stochastic bias from the estimation of $\boldsymbol{\delta}$; thus, we estimate the corresponding Haj10 measure which will not suffer this issue. We also simulate diff IV mesures and obtain the similar results.

A.5 Simulating Hypothetical Fund Strategies

This section describes simulations of hypothetical funds undertaking buy-and-hold and momentum trading strategies. In the real data, most stocks have returns that exist for a number of consecutive months, and then disappear. We wish to replicate this feature. Some short-lived stocks have extreme sample average returns and alphas with respect to the various models. If these stocks were included for more periods in the simulation than in the actual data, it will lead to inaccuracies in the simulations. Thus, we strive to keep the number of months during which a stock exists the same in the simulations as in the real data. The cross-sectional covariance among different stocks should also be preserved in the simulations.

We bootstrap the stock returns with the following method. Consider the DGTW model as an example. The bootstrap is based on $r_t^i = r_t^{D^i,t-1} + (r_t^i - r_t^{D^i,t-1})$, where $r_t^i - r_t^{D^i,t-1}$ are the DGTW benchmark adjusted returns, and the two parts of the returns are bootstrapped independently. (The approach using the other benchmarks is similar.) The first step is to construct a pool of 125 DGTW benchmark returns, $r_t^{D^i,t-1}$. We also create another pool of the DGTW benchmark adjusted returns, $r_t^i - r_t^{D^i,t-1}$, covering only those periods during which the stock return r^i_t exists. In each simulation draw we select all the benchmark returns at a randomly selected time point from the benchmark pool. Note that we draw all 125 DGTW benchmark returns (or all the benchmark excess returns, r_B), thus preserving the cross-sectional covariance between the benchmark returns. For each stock that exists at this time point in calendar time in the real data, we independently bootstrap the corresponding benchmark adjusted return from the pool of adjusted returns for that stock, picking a separate time period at random for the adjusted return.

This is consistent with a regression residual being independent of the regressors in the factor model. If a stock does not exist on the (calendar) date, the adjusted return for that stock is set as missing. This assures that the stock returns exist during the same periods in the simulation as in the real data. The bootstrapped stock return for its nonmissing dates is the summation of its bootstrapped benchmark return and the benchmark-adjusted return. This procedure captures the correlations of the stocks only through their benchmarks. The benchmark-adjusted return captures the alpha of the stock relative to the benchmark, which enables us to examine the Average Alpha Effects with realistic stock alphas.

Under the null hypothesis of no performance, the holdings are uninformed about future stock returns. Our simulated buy-and-hold strategies have this feature. The buy-and-hold weights contain only information about past stock returns, and so are uninformed under weak-form informational efficiency (Fama, 1970). The momentum strategy weights use similar information, except they also "know" about the momentum effect. To make the strategies somewhat realistic we keep the weights between 0 and 1, and we keep the number of stocks in the portfolios similar to those in the real data.

A.6 Simulated Buy-and-hold strategy weights

The weights of the buy-and-hold strategy are simulated as follows: In each trial, we randomly select 1000 stocks and assume that the fund chooses the stocks from this pool. At time $t = 120$ (which corresponds to a fund that starts in 1984 in our simulation), the fund selects stocks and equally weights them. If a stock does not exist, the fund will not hold it. On average, there are 150 stocks in the starting portfolio (150 stocks exist in the pool at t=120); thus obtaining a similar number as the number of stocks held by funds in the real data. Next, for each time point

$t > 120$, some new stocks start to exist and some old stocks disappear. For the new stocks in the pool, the manager will hold the stocks, and the weights are random numbers between 0% and 5%. For stocks in the pool that disappear, the fund will not hold these stocks.[15] For stocks that exist in the previous period and do not disappear, the fund continues to hold them without changing the holdings (the weights are changing because prices fluctuate). The simulated weights thus contain only past price information. The number of stocks in each subsequent portfolio is between 150 and 200.

The Investment Company Act of 1940 requires that the weight of each stock cannot be larger than 5% without triggering reporting requirements. As the simulation evolves there are multiple stocks with weights that could exceed 5%, so we use the following procedure to adjust the weights. We first rank the stocks by their weights in the portfolio. Then we select the stock with the highest weight and denote this weight by w_1 (by assumption, $w_1 > 5\%$). Next, we decrease the weight of this stock to 5%, and let the weights of lower-ranked stocks increase by $(w_1 - 5\%)/(N-1)$, where N is the number of stocks in the portfolio. Next, we select the stock with the second highest weight (denote this weight by w_2). If $w_2 > 5\%$, we decrease the weight of this stock to 5% and let the weights of lower-ranked stocks increase by $(w_2 - 5\%)/(N-2)$ (there are $N-2$ stocks with lower rank). We repeat this process until all the stocks have weights less than or equal to 5%. The only exception is the rare case when there are less than 20 stocks in the portfolio. In this case, we do not reduce the weights to 5%.

A.7 Simulated Momentum strategy weights

Similar to the buy-and-hold strategy, the momentum strategy starts off holding all the

[15] The delisting returns are included in the return data to avoid selection bias.

stocks in the pool that have non-missing returns at time *t*=120. After the first month, the fund manager rebalances her portfolio. She ranks the stocks by average returns from the preceding 2 to 12 months, and removes 6% of the stocks with the lowest past returns. This is to match the average monthly turnover in the data of 6% per month[16]. The stocks that have been removed are replaced by the stocks with the largest average past returns. The stocks added are weighted in proportion to their average past returns. On average, the momentum portfolios also contain 150 to 200 stocks.

[16] We would expect that the momentum manager may have a higher turnover ratio than an average fund manager. Therefore, we also simulate the momentum strategy with a 10% or 20% turnover ratio.

References

Amihud, Yakov and Rusian Goyenko, 2013, Mutual Fund R^2 as Predictor of Performance, *Review of Financial Studies* 26, 2013, 667-695.

Amihud, Yakov and Clifford Hurvich, 2004, Predictive Regressions: A Reduced-Bias Estimation Method, *Journal of Financial and Quantitative Analysis* 39, 813-841.

Amihud, Yakov, Cliffird Hurvich and Yi Wang, 2008, Multiple-Predictor Regressions: Hypothesis Testing, *Review of Financial Studies* 22, 414-434.

Amihud, Yakov, Clifford Hurvich and Yi Wang, 2010, Predictive Regression with Order-p Autoregressive Predictors, *Journal of Empirical Finance* 17, 513-525.

Anderson, T.W and Cheng Hsiao, 1981, Estimation of dynamic models with error components, *Journal of the American Statistical Association*, 589-606.

Barber, Bradford, X. Huang and Terry O'Dean, 2016, Which risk factors matter to investors? Evidence from mutual fund flows, *Review of Financial Studies* 29, 2600-2642.

Berk, Jonathan and Richard Green, 2004, Mutual fund flows and performance in rational markets. *Journal of political economy* 112, 1269-1295.

Berk, Jonathan and Jules vanBinsbergen, 2016, Assessing asset pricing models using revealed preference, *Journal of Financial Economics* 119, 1-23.

Brennan, Michael J., 1970, Taxes, market valuation and corporate financial policy. *National tax journal* 23, 417-427.

Busse, Jeffrey and Qing Tong, 2012, Mutual fund industry selection and persistence, Review of Asset Pricing Studies 2, 245-272.

Carhart, Mark, 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57-82.

Chen, Nai-Fu, Bruce Grundy, and Robert F. Stambaugh. "Changing risk, changing risk premiums, and dividend yield effects." *Journal of Business* (1990): S51-S70.

Chan, Lous, Stephen Dimmock and Josef Lakonishok, 2009, Benchmarking money manager performance: Issues and evidence, *Review of Financial Studies* 22-11, 4553-4599.

Chretien, S. and M. Kammoun, 2017, Mutual fund performance and best clienteles, *Journal of Financial and Quantitative Analysis* 52, 1577-1604.

Cremers, Martijn and Antii Petajisto, 2009, How active is your fund manager? A new measure that predicts performance, *Review of Financial Studies* 22, 3329-3265.

Cremers, Martijn, Antii Petajisto and Eric Zitzewitz, 2012, Should benchmark indices have alpha? Revisiting Performance Evalution, *Critical Finance Review* 2, 1-48.

Daniel Kent, Mark Grinblatt, Sheridan Titman, and Russ Wermers, 1997, Measuring Mutual Fund Performance with Characteristic Based Benchmarks, *Journal of Finance* 52, 1035-1058.

Del Guercio, Diane and Paula Tkac, 2002, The determinants of the flow of funds of managed portfolios: Mutual funds versus pension funds, *Journal of Financial and Quantitative Analysis* 37, 523-57.

Doshi, Hitesh, Redouane Elkamhi and Mikhail Simutin, 2015, Managerial activeness and mutual fund performance, *Review of Asset Pricing Studies* 2, 156-184.

Elton, Edwin J., and Martin J. Gruber. "Marginal stockholder tax rates and the clientele effect." *The Review of Economics and Statistics* (1970): 68-74.

Elton, Edwin J., Martin J. Gruber, and Christopher R. Blake, 2011, "Holdings data, security returns, and the selection of superior mutual funds." *Journal of Financial and Quantitative Analysis* 46, 341-367.

Evan, Richard, 2010, Mutual fund incubation, Journal of Finance 65, 1581-1611.

Fama, Eugene, F., 1970, Efficient Capital Markets: A Review of Theory and Empirical Research, Journal of Finance

Fama, Eugene, and Kenneth French, 2010, Luck versus skill in the cross-section of mutual fund returns, *Journal of Finance* 65, 1915-1947.

Fama, Eugene F., and James D. MacBeth. "Risk, return, and equilibrium: Empirical tests." *Journal of political economy* 81.3 (1973): 607-636.

Ferson, Wayne and Kenneth Khang, 2002, Conditional Performance Measurement Using Portfolio Weights: Evidence for Pension Funds, *Journal of Financial Economics* 65, 249-282.

Ferson, Wayne and Haitao Mo, 2016, Measuring Performance with Market and Volatility timing and Selectivity, *Journal of Financial Economics* 121, 93-110.

Ferson, Wayne, Timothy Simin and Sergei Sarkissian, 2003a, Spurious regressions in Financial Economics, *Journal of Finance* 58, 1393-1414.

Ferson, Wayne, Timothy Simin and Sergei Sarkissian, 2003b, Is stock return predictability spurious? *Journal of Investment Management* vol. 1, no. 3, 10-19.

Ferreira, Kwswanit, Miguel and Ramos, 2013,

Frisch, Ragnar; Waugh, Frederick V. (1933). "Partial Time Regressions as Compared with

Individual Trends, *Econometrica 4, 387–401*.

Grinblatt Mark, and Sheridan Titman, 1989, Mutual fund performance: an analysis of quarterly portfolio holdings, *Journal of Business* 62, 393-416.

Grinblatt Mark, Sheridan Titman, 1993, Performance measurement without benchmarks: an examination of mutual fund returns, *Journal of Business* 60, 97-112.

Grinblatt, Mark, Sheridan Titman and Russ Wermers, 1995, Momentum investment strategies, portfolio performance and herding: A study of mutual fund behavior, *American Economic Review* 85, 1088-1105.

Harris, Hartzmark and Solomon, 2017.

Hjalmarsson, Erik, 2010, Predicting Global Stock returns, *Journal of Financial and Quantitative Analysis* 45, 49-80.

Hjalmarsson, Erik, 2008, The Stambaugh Bias in Panel Predictive Regressions, *Finance Research Letters* 5, 47-58.

Ippolito, Richard A. "Consumer reaction to measures of poor quality: Evidence from the mutual fund industry." *The Journal of Law and Economics* 35.1 (1992): 45-70.

Jarque, Carlos, and Bera Anil, 1986, Efficient tests for normality, homoscedasticity and serial independence of regression residuals, *Economics Letters*, 6(3), 255-259.

Jones, Chris and Haitao Mo, 2017, Out of sample performance of mutual fund predictors, working paper, University of Southern California.

Jordan, Bradford D., and Timothy B. Riley. "Volatility and mutual fund manager skill." *Journal of Financial Economics* 118.2 (2015): 289-298.

Jiang, George, Tong Yao and Tong Yu, 2007, Do Mutual Funds Time the Market? Evidence from Portfolio Holdings, *Journal of Financial Economics* 86, 724-758.

Kacperczyk, Marcin, Clemens Sialm, and Lu Zheng, 2008, Unobserved actions of mutual funds, *Review of Financial Studies* 21, 2379-2416.

Kothari, S.P. and Jerold Warner, 2001, Evaluating Mutual Fund Performance, *Journal of Finance* 56, 1985-2010.

Lilliefors, H. 1967, On the Kolmogorov-Smirnov test for normality with mean and variance unknown, *Journal of the American Statistical Association*, 62, 399-402.

Magkotsios, Georgios, 2017, Economies of Scale in fund Management, working paper, University of Southern California.

Newey, Whitney and K. West, 1987, A simple, autocorrelation and heteroskedasticity consistent covariance matrix, Econometrica

Pastor, Lubos and Robert Stambaugh, 2009, Predictive Systems: Living with Imperfect Predictors, *Journal of Finance* 64, 1583-1628.

Pastor, Lubos, Robert F. Stambaugh and Lucian Taylor, 2015, Scale and skill in active management, *Journal of Financial Economics* 116, 23-45.

Petajisto, Antti, 2013, Active Share and Mutual Fund Performance, *Financial Analysts Journal* 69, 73-93.

Petersen, Mitchell, 2009, Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches, *Review of Financial Studies* 22, 435-480.

Sirri, Eric and Peter Tufano, 1998, Costly search and mutual fund flows, *Journal of Finance* 53, 1589-1622.

Stambaugh, Robert, 1999, Predictive regressions, *Journal of Financial Economics* 54, 315-421.

Wermers, Russ, 2006, Performance Evaluation with Portfolio Holdings Information, *North American Journal of Economics and Finance* 17, 207-230.

Wermers, Russ, 2000, Mutual fund performance: An empirical decomposition into stock-picking talent, style, transactions costs and expenses, *Journal of Finance* 60, 1655-1695.

Wang, Junbo L., 2015, Can weight-based measures distinguish between informed and uninformed fund managers, PhD dissertation, University of Southern California.

Table 1: Summary Statistics

In panel A, for each fund in the sample we compute the time-series average of its total net assets (TNA) in millions of dollars, the average number of stocks held, the return gap and the active weight. We compute the sample standard deviations of the reported fund returns, σ, and the R-squares of factor model regressions in the Carhart (1997) four-factor model. We also compute the sample autocorrelations of the funds' portfolio weights, ρ_j, averaged across the holdings, at various lags j, j=1,…5. The means, std, max and min are taken across the funds in the sample. Error Corr is the correlation between the errors of the portfolio weight autocorrelation regression and the panel regression of future stock returns on the weights, averaged across the holdings. The sample period is from 1984 to 2012, and the number of funds is 3596. In Panel B, we compute the autoregression for the weights for each stock and average across the funds that hold the stock. The descriptive statistics are calculated at the stock level. In Panel C, we compute the autocorrelation of the first differences in the weights.

Panel A: Descriptive statistics at the fund level

	Mean	Std	Max	Min
TNA ($million)	684	2102	44496	1.03
Number of stocks	114	198	3317	10
Return Gap (%)	0.03	0.23	1.85	-2.6
Active Weight	0.40	0.10	0.90	0.01
Return σ	0.01	0.01	0.21	0.00
R-squares	0.87	0.14	0.99	0.06
ρ_1	0.9424	0.0398	0.9981	0.1957
ρ_2	0.8854	0.0730	0.9966	-0.0306
ρ_3	0.8282	0.1092	0.9962	-0.5256
ρ_4	0.7823	0.1277	0.9962	-0.5187
ρ_5	0.7366	0.1478	0.9962	-0.5187
Error Corr	-0.0043	0.0154	0.2430	-0.1868

Panel B: AR coefficients of the weights at the stock level

	Mean	Std	Max	Min
ρ_1	0.9027	0.0380	0.9971	0.2638
ρ_2	0.8132	0.0656	0.9921	0.1978
ρ_3	0.7289	0.0901	0.9868	-0.5539
ρ_4	0.6553	0.1040	0.9853	-0.4970
ρ_5	0.5856	0.1172	0.9852	-0.3851

Panel C: Average Autoregressive coefficients for weight differences

	Mean	Std	Max	Min
ρ_1	-0.0091	0.0180	0.0627	-0.1223
ρ_2	-0.0090	0.0185	0.0634	-0.1833
ρ_3	-0.0271	0.0422	0.1211	-0.4513
ρ_4	-0.0091	0.0184	0.0650	-0.1897
ρ_5	-0.0086	0.0177	0.0638	-0.0910

Table 2: Performance Measures with and without Carhart Adjustment

DGTW is the holdings-based performance measure from DGTW (1997) and GT is the Grinblatt and Titman (1993) measure. The subscript H indicates a measure in a model with fixed effects, estimated using the Hjalmarsson (2010) bias adjusted method. The Raw measures are without further adjustment, while the Carhart adjusted measures are the intercepts from regressing the monthly performance measures on the four Carhart (1997) factors. Avg α is the difference between a classical and a Hjalmarsson estimate. T-ratios are on the second line in parentheses, calculated similar to DGTW (1997) as the time series standard errors of the monthly performance measures for the average taken across the funds in the group. The Index Funds in Panel B are a sample of 201 index funds, 1994-2012. The units are annual percent.

Panel A: 1980-1994

	Measures					
	GT	GT_H	Avg α	DGTW	$DGTW_H$	Avg α
Raw Measures:						
All Active funds	2.08	1.45	0.63	0.40	0.55	-0.15
	(5.89)	(4.21)		(1.76)	(1.54)	
Aggressive Growth	4.06	3.13	0.93	1.15	1.62	-0.47
	(6.97)	(5.72)		(1.66)	(2.34)	
Growth Funds	1.98	1.31	0.67	0.38	0.48	-0.10
	(5.74)	(3.73)		(1.77)	(1.46)	
Carhart Adjusted:						
All Active funds	0.59	0.41	0.18	0.49	0.23	0.26
	(1.95)	(1.19)		(1.68)	(0.57)	
Aggressive Growth	1.90	1.40	0.50	1.59	1.03	0.56
	(3.28)	(2.52)		(2.35)	(1.31)	
Growth Funds	0.44	0.26	0.18	0.43	0.16	0.27
	(1.48)	(0.74)		(1.31)	(0.48)	

Table 2 page 2

Panel B: 1980-2012

	Measures					
	GT	GT_H	Avg α	DGTW	$DGTW_H$	Avg α
Raw Measures:						
All Active funds	1.32	0.96	0.36	0.30	-0.21	0.51
	(2.51)	(2.13)		(1.53)	(-0.57)	
Aggressive Growth	2.55	2.11	0.44	0.98	-0.07	1.05
	(2.62)	(2.29)		(1.90)	(-0.10)	
Growth Funds	1.28	0.85	0.43	0.32	-0.27	0.59
	(2.19)	(1.67)		(1.49)	(-0.66)	
Index Funds	0.07	0.069	0.01	0.32	-1.03	1.35
	(0.14)	(0.17)		(1.13)	(-3.02)	
Carhart Adjusted:						
All Active funds	-0.34	-0.28	-0.07	0.07	-0.64	0.72
	(-1.30)	(-1.71)		(0.38)	(-2.06)	
Aggressive Growth	-0.09	-0.16	0.07	0.69	-0.77	1.46
	(-0.17)	(-0.23)		(1.58)	(-1.21)	
Growth Funds	-0.52	-0.57	0.05	0.06	-0.72	0.79
	(-1.75)	(-1.71)		(0.29)	(-2.13)	
Index Funds	0.41	0.34	0.07	0.14	-1.18	1.32
	(0.55)	(0.75)		(0.36)	(-3.45)	

Table 3: Simulations of Hypothetical Strategies

Two strategies are simulated: buy-and-hold and momentum, following section 6.1. The average value of the measures (annualized percentage) are shown in the first row of each panel. The distributions of the T-statistics are summarized by reporting the values at the 2.5%, 5%, 50%, 95%, and 97.5% tails of the 1000 simulations. The percentage of simulated T-statistics larger than 1.96 are presented in the bottom row of each panel. GT and DGTW are the original estimates, while DiffIV uses the differenced IV method and Haj10 uses the Hjalmarsson (2010) estimator for the models with fixed effects.

Panel A: Simulated Buy-and-Hold strategy for GT and DGTW measures

	GT	DGTW	DiffIV GT	DIffIV DGTW	Haj10 GT	Haj10 DGTW
Average Measure	1.04	9.38	-0.03	0.03	0.04	-0.03
2.5%	-0.96	3.00	-2.37	-2.10	-2.24	-2.33
5.0%	-0.74	3.51	-1.97	-1.76	-1.98	-2.04
50%	1.18	7.50	-0.09	-0.01	-0.15	-0.34
95%	2.50	10.08	1.78	1.69	1.52	1.26
97.5%	2.77	10.48	2.10	1.86	1.74	1.52
>1.96	16.80%	99.80%	3.10%	1.90%	1.40%	0.50%

Panel B: Simulated Momentum strategy for GT and DGTW measures

	GT	DGTW	DiffIV GT	DiffIV DGTW	Haj10 GT	Haj10 DGTW
Average Measure	0.42	7.57	-0.06	-0.10	0.04	0.00
2.5%	-1.14	9.61	-2.31	-2.45	-2.20	-2.37
5.0%	-0.85	9.87	-2.00	-2.16	-1.83	-1.96
50%	0.84	11.39	-0.09	-0.38	0.10	-0.09
95%	2.70	13.38	1.90	1.48	1.87	1.73
97.5%	3.15	13.79	2.33	1.87	2.25	2.09
>1.96	16.70%	100.00%	4.80%	1.90%	4.20%	3.10%

Panel C: Simulated Buy-and-Hold strategy for CWM measures

	CWM	DiffIV CWM	Haj10 CWM
Average Measure	-1.38	-0.05	-0.04
2.5%	-1.94	-1.94	-2.01
5.0%	-1.82	-1.69	-1.80
50%	-1.01	-0.04	-0.04
95%	0.95	1.59	1.48
97.5%	1.20	1.78	1.79
>1.96	0%	1.9%	1.7%

Panel D: Simulated Momentum strategy for CWM measures

	CWM	DiffIV CWM	Haj10 CWM
Average Measure	3.50	-0.01	0.04
2.5%	-0.43	-2.20	-1.92
5.0%	-0.17	-1.85	-1.64
50%	1.81	0.00	0.05
95%	3.07	1.77	1.63
97.5%	3.29	2.08	1.91
>1.96	43%	3.5%	2.3%

Panel E: Simulated Buy-and-Hold strategy for FM measures

	FM	DiffIV FM	Haj10 FM
Average Measure	9.37	0.00	0.31
2.5%	1.76	-1.97	-1.98
5.0%	2.12	-1.81	-1.68
50%	4.09	-0.05	0.03
95%	6.09	1.57	1.39
97.5%	6.49	1.82	1.67
>1.96	96%	1%	1%

Panel F: Simulated Momentum strategy for FM measures

	FM	DiffIV FM	Haj10 FM
Average Measure	9.07	-0.02	0.02
2.5%	5.51	-2.07	-2.10
5.0%	5.76	-1.80	-1.75
50%	7.11	-0.05	-0.01
95%	9.06	1.81	1.86
97.5%	9.41	2.17	2.21
>1.96	100%	4%	4%

Table 4: The Cross-section of Holdings-based Performance Measures

The various performance measures are estimated for each fund in the sample using panel regressions. GT is the portfolio change measure, DGTW is the DGTW characteristic selectivity measure, FM is the Ferson and Mo SDF-based measure and CWM is the conditional weight based measure. Funds are sorted into five groups on the basis of the original measures, shown in the first column. The third quintile is the Median quintile. The second column, subscripted with H, indicates an average of the estimates for the funds in that quintile, including stock fixed effects in the panel regression and using the Hjalmarsson (2010) method. The third column is the average alpha effect extracted by the fixed effects. The average style change component of the GT measure is described in the text. The units of the performance measures are percent per month. HML is the difference between the top and bottom quintile measure. The sample period is from 1980 to 2012, and the number of active funds is 3596. Index Funds refers to the average in a sample of 201 index funds, 1994-2012.

Panel A: Grinblatt Titman Portfolio Change Measure

Fund Quintile:	GT	GT_H	Avg. Alpha	Avg. Style
Low	-3.49	-1.01	-2.70	0.21
2	-0.72	-0.63	-0.26	0.18
Median	0.19	0.00	0.04	0.15
4	1.20	0.76	0.24	0.20
High	4.19	3.14	0.81	0.25
HML	7.69	4.15	3.50	0.04
Index Funds	-0.84	-0.28	-0.70	0.14

Table 4, page 2

Panel B: DGTW Characteristic Selectivity Measure

Fund Quintile:	DGTW	$DGTW_H$	Avg. Alpha
Low	-3.65	-1.39	-2.26
2	-0.80	-1.03	0.23
Median	0.04	-0.68	0.71
4	0.82	-0.65	1.47
High	2.97	0.01	2.96
HML	6.61	1.40	5.21
Index Funds	0.31	-1.07	1.37

Panel C: Conditional Weight-based Measure

Fund Quintile:	CWM_H
Low	-0.93
2	-0.17
Median	0.03
4	0.51
High	0.98
HML	1.91
Index Funds	-0.35

Panel D: Ferson and Mo Stochastic Discount Factor Measure

Fund Quintile:	FM	FM_H	Avg. Alpha
Low	-19.36	-2.13	-17.23
2	-5.29	-2.48	-2.80
Median	-1.30	-1.69	0.39
4	1.56	-1.32	2.88
High	8.36	-1.28	9.64
HML	27.72	0.85	26.87
Index Funds	-11.53	-1.95	-9.58

Table 5: Holdings-based Performance Measures and Active Management
Funds are sorted by return-gap, R-square, active weight and volatility of returns in quintiles. Holdings-based measures are estimated for each fund using its available data and the averages are shown for each quintile. GT is the portfolio change measure, DGTW is the DGTW characteristic selectivity measure, and FM is the Ferson and Mo SDF-based measure and CWM is the conditional weight based measure. For each measure the second column, subscripted with *H*, indicates an estimate that includes stock fixed effects in the panel regression, estimated according to the method of Hajmarlsson (2010).

Panel A: Return-gap (1980-2012)

Fund Quantile	GT	DGTW	FM	CWM_H	$GT-GT_H$	$DGTW-DGTW_H$	$FM-FM_H$
High	2.37	0.45	1.69	1.32	0.32	0.49	2.39
	(2.74)	(1.22)	(0.53)	(1.37)	(1.12)	(1.07)	(0.89)
2	1.50	0.51	1.70	0.99	0.41	0.70	2.20
	(2.60)	(2.05)	(0.53)	(1.19)	(1.63)	(1.92)	(0.82)
3	1.02	0.29	1.43	0.75	0.34	0.74	2.28
	(2.05)	(1.48)	(0.45)	(0.98)	(1.91)	(2.15)	(0.86)
4	0.96	0.51	1.29	0.72	0.14	0.90	2.19
	(1.94)	(1.65)	(0.40)	(0.87)	(0.75)	(2.34)	(0.81)
Low	1.38	0.18	0.85	1.03	0.36	0.48	1.78
	(2.45)	(0.85)	(0.27)	(1.15)	(1.34)	(1.32)	(0.67)
Low-High	-0.98	-0.38	-0.70	-0.28	0.04	0.00	-0.61
	(-2.45)	(-1.36)	(-1.13)	(-0.89)	(0.22)	(-0.01)	(-1.41)

Panel B: R-square (1999-2012)

Fund Quantile	GT	DGTW	FM	CWM_H	$GT-GT_H$	$DGTW-DGTW_H$	$FM-FM_H$
Low	0.64	0.96	-0.65	0.12	0.29	1.80	1.57
	(0.92)	(1.88)	(-0.11)	(0.10)	(0.91)	(3.32)	(0.31)
2	-0.52	0.80	-1.52	-0.40	-0.09	2.41	2.19
	(-0.68)	(1.41)	(-0.26)	(-0.37)	(-0.16)	(2.93)	(0.39)
3	-0.77	0.24	-2.75	-0.09	-0.82	1.67	1.58
	(-1.01)	(0.46)	(-0.48)	(-0.10)	(-0.94)	(2.49)	(0.28)
4	-1.04	0.02	-3.46	-0.03	-0.98	1.56	0.93
	(-1.37)	(0.03)	(-0.62)	(-0.04)	(-1.03)	(2.37)	(0.17)
High	-0.87	0.01	-4.22	-0.26	-0.62	1.44	-0.61
	(-1.43)	(0.04)	(-0.74)	(-0.36)	(-0.87)	(2.27)	(-0.11)
High-Low	-1.50	-0.94	-3.58	-0.38	-0.91	-0.36	-2.18
	(-2.47)	(-1.84)	(-3.28)	(-0.54)	(-1.34)	(-0.83)	(-1.80)

Panel C: Active weight (1980-2012)

Fund Quantile	GT	DGTW	FM	CWM_H	$GT-GT_H$	$DGTW-DGTW_H$	$FM-FM_H$
Low	0.94	-0.14	0.24	0.67	0.10	0.40	1.17
	(1.51)	(-0.73)	(0.08)	(0.49)	(0.55)	(0.99)	(0.49)
2	0.78	0.24	0.71	0.65	0.20	0.82	1.93
	(1.57)	(1.09)	(0.22)	(0.52)	(1.10)	(2.23)	(0.79)
3	1.25	0.23	0.95	0.55	0.45	0.69	1.79
	(2.49)	(1.09)	(0.30)	(0.41)	(3.19)	(1.89)	(0.74)
4	1.13	0.38	1.22	0.62	0.28	0.90	2.14
	(2.06)	(1.43)	(0.38)	(0.49)	(1.82)	(2.43)	(0.88)
High	1.69	0.75	2.07	1.53	0.60	0.75	2.18
	(3.23)	(2.00)	(0.63)	(1.18)	(2.96)	(1.79)	(0.89)
High-Low	0.75	0.89	1.82	0.86	0.50	0.35	1.01
	(2.06)	(2.95)	(2.49)	(1.44)	(2.65)	(1.44)	(2.44)

Panel D: Return volatility (1999-2012)

Fund Quantile	GT	DGTW	FM	CWM_H	$GT-GT_H$	$DGTW-DGTW_H$	$FM-FM_H$
Low	0.35	0.42	-1.53	-0.08	0.37	1.49	0.78
	(0.75)	(0.88)	(-0.28)	(-0.14)	(1.48)	(2.91)	(0.16)
2	-0.45	0.40	-2.33	-0.08	-0.18	1.70	0.43
	(-0.86)	(0.97)	(-0.43)	(-0.12)	(-0.48)	(2.62)	(0.09)
3	-0.54	0.28	-3.16	-0.08	-0.66	1.63	0.42
	(-0.80)	(0.78)	(-0.55)	(-0.09)	(-1.07)	(2.46)	(0.08)
4	-0.82	0.54	-2.96	-0.03	-0.80	2.01	1.28
	(-0.94)	(0.86)	(-0.50)	(-0.02)	(-0.93)	(2.31)	(0.23)
High	-1.15	0.35	-2.68	-0.42	-1.00	2.05	2.72
	(-1.09)	(0.26)	(-0.42)	(-0.32)	(-0.81)	(1.70)	(0.40)
High-Low	-1.50	-0.06	-1.15	-0.33	-1.37	0.57	1.94
	(-1.60)	(-0.04)	(-0.45)	(-0.43)	(-1.24)	(0.39)	(0.57)

Table 6: Fund Characteristics and Average Alpha Effects

The Average Alpha Effects in classical performance measures are regressed in a panel on lagged fund characteristics. DGTW is the performance measure from DGTW (1997) and GT is the Grinblatt and Titman (1993) measure, FM is the Ferson and Mo (2016) measure. TBH is a measure of the tendency towards buy-and-hold, measured as the average absolute difference between the funds' portfolio weights and what they would have been had the fund had the same holdings as they did 12 months ago. The monthly measure is averaged over the past year to reduce noise. LM is the lagged momentum measure of Daniel, Titman and Wermers (1995), capturing the relation between a fund's deviation from buy-and-hold weights and the average returns on the stocks over the last 12 months. Div Yield is a fund portfolio weighted average of the dividend per share of a stock during the past 12 months divided by its price per share. HLDSize is the portfolio weighted market capitalizations of the stocks held by the fund. Aggressive is a dummy variable indicating an aggressive growth style fund. Log(TNA) is the log of total net assets, Exp is expense ratio, Fund Age is the number of month from fund's first offering date, and turnover is the reported annual turnover. T-ratios are on the second line, calculated by clustering by time and using Newey-West (1987) covariance terms to 30 lags. The units of the Average Alpha Effects are annual percent.

Panel A: 1980-1994

	GT alpha	DGTW alpha	FM alpha
Const	0.00	0.00	0.00
T Stat	0.06	-0.97	-1.28
Fund age	0.00	0.00	0.00
T Stat	-0.74	-0.75	-1.59
Turn ratio	0.00	0.00	0.00
T Stat	0.93	0.38	0.31
Aggressive_ind	0.00	0.00	0.00
T Stat	-0.74	-0.31	-0.44
Expense ratio	-0.01	0.03	0.08
T Stat	-0.41	1.13	0.71
log(TNA)	0.00	0.00	0.00
T Stat	-0.40	0.27	-2.00
Div yield	0.00	0.00	0.02
T Stat	**4.42**	1.40	1.68
log(Mkt_cap)	0.00	0.00	0.00
T Stat	-0.99	-1.08	-1.14
BHL	0.00	0.00	0.01
T Stat	0.37	**7.47**	**5.81**
LM	-0.01	-0.02	0.08
T Stat	-0.73	-0.95	1.39

Panel B: 1980-2012

	GT alpha	DGTW alpha	FM alpha
Const	0.00	0.00	0.00
T Stat	-0.23	-0.29	0.46
Fund age	0.00	0.00	0.00
T Stat	1.16	-0.45	-0.98
Turn ratio	0.00	0.00	0.00
T Stat	-0.12	**2.85**	1.34
Aggressive_ind	0.00	0.00	0.00
T Stat	-0.81	0.95	0.59
Expense ratio	0.00	0.00	0.04
T Stat	-0.37	-0.03	0.55
log(TNA)	0.00	0.00	0.00
T Stat	0.32	-0.50	-1.94
Div yield	0.00	0.00	0.03
T Stat	**2.62**	**2.16**	**2.36**
log(Mkt_cap)	0.00	0.00	0.00
T Stat	-1.14	0.14	-1.26
BHL	0.00	0.00	0.00
T Stat	0.55	**2.80**	0.94
LM	-0.04	-0.01	0.06
T Stat	**-4.15**	-1.09	1.52

Table 7: Fund Flows, Average Alpha Effects and Characteristics

The annual new money flows are regressed on fund performance, measured as their Average Alpha Effects in various performance measures over the past year. The Average Alpha Effects are estimated as in Table 3 and are used to form ranked performance measures within each of five performance quintiles, as in Sirri and Tufano (1998). In panel A there are no controls for fund characteristics, and in panel B fund characteristics are included as control variables. The sample period is 1980-2012. DGTW is the performance measure from DGTW (1997) and GT is the Grinblatt and Titman (1993) measure, FM is the Ferson and Mo (2016) measure and Panel CWM is the panel version of the conditional weight measure of Ferson and Khang (2002). TBH is a measure of the tendency towards buy-and-hold, measured as the average absolute difference between the funds portfolio weights and what they would have been had the fund had the same holdings as they did 12 months ago. The monthly measure is averaged over the past year to reduce noise. LM is the lagged momentum measure of Daniel, Titman and Wermers (1995), capturing the relation between a fund's deviation from buy-and-hold weights and the average returns on the stocks over the last 12 months. Div Yield is a fund portfolio weighted average of the dividend per share of a stock during the past 12 months divided by its price per share. HLDSize is the portfolio weighted market capitalizations of the stocks held by the fund. Aggressive is a dummy variable indicating an aggressive growth style fund. Log(TNA) is the log of total net assets, Fund Age is the log of number of month from fund's first offering date, Exp is expense ratio, turnover is the reported annual turnover, style flow is the aggregate flows of funds within the same style, and return std and autocorrelations are standard deviation and autocorrelation of the fund returns. T-ratios are on the second line in parentheses, calculated by clustering by time and using Newey-West (1987) covariance terms to 30 lags. The units of the Average Alpha Effects are annual percent.

Panel A: Average Alpha Effects Only

Measure	GT	DGTW	FM
Coeff	Flow	Flow	Flow
Const	0.02	0.01	0.01
T	4.01	4.99	4.86
alpha Bottom	-0.09	0.05	-0.14
T	-2.41	0.39	-5.87
alpha 2nd Quantile	-0.11	-0.07	-0.05
T	-3.38	-6.27	-1.40
alpha 3rd Quantile	-0.09	-0.01	-0.01
T	-2.72	-0.37	-0.35
alpha 4th Quantile	-0.04	0.01	-0.04
T	-2.29	0.48	-1.89
alpha top	-0.01	**0.07**	**0.07**
T	-0.34	**7.25**	**3.69**

Table 8 panel B: Fund Characteristics Included

Measure	GT	DGTW	FM
Coeff	Flow	Flow	Flow
Const	0.04	0.02	0.03
T	3.85	2.43	2.72
Fund Age	-0.01	0.00	-0.01
T	-2.43	-3.53	-4.11
Fees	0.19	0.20	0.19
T	0.95	0.94	0.91
Log(TNA)	0.00	0.00	0.00
T	2.70	2.49	2.48
Style flow	1.35	1.35	1.36
T	25.63	25.35	25.31
Return std	-0.16	-0.17	-0.17
T	-2.03	-2.04	-2.09
Autocorrelation	0.00	0.00	0.00
T	1.17	1.18	1.19
Turnover	0.01	0.01	0.01
T Stat	6.27	6.37	6.29
Aggressive	0.00	0.00	0.00
T Stat	1.63	1.56	1.51
Div yield	-0.01	-0.01	-0.01
T Stat	-0.65	-1.12	-1.22
HLDSize	0.00	0.00	0.00
T Stat	-2.53	-1.71	-1.82
BHL	-0.03	-0.03	-0.03
T Stat	-4.31	-2.81	-3.23
LM	-0.66	-0.66	-0.66
T Stat	-1.79	-1.75	-1.77
alpha Bottom	-0.03	-0.06	-0.07
T	-1.15	-2.61	-3.03
alpha 2nd Quantile	0.01	-0.09	-0.10
T	0.27	-1.59	-1.71
alpha 3rd Quantile	-0.02	-0.03	-0.05
T	-1.16	-0.42	-1.11
alpha 4th Quantile	0.25	-0.03	0.01
T	1.24	-0.77	0.35
alpha top	0.03	-0.02	0.02
T	1.74	-0.39	0.62

Table A.1: Simulations to Address Stochastic Regressor Bias

The table reports the average across 1000 simulation trials, of estimated holdings based performance measures. The units are percent excess return per quarter. True denotes the actual values of the measures used to calibrate the simulations, bootstrapping from versions of Equations (8) and (14). These are found at various fractiles in the cross-section of actual fund data. The "No alpha" OLS estimates the slope coefficient in the baseline panel regression of (6), Diff IV is the differenced IV estimator, Diff no IV is the classical difference estimator of the regression (8) with stock dummies and Within is the classical within-group (least squares with dummy variables) estimator. OLS with alpha is based on the model in footnote 7. Haj2010 is the bias adjusted estimator in (18) and Haj2008 is the bias adjusted estimator in (16). GT is the weight change measure of Grinblatt and Titman (1993), DGTW is the measure of Grinblatt, Titman and Wermers (1997). CWM is the panel conditional weight measure introduced in this paper. FM is the stochastic discount factor measure of Ferson and Mo (2016).

Panel A: Alternative Holdings-based Performance Estimators

True	No alpha OLS	OLS with alpha	Haj2008	Within	Diff	Diff IV	Haj2010
-0.29	0.13	-0.16	-0.25	-0.37	-0.44	-0.30	-0.30
-0.19	0.17	-0.01	-0.14	-0.32	-0.22	-0.22	-0.20
0.06	0.08	0.04	0.17	-0.02	0.00	0.04	0.05
0.40	0.48	0.39	0.42	0.37	0.35	0.40	0.41
0.58	0.92	0.58	0.60	0.46	0.40	0.60	0.58

Panel B: Classical Measures when there are no Fixed Effects

True (GT)	GT	True (DGTW)	DGTW	True (CWM)	Haj10(CWM)	True (FM)	FM
-0.29	-0.29	-0.26	-0.29	-0.15	-0.15	-0.54	-0.54
-0.19	-0.17	-0.18	-0.18	-0.08	-0.08	-0.37	-0.37
0.06	0.06	0.00	0.03	0.07	0.06	0.00	0.01
0.40	0.41	0.22	0.23	0.20	0.20	0.23	0.24
0.58	0.59	0.35	0.32	0.28	0.26	0.38	0.38