An Anatomy of Fundamental Indexing*

Lieven De Moor,† Fang Liu,‡ Piet Sercu,§ and Tom Vinaimont¶

First draft: Summer 2011; this version Monday 29th April, 2013
Preliminary—comments very welcome

*We thank Anup Basu, Gaël Imad Eddine, Aurobindo Ghosh, Hening Liu, Vestman Roine, and Martina Vandebroek, and participants in presentations at the authors' home institutions, the 2011 Indian Finance Conference (Bangalore), the 2012 BRForum (Antwerp), the 2012 Paris Finance Meeting, and the 2012 Symposion on Banking and Finance (Shanghai) for useful comments. Any remaining errors are the authors' sole responsibility.

†Hogeschool-Universiteit Brussel, and KU Leuven, Faculty of Economics and Business; lieven.demoor@hubrussel.be

‡Central University for Finance and Economics, Beijing; liufang8000@yahoo.com.cn

§KU Leuven, Faculty of Economics and Business; piet.sercu@econ.kuleuven.be

¶City University of Hong Kong; tom.vinaimont@cityu.edu.hk

Abstract

Proponents of Fundamental Indexing (FI) suggest that it is more profitable to base portfolio weights on indirectly size-related indicators like accounting data rather than directly on market caps. In noisy markets *à la* Roll (1984), it is argued, underpriced stocks (which typically overperform) get underweighted and *vice versa*. This negative interaction implies a 'drag', which FI claims to avoid.

The key question is to what extent the extra return that FI pays really reflects drag avoided rather than style shifts. Multifactor regression estimates of alpha are not robust to alternative specifications, and suffer from demonstrably non-constant factor sensitivities. Generalized Roll (1984) estimates of the standard deviation and autocorrelation of mispricing do confirm there must be more going on than drag avoidance. Cross-sectional regression studies of the differences between FI weights and value weights likewise tell us that the adjustments are substantially larger than would be necessary just to avoid drag; in addition, the cross-sectional adjustment patterns are found to be quite variable over time. In short, not only there must be style shifts, but they are also unstable.

To estimate the pure benefits from drag avoidance, purged of style shifts without having to rely on style regressions, we study an investment strategy that should be immune to drag, but without much style shift: we sort stocks into twenty size buckets (vigintiles), and form a portfolio where a stock's weight equals the average value-weight of all stocks in its vigintile. We find no meaningful extra return, whether at the total-portfolio level or per vigintile. Thus, avoiding drag is not why FI does well: drag is empirically unimportant. Most or all of the *prima facie* benefits must be from time-varying style shifts.

Contents

JEL classification: G11, G12, G14
Key words: fund management, drag, noisy prices.

Introduction

In the portfolio-theory/efficient-markets view, the familiar recommendation is to mimic the market portfolio, i.e. to choose weights proportional to the firms' relative market capitalizations. Proponents of the Fundamental Indexing (henceforth FI) investment strategy beg to disagree: while big stocks should still be allocated more money, their view is that weights should be calculated not from market capitalisations but from instrumental-variable-like alternatives to market value, such as book value, payout, cash flow, or sales. In noisy markets, the argument goes, directly using market-cap-based weights creates a 'drag', a negative interaction between weighting errors and subsequent return: tautologically, a value-based weight will over-weight overpriced stocks (the very stocks that tend to underperform) and will under-weight the underpriced stocks (the stocks that tend to overperform). If the market-value weight is replaced by an instrument, the interaction between the weighting error and the degree of mispricing is avoided or at least reduced, so that expected returns improve.

In most empirical studies, including ours, the proposed alternative weighting schemes are effectively found to boost returns. But generalized Jensen regressions also reveal that the shifts in the weights do affect exposures to risks and that FI's exposures are unstable. Thus, at least part of the extra return that FI offers is due to style shifts and to market timing in the Hendriksson-Merton sense. In addition, the alpha estimates are sensitive to the factor specification and the time period. In short, regression analysis is unlikely to provide a definitive answer as to whether avoiding drag is a significant contributor of FI's extra return.

In this paper we present three novel approaches to the question how large drag really is. One approach is to tentatively interpret the FI extra returns as pure drag, and then use return autocovariances to back out, in the Roll (1984) tradition, a standard deviation and autocorrelation of the pricing errors. The implied noise turns out to be so high that we can safely reject the notion that FI's extra return is pure drag indeed. In our second approach we study not the time series of extra returns but the cross sections of FI weight adjustments, and we test the idea that these are just a rational Bayesian reaction to noise in prices. We find that FI's adjustments to the weights are substantially larger than a rational Bayesian response, which again tells us that FI is not all about avoiding drag. We also find that the adjustments are highly unstable over time. This is in line with the conclusions from the regression tests: not only there are style shifts, but these shifts are also far from constant. As our third contribution, we test a variant of the alternative proposed by Basset, Chen and Chen (2003). They advocate the use of lagged value weights rather than fundamental weights, arguing that old weighting errors should be less correlated with current weighting errors and, therefore, with the subsequent return. Instead of using one set of lagged weights, as they do, we work with either one-, three- or six-months-lagged value weights. The older the weights, the less their error should resemble the current degree of mispricing and, therefore, the weaker the interacting with the current period's return, and the better the return. Lastly, we propose to study a new, 'mixed' investment strategy, that should avoid the drag effect with far less style shift than FI. Specifically, we sort stocks into 20 size buckets and then use equal weighting within each bucket but value weighting across buckets. It turns out that this 'VW/EW' mixed strategy offers no extra returns relative to full value weighting. That is, the direct gains from purely avoiding drag are economically insignificant, implying that virtually all of FI's extra return must be from other sources, like unstable style shifts and timing.

The issues

In the remainder of this introduction we review the above ideas in more detail. FI, as explained above, makes two claims (Arnold, Hsu and Moore, 2005): noisy valuations impose a drag on the portfolio return, but that drag can be avoided. Specifically, if the portfolio weights are set on the basis of instruments for market value (like book values, cash flows, dividends, or sales, or a combination of all these), the errors in the weights are no longer correlated with the unexpected returns. But drag avoidance might not be the sole channel through which extra returns are generated. As sceptics stress, FI puts more weight on small and 'value' stocks, so part of the extra return reflects style shifts. For fundamental analysis to be really useful there must be a residual average effect, not explained by the style shifts and market timing.

When the objective is to avoid drag, there are alternatives to FI. Treynor (2005) already notes that equal weighting (EW) should dilute away almost all mispricing problems, thus efficiently avoiding any interaction between weighting errors and abnormal returns. FI proponents reply that EW is infeasible in practice and creates a gigantic style shift, two problems that are substantially reduced when FI is adopted. But Basset, Chen and Chen (2007) suggest another alternative: use lagged weights (LW). This is simpler than FI, and probably largely preserves the style chosen by conventional value weighting (VW) because weights tend not to change drastically over a few months.

All of the proposed alternative weighting schemes (EW, FI and LW) do increase performance relative to VW. As a prelude to our review of the empirical literature we first show our own findings. (The data, described in Section 1.1, bear on U.S. stocks, 1990-2009.) Table 1 shows the *prima facie* benefits from FI, in percent per month, when portfolio weights are

Table 1: Performance of Fundamental Indices compared to the value-weighted (VW) index, in %.

	Avge (%/mo)	Stand. dev.	Sharpe ratio	Δ vis-à-vis VW (%/mo)	t-stat	*p.a.*. (%/yr)	returns extra
VW	0.733	4.59	0.089			7.83	
BV	1.045	4.90	0.147	0.312	3,59	11.62	3.79
SL	0.977	4.84	0.135	0.244	2,02	10.82	2.99
CF	0.961	4.87	0.131	0.228	2,09	10.56	2.73
EW	1.187	5.99	0.144	0.454	2.20	12.82	4.99
LW1	0.760	4.69	0.093	0.026	1.09	5.05	0.26
LW3	0.753	4.83	0.089	0.019	0.43	7.96	0.13
LW6	0.740	4.98	0.083	0.007	-0.10	7.70	-0.13

Key The table shows average monthly returns, standard deviation, Sharpe ratios and monthly excess returns *v* the value-weighted (VW) Index for Fundamental Indexing strategies with weights based on book value (BV), sales (SL) and the (absolute value of) the free cash flow (CF). Also shown are results for portfolios where weights are equal (EW), or set equal to the value weight as observed 1, 3 or 6 months ago (LW1-6). The Sharpe ratios are not annualized. The last two columns show the geometric *p.a.* mean returns from following the stratagy for 20 years, and the differences of the latter relative to the VW figure.

based on relative Book Value, Sales and the absolute value of Free Cash Flow, respectively.[1] We also show the returns from applying EW and lagged weights (LW1, LW3 and LW6). These three LW strategies are different from the Basset *et al* rule, which took as its lagged weight the median weight over the preceding 12 months. As motivated above we lag the weights by either one, three or six months. In the logic underlying FI, older weights should provide better returns, consistent with lower drag.

The returns look good for EW and FI, but nor for LW. Monthly returns increase on average by 0.312% when weighting is based on Book Values, and by 0.244% and 0.228% when weighting is by Sales and the absolute value of Free Cash Flow, respectively. Annualized, this means 2.7-3.8% extra.[2] In our experiment, this comes with a small increase in risk but, unlike the rise in return, an increased volatility is not a common finding in empirical work of this kind. And even in our calculations the Sharpe ratios are still up; that is, by de-levering, one could have maintained volatility and still come out ahead in terms of return. EW pays an even

[1] Absolute cash flow works better than the standard versions, CF itself or Max(CF, 0). A hugely negative cash flow tends to occur only in big firms.

[2] The analysis is systematically done in terms of monthly returns, and the averages we show are the simple averages as used in standard statistics. However, these mean monthly returns could be poor indicators of long-term cumulative returns if there are serious autocorrelation patterns. The last two columns accordingly show annual returns, computed as internal rates of return *per annum* from the cumulative compound outcomes. Across the seven strategies, there is a 0.997 correlation between simple monthly averages and geometric annual averages. Thus, it looks correct to study just the mean monthly returns.

more generous extra return, nigh 1.2 percent per month, with a predictably larger standard deviations (6%, against 4.5-5% for the other strategies). While its Sharpe ratio is not as good as the BV-based one, it still beats the other FI variants and VW. Lagged returns, finally, barely (and insignificantly) beat VW in terms of mean return and Sharpe ratio. Actually, against the FI prediction, the extra return falls with the lag length. Whether this reflects an autocorrelation pattern in the true returns, or changing style shifts, or just a coincidence is far from clear: the drop is not significant. Still, our results for the Basset *et al.* test, if valid, already suggests that drag may be statistically and economically insignificant and, by implication, that FI's extra return probably has another source than drag avoided.

The issue, then, is how we can reliably find out which tests are right about the drag, and what explains the substantial differences between the LW and the FI results.

Prior Work

Our results, above, fit into an extensive literature. Most studies find that FI-based portfolios do outperform value-weighted ones. Arnott, Hsu and Moore (2005), first, note that fundamental indexation portfolio outperform the S&P 500 by 2% on average per year between 1962 and 2004 while the volatility is very close to that of a cap-weighted index. Hemminki and Puttonen (2008) reweight the components of the Dow Jones Euro Stoxx 50 and find a similar return advantage. Stotz, Döhnert and Wanzenfried (2007) confirm these conclusions for the DJ Euro Stoxx 600 stocks. Neukirch (2008) compares an internationally diversified portfolio of FI-weighted ETFs to the MSCI cap-weighted world index and finds overperformance. Similarly promising extra returns have also been documented by e.g. Jun and Malkiel (2008), Asness (2006), Walkhäusl and Lobe (2010), Houwer and Plantinga (2009), Peltomäki (2010), and Mihm and Locarek-Junge (2010). Sometimes gains seem to be huge: Ferreira and Krige (2011) come forth with an extra return of 4.7% for South Africa, while Forbes and Basu (2011) conclude that in Australia the gain was above 5%. Biltz, Van de Grient and Van Vliet (2010) obtain a downright impressive 10% extra return if re-weighting is done as early as March. This may be optimistic, as they hasten to add: if on March 1 the accounting information is not actually available yet, these calculations would definitely create a look-ahead bias. But, Biltz *et al.* note, reweighting in September instead of in spring, for instance, kills all extra returns, a result that flatly contradicts the FI logic.

Biltz *et al.* are not the only skeptical authors. Chen, Chen and Basset (2007), as mentioned, propose past stock prices as an alternative to accounting-based weights. Unlike us, they find that this trading rule also outperforms the cap-weighted index;[3] still, the extra return is below that of a FI strategy. Other authors, however, question FI's very logic. Kaplan (2008) notes that the FI weights may still be correlated with the markets valuation errors, in which case FI works less well than promised. Perold (2007) shows analytically that a strategy of beefing up the weights for lower-cap stocks and *vice versa* does not produce any extra expected return—at least when, critically, the Bayesian prior about the true values is diffuse and returns are measured by log changes rather than orthodox simple returns. (We return to this analysis later.) But, as mentioned, most critics take the line that FI introduces size and value biases, and that the extra return is just a reward for non-beta exposures (e.g. Fama and French, 2007; Dopfel, 2008; Subramaniam, Kulkarni, Kouzmenko and Melas, 2011; Perold, 2007; and others cited below).

The empirical findings on risk-adjusted returns are mixed. Many studies do find that, after the standard regression-based correction for size and value exposure, there still is an alpha return left (Stotz *et al.*, 2007; Houwer and Plantinga, 2009; Peltomäki, 2010; Mihm and Locarek-Junge, 2010; and Forbes and Basu, 2011). Others disagree, like Jun and Malkiel (2008), Amenc, Goltz and Le Sourd (2008), and Walkhäusl and Lobe (2010).

One possible explanation of the conflicting findings is that the factor exposures of FI-weighted portfolios are not constant and that changes in the betas may sometimes be correlated with factor returns, in which case the standard regression intercept does not provide a reliable estimate of the abnormal return. Walkhäusl and Lobe (2010), in fact, provide evidence that correlations between exposures and factor returns do exist: the regression coefficients for Henriksson-Merton squared factor returns are significant. We come to the same conclusion in Section 1. Our analysis, in Section 2.2 below, of the time series patterns in the weight shifts as administered by FI similarly shows that the style changes are very variable over time, which again make it implausible that factor exposures would be constant over time. The challenge, then, is how to eliminate this nuisance factor in the diagnosis. The Henriksson-Merton (1981) squared factor return as an added return is able to detect, at best, linear relations between exposure and expected return; that is, it may document that variations in exposure are correlated with changes in expected factor returns (market timing), but it may miss other changes in exposure and does not allow any generally valid conclusions about whether there actually is a genuine alpha.

[3] There are important design differences. Basset *et al.* focus on the 1000 largest stocks; their sample period is 1962-2003; and they use the 12-month median of the cap-weights rather than weights for fixed lags. We have more stocks (3400 per cross-section, on average), a different period (1990-2009), and three fixed lags.

Contribution and structure of the paper

In Section 1 we report the results from our own standard performance analysis. We obtain rather unsatisfactory conclusions: the results depend on the choice of the factors and the period, and exposures are correlated with factor returns. So in the rest of the paper we try to find out what part of FI's extra returns stems from drag avoidance as opposed to style shift, but without having to rely on regressions.

A streamlined version of the math is provided in Appendix and reviewed in Section 2.1. It provides the basis for three tests. First, we combine FI extra returns with autocovariances, and we back out lower bounds on the unconditional standard deviation and persistence of pricing errors under the Null of no style shift. Second, we do a cross-sectional analysis of weight shifts (Section 2.3) which again gives us bounds on the noise variance, but this time per size class and per period. These tests also give strong hints about the variability of size-related style shifts. Third, we provide size-controlled comparisons of FI, LW and EW returns (Section 3). The main findings are as follows. First, the amount of pricing noise required to reconcile extra returns with observed autocovariances under the Null of no style shift is implausibly high. Thus, we conclude there is style shift. Second, the cross-sectional analysis of FI's weight corrections reveals huge shifts towards the mean. In fact, the mean reversion totally overshoots the degree of Bayesian shrinkage one expects if the portfolio manager's objective just were to avoid drag. In addition, the degree of shrinkage waxes and wanes over time, with episodes where the relation with size is quadratic rather than just linear and other episodes where no link with size seems to be present. This makes it implausible that factor exposures would be constant. Third, the direct estimates of drag show it to be economically insignificant. The optimistic interpretation is that price noise is smaller than one might fear; but an alternative view is that too much of the noise is market-wide or persistent, making the drag in monthly returns small even if noise is important. Whatever the interpretation, though, the amount of drag in monthly returns created by value weighting is *not* a serious problem.

1 A conventional performance analysis

1.1 Data

We obtain end-of-month dollar returns for US common stocks adjusted for stock splits and dividend payments as well as their beginning-of-the-month market caps from Thomson Reuters Datastream (TRD) for the period January 1990 till May 2009. The data are carefully filtered for errors following the procedures adopted in *e.g.* Ince and Porter (2006). The corresponding beginning-of-the-month fundamentals (*i.e.* book values, sales and free cash flows) are also from TRD. In any given month we eliminate stocks with a market capitalization smaller than \$10 million, a monthly trading volume smaller than \$100,000 or a price smaller than \$1.[4] Each month we also exclude stocks with negative or zero book value or sales.[5] Note that both the market cap and all fundamental data need to be available if we want to work with one common sample for all strategies. Working with such a common sample has the advantage that differences in performance across the weighting schemes cannot be driven by differential data availability but, on the other hand, has the disadvantage that the sample size depends on the least available fundamental, namely the free cash flow. The average cross-section contains roughly 3,400 ongoing stocks (1,000 in 1990, 5,000 as of 2000) over a period of almost 20 years or, more precisely, 232 months. For the rebalancing frequency of the Fundamental Indices we go for monthly revisions, the standard in this line of research. Monthly rebalancing of the FI portfolios seems realistic for both institutional and private investors who need to weigh transaction costs against tracking error. Obviously this is only relevant for the fundamental indices, not for the cap-weighted one, where rebalancing is necessary only in case of stock repurchases or share issues. We do not correct for transaction costs and the dataset is free of survivorship bias. The asset-pricing factors and the T-bill rates, lastly, are from French's website.

1.2 Results of a conventional performance analysis

For the FI, EW, and LW portfolios we compute returns in excess of the VW return, we regress them on the Fama-French factors augmented with the momentum factor, following Carhart (1997), and we do the standard Newey-West t-test for a zero value of the intercept, alpha.

The results are summarized in Panel A of Table 2. Since the left-hand variable is a return in excess of the VW return, the coefficients estimate the exposures to market, SMB, HML and Momentum as differentials relative to those of VW. These numbers do reveal some style shift, as usual. Specifically, relative to VW, FI imparts a small but significant upward boost to the

---

[4] This is to eliminate tiny, illiquid and penny stocks which are reasonably more likely to contain data errors. Penny stocks are often fallen angels (Chan and Chen, 1991) which are highly speculative and illiquid. Tiny companies likewise have limited liquidity, can be subject to high price pressure or price manipulation, and often represent too little value to warrant attention.

[5] Negative and zero book values and sales are almost surely mistakes. Negative free cash flows, in contrast, make more sense and are much more prevalent. We do not weight on the basis of dividends because too many stock/year combinations then produce zero weights, leading to a portfolio that is biased against small and distressed stocks.

portfolio's sensitivity to the market, somewhat more to its exposure to SMB, and a lot to its sensitivity to HML. More unexpectedly perhaps, there is negative exposure to momentum, a phenomenon we return to below. Among FI's competitors, EW is especially exposed to SMB, as one would expect, but hardly more to market and HML than the VW portfolio. Using lagged weights affects style far less than adopting FI, again as expected, except for some exposure to SMB (LW3 and LW6). We conclude there are style shifts. But the alphas tell us that these increased exposures do not seem to explain away all of the extra return we have noted before. In fact, all FI strategies provide a positive risk-adjusted return of 0.2-0.3 percent per month, which is large in both statistical and economic terms; EW, with almost 0.4 percent per month, does even better; and also LW6 seems to add value, once its lower risk is accounted for, while LW1 and LW3's alphas approach significance ($p = 0.069$ and 0.049, one-sided).

But the conclusion may be less straightforward than it seems at this stage. The strongly negative extra exposure to momentum, for one, is somewhat puzzling, and may really mean a positive exposure to reversal. The potential for a link between momentum/reversal and the weighting scheme is logically clearest for the LW strategies. The difference between the LW and VW returns is driven by the change of weights over the preceding L months, which in turn is determined by the stock's return in excess of the market return during that period; so weighting on the basis of the change in relative market cap means betting on reversal, against momentum.[6] For FI strategies there is a link with reversal too, albeit weaker: the FI weights are insensitive to recent changes in market value, so the differences with the VW return are driven by two elements: the initial difference between the lagged VW weights and the FI ones, and the changes in the value weights between $t - L$ and t. This last item is the same as for LW, so the presence of the first extra source of deviations may obscure the link with recent changes in value weights but should not wipe it out.

In order to determine whether momentum might be at work in the lagged-weight strategy, we allocate stocks at the beginning of each month t to either of three performance categories. The three categories contain respectively the 25% worst, 50% middle and 25% best performing stocks as measured via the average return for the months $t-7$ until $t-2$.[7] For each performance category, we calculate returns for portfolios *à la* Basset *et al.*, weighted by market caps lagged over 1 month, 3 months and 6 months, and we compare them to a a standard VW portfolio

[6] For example, if a stock has underperformed the market, then the lagged weight it gets is higher than the current one. If giving a positive differential weight to past losers seems to pay, as it does here, we can conclude there was reversal.

[7] It is standard to limit the evaluation period to $t - 2$. If not, a possible pricing error in V_{t-2} would affect both the sorting procedure and the subsequent portfolio return.

Table 2: Performance analysis of competing investment strategies (returns in excess of VW)

Panel A: FF factors plus Momentum

	α (%/month)		market (β)		SMB (γ)		HML (δ)		Momentum (ζ)	
	coeff	tstat	coeff	tstat	coeff	tstat	coeff	tstat	coeff	tstat
BV	0.298	6.044	0.030	2.504	0.078	5.311	0.266	16.280	-0.120	-12.539
SL	0.231	3.289	0.004	0.204	0.073	3.473	0.378	16.198	-0.152	-11.060
CF	0.221	3.451	0.034	2.172	-0.030	-1.583	0.322	15.188	-0.125	-10.046
EW	0.380	3.468	0.001	0.023	0.772	23.495	0.125	3.437	-0.151	-7.070
LW1	0.035	1.481	0.012	2.086	-0.012	-1.698	0.002	0.318	-0.016	-3.558
LW3	0.063	1.654	0.022	2.394	-0.032	-2.777	-0.007	-0.544	-0.062	-8.310
LW6	0.113	2.381	0.019	1.683	-0.071	-4.992	-0.021	-1.361	-0.132	-14.307

Panel B: FF factors plus Short-term reversal

	α (%/month)		market (β)		SMB (γ)		HML (δ)		Sh-t revrsl (η)	
	coeff	tstat	coeff	tstat	coeff	tstat	coeff	tstat	coeff	tstat
BV	0.176	2.996	0.055	3.821	0.071	3.965	0.301	15.501	0.092	5.817
SL	0.079	0.967	0.039	1.941	0.064	2.561	0.423	15.593	0.100	4.504
CF	0.094	1.314	0.059	3.349	-0.038	-1.752	0.358	15.187	0.101	5.236
EW	0.226	1.981	0.025	0.903	0.763	21.979	0.167	4.437	0.142	4.618
LW1	0.014	0.812	0.001	0.349	-0.014	-2.612	0.003	0.457	0.067	14.249
LW3	-0.004	-0.121	0.020	2.389	-0.037	-3.526	0.006	0.544	0.107	11.658
LW6	-0.022	-0.386	0.042	3.017	-0.079	-4.547	0.016	0.834	0.119	7.702

Panel C: FF factors plus a dot.com dummy

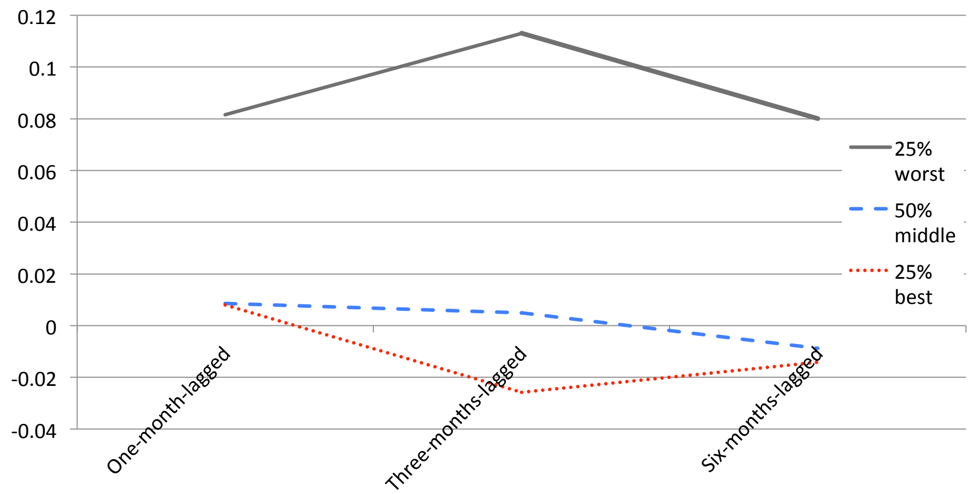
	α (%/month)		market (β)		SMB (γ)		HML (δ)		dot.com (ζ)	
	coeff	tstat	coeff	tstat	coeff	tstat	coeff	tstat	coeff	tstat
BV	0.083	1.270	0.086	5.971	0.066	3.578	0.290	14.198	0.756	4.188
SL	-0.022	-0.247	0.072	3.653	0.059	2.307	0.412	14.582	0.826	3.309
CF	-0.014	-0.179	0.093	5.359	-0.044	-1.951	0.346	14.040	0.871	4.012
EW	0.177	1.384	0.066	2.348	0.761	20.987	0.168	4.202	0.454	1.285
LW1	-0.013	-0.519	0.021	3.801	-0.015	-2.080	0.002	0.317	0.244	3.525
LW3	-0.047	-1.034	0.051	5.119	-0.038	-2.956	0.006	0.424	0.387	3.080
LW6	-0.056	-0.810	0.076	5.025	-0.080	-4.107	0.018	0.844	0.325	1.707

Key Panel A of the table shows Carhart-style (1997) performance analysis regressions, using percentage monthly returns. The left-hand-side variable is the return in excess of the VW return for Fundamental Indexing strategies with weights based on book value (BV), sales (SL) and the (absolute value of) the free cash flow CF). Also shown are results for portfolios where weights are equal (the 1/N rule), or set equal to the value weight as observed 1, 3 or 6 months ago. In Panals B and C, Momentum is replaced by short-term reversal and a dot.com dummy (March 2000 till October 2002, respectively. T-stats are Newey-West-corrected.

containing the same stocks (the same performance class).

Figure 1 displays, for each of these performance classes, the extra average monthly returns realized by using lagged weights rather than contemporaneous ones. That is, we calculate, for every observation j, t, the number $(w_{j,t-L} - w_{j,t-1})r_{j,t}$ and provide averages per performance class. Unsurprisingly, there appears to be no extra return when investing in middle-of-the-road performers. For winners, investing on the basis of 3- or 6-month lagged weights yields lower returns than investing on the basis of contemporaneous weights. This is consistent with

Figure 1: Average extra monthly returns from using lagged weights instead of value weights for three past-performance classes of stocks



Key At the start of each month t stocks are allocated to 3 momentum buckets. Bucket 1, 2 and 3 contain respectively the 25% worst, 50% middle and 25% best past performing stocks. The past performance is measured as the average return for the months $t-7$ until $t-2$. For each performance bucket, monthly index returns are calculated based on first-of-the-month value weights lagged 1, 3 or 6-months. Average monthly returns are in excess of the first-of-the-month value-weighted index within each performance category. The top schedule shows mean extra returns for the losers, using either one-, two- or three-month lagged weights; the middle-of-the-road performers's extra returns are below, while the lowest schedule refers to winners.

momentum: winners go on winning, but by using lagged weights we assign too little capital to those stocks. The eye-opener, however, is the substantial extra return from using lagged weights in the case of losers. This suggests strong reversal rather than momentum: when using lagged weights for losers, we overinvest in them, by current-cap standards, which turns out to pay off handsomely when these stocks recover. Reversal may to some extent reflect a single outlier event, like the rebound after the 2008 crash, but we will show that at least part is also due to an ongoing fallen-angel effect.[8] But whatever the explanation, the reversal identified by the regressions now seems to stem from losers that rebound, not from winners whose prices fall back.[9]

[8]The fallen-angel effect is documented by e.g. Chan and Chen (1991), Chen and Zhang (1998), Lakonishok and Vermaelen (1990), Ikenberry, Lakonishok and Vermaelen (2002) or Peyer and Vermaelen (2009). Bearing in mind their own resumé, the story goes, professional portfolio managers are reluctant to invest in stocks that did quite badly recently, so prices are depressed and expected returns high.

[9]The asymmetry between winners and losers also suggests different correction speeds. Winners seem to be either initially undervalued firms whose prices slowly correct, or firms that gradually get more overvalued and whose eventual price correction takes place beyond the our six-month horizon. Losers, in contrast, are much more likely to be briefly undervalued. With a faster correction, chances of capturing that correction within our horizon are better; and even if all corrections do happen inside the horizon, a sudden correction produces a higher mean monthly return than a gradual one. For instance, if a price goes from 50 to 100 in five equal steps, the average per period return is 0.15. In contrast, if it jumps from 50 to 100 in period 1 and then pays four

The finding of strong reversal prompted us to replace, in the performance analysis regressions, the momentum factor by French's short-term reversal factor, STR. Panel B of Table 2 shows the new results. We find clear evidence of positive extra exposure to STR in all strategies, and most clearly so in the LW portfolios. In fact, the t-ratios for STR are all bigger, in absolute terms, than those for momentum. In addition, all alphas except BV's are now insignificant. So it seems that the performance analysis results are not robust with respect to the choice of the factors, even if the versions are as closely related as momentum and reversal.

There may be other problems with the standard regressions. Familiarly, exposures are assumed to be constant. Variability over time is especially problematic if it is correlated with its price of risk: then the cross-product gets mistaken for genuine alpha. In our analysis of the weight adjustments (Section 2.3) we do find that FI's size mix varies substantially over time, with the end of the dot.com period being especially atypical. That period also had unusual SMB returns. When we accordingly add a dot.com dummy (March 2000-October 2002)[10] as a crude way to separate this unusual period, see Panel C of Table 2, all alphas become insignificant.

Another illustration of changing exposures is provided in Table 3, which shows t-tests for the parameters in a Carhart regression augmented with interactions between the market and the four factors, *à la* Henriksson-Merton (1981). The regressee is still the extra return for each FI strategy relative to VW, and a significant t-ratio signals that changes in the differential beta are correlated with unusual returns in the factor. For all three FI strategies, β is significantly down when small stocks do well. For BV, β is down also when value stocks pay off more, and maybe even when momentum is high ($p = 0.086$, two-sided). The test is exploratory only, as the other exposures are still treated as constant, but for our purpose the results suffice to reject a regular Carthart test as unreliable in this data set.

This is not a comfortable situation. To FI supporters, the above tests may look like unscientifically experimenting with alternative models until results do conform to orthodoxy's priors. For this reason we now propose alternative ways to find out whether FI's extra returns largely reflect drag avoided.

---

zero returns, the average return is 0.20.

[10] The dot-com bubble burst on Friday, March 10, 2000. The technology heavy NASDAQ Composite index had just peaked at 5,048.62 (intra-day peak 5,132.52), more than double its value just a year before. USD 5 trillion in market cap disappeared from March 2000 to October 2002.

Table 3: **T-tests for Carhart regressors and their interactions with the market factor**

	a	MKT	interactions of market with				SMB	HML	MOM
			MKT	HML	SMB	MOM			
BV	4.819	1.298	1.406	-2.221	-4.886	-1.713	6.219	14.527	-11.699
SL	2.826	-0.658	1.146	-0.357	-4.176	-0.882	4.483	14.705	-10.073
CF	2.988	1.128	1.089	-0.048	-4.440	-1.544	-0.260	13.904	-8.885

Key For each of the FI return series we run a Carhart regression augmented by interactions with the market. The table shows the t-ratios.

Table 4: **Notation**

$V_{j,t}$	the observed market value of stock j at time t
$v_{j,t}$	the unobservable true market value of stock j at time t
$\epsilon_{j,t}$	the percentage pricing error: $V_{j,t} = v_{j,t}(1 + \epsilon_{j,t})$
$r_{j,t}$	the true return, i.e. $v_{j,t}/v_{j,t-1} - 1$
$R_{j,t}$	the noisy observed return, i.e. $R_{j,t} = r_{j,t}(1 + \epsilon_{j,t})/(1 + \epsilon_{j,t-1})$
$W_{j,t}$	the observed market weight of stock j at time t
$w_{j,t}$	the unobservable true value weight of stock j at time t
$\epsilon_{m,t}$	the market-wide pricing error: $W_{j,t} = w_{j,t}\frac{1+\epsilon_{j,t}}{1+\epsilon_{m,t}}$
ρ_j	the autocorrelation in $\epsilon_{j,t}$
γ_j	the cross-correlation between $\epsilon_{j,t}$ and $\epsilon_{m,t-1}$
$w^*_{j,t}$	a proposed alternative weight of stock j at time t
$\epsilon^*_{j,t}$	the error in the alternative weight relative to the true weight: $w^*_{j,t} = v_{j,t}(1 + \epsilon^*_{j,t})$

2 Two preliminary plausibility checks

Arnott and Markowitz (2008) provide the mathematical background. The appendix presents their logic, adding a modestly more transparent way to identify the drag. The first subsection, below, just reviews the key elements. Two tests then follow. The notation is summarized in Table 4.

2.1 FI's math in a nutshell

There are three main points. First, if prices are noisy and mispricing is unrelated to true value v or true return r, the expectation of the noisy observed return R is biased upward by a Jensen's convexity effect (Brennan and Wang, 2010):

$$1 + R_{j,t} = (1 + r_{j,t})\frac{1 + \epsilon_{j,t}}{1 + \epsilon_{j,t-1}}, \tag{1}$$

$$\Rightarrow \mathrm{E}(R_{j,t}) \approx \mathrm{E}(r_{j,t}) + \mathrm{var}(\epsilon)(1 - \rho_j). \tag{2}$$

Second, mispricing also drives a wedge between observed value weights W and the true ones, w. If the errors are not diversified away at the market level, also the market-wide average mispricing enters into the picture:

$$W_{j,t-1} = w_{j,t-1} \frac{1 + e_{j,t-1}}{1 + \epsilon_{m,t-1}}. \tag{3}$$

Value weighting then means that the upward bias from $\mathrm{E}(1 + \epsilon_{j,t-1})^{-1}$ disappears — this loss is FI's 'drag' — only to be replaced by a similar but presumably smaller market-wide valuation error:

$$\begin{aligned} (1 + R_{j,t}) W_{j,t-1} &= (1 + r_{j,t})\, w_{j,t-1} \frac{1 + \epsilon_{j,t}}{1 + \epsilon_{m,t-1}}, & (4) \\ \Rightarrow \mathrm{E}[(1 + R_j)(w_{j,t-1} - W_{j,t-1})] &= \mathrm{E}\left((1 + r_{j,t}) w_{j,t-1} \left[\frac{1 + \epsilon_{j,t}}{1 + \epsilon_{j,t-1}} - \frac{1 + \epsilon_{j,t}}{1 + \epsilon_{m,t-1}} \right] \right); \\ \frac{\mathrm{E}[(1 + R_j)(w_{j,t-1} - W_{j,t-1})]}{w_{j,t-1}} &\approx \mathrm{E}(1 + r_{j,t}) \left[\mathrm{var}(\epsilon_j)(1 - \rho_j) - \mathrm{var}(\epsilon_m)(1 - \gamma_j) \right]. & (5) \end{aligned}$$

The above shows that if one could observe the noise-free weights w, one would gain and get back the original expected return, Equation (1). The third result is that, if one works with an alternative weight w^* whose error ϵ^* is uncorrelated with the valuation error, in terms of expectations this is as good as the true weights.

From the above, the drag is driven by the variance of the percentage mispricing and its autocorrelation, in excess of the market-wide valuation error variance and its cross-correlation. Most of the FI literature ignores the marker-wide valuation errors, and so do some of the tests below. (This is conservative, here: we find that the gross drag is economically meaningless, so that result holds *a fortiori* if we would have subtracted the market-related part.)

In this section the question of interest is whether it is plausible that most of FI's extra return stems just from drag avoidance and, for this to be so, what properties of noise would be necessary. First we study the time series of returns, computing autocovariances to obtain rough estimates of noise variances (Roll, 1984) that can be compared to FI's extra return and provide a first plausibility check. Then we look at the cross-sections of weight adjustments administered by FI, and test whether it is plausible that they just introduce an alternative noise without creating style shifts.

2.2 A lower bound on noise a generalized Roll (1984) test

How much potential return are we talking about? From Section 2.1, drag amounts to at most $\mathrm{var}(\epsilon_j)(1 - \rho_j)$, the unconditional noise variance corrected for autocorrelation. (The actual amount of recoverable drag may be lower if there tend to be market-wide valuation errors.)

This term has been studied by Brennan and Wang (2010).[11] For the CRSP population they estimate the median and mean of the standard deviations for mispricing to be 4 and 6 percent, and ρ 0.33 or less. So the gross effect of the noise on mean returns could be $0.06^2 = 0.0036$ or 36 bp for average returns, and less than 0.0024 after correction for autoccorrelation. This is just half of our 45bp extra return from EW. For FI strategies, the unexplained gap must be even bigger. The reason is that FI strategies put most of their cash into the biggest and safest stocks, for which the Brennan-Wang's noise estimates are much lower (0.01, as opposed to 0.15 for idiosyncratic stocks). If the weighted average drag, following Brennan and Wang, is, say, $0.02^2 = 0.0004$ or 4bp instead of $0.06^2 = 0.0036$ or 36 bp, the extra returns we saw are even harder to explain as drag.

Instead of starting from an external estimate of the noise, we can rely on a Roll (1984)-type analysis of autocorrelations. Ignoring third and higher moments and adopting Roll's assumptions except for autocorrelation in the noise, the autocovariance remains proportional to the noise variance:

$$\begin{aligned}
\mathrm{cov}(R_{j,t}, R_{j,t-1}) &\approx \mathrm{cov}(r_{j,t} + \epsilon_{j,t} - \epsilon_{j,t-1}..., r_{j,t-1} + \epsilon_{j,t-1} - \epsilon_{j,t-2}...) \\
&= \mathrm{cov}(\epsilon_{j,t}, \epsilon_{j,t-1}) - \mathrm{cov}(\epsilon_{j,t}, \epsilon_{j,t-2}) - \mathrm{var}(\epsilon_{j,t-1}) + \mathrm{cov}(\epsilon_{j,t-1}, \epsilon_{j,t-2}) \\
&= \mathrm{var}(\epsilon)(-1 + 2\rho_j - \rho_j^2) \\
&= -\mathrm{var}(\epsilon)(1 - \rho_j)^2, \qquad (6)
\end{aligned}$$

This depends on the same parameters as the gross drag. So if we would obtain sufficiently reliable estimates of the autocovariances, we can combine these numbers with the extra returns from each investment strategy to obtain lower bounds for the noise and the autocorrelation that would be compatible with the absence of style shifts. Denoting the extra return from the alternative investment strategy by Δ_j, we have two relations, and two implications:

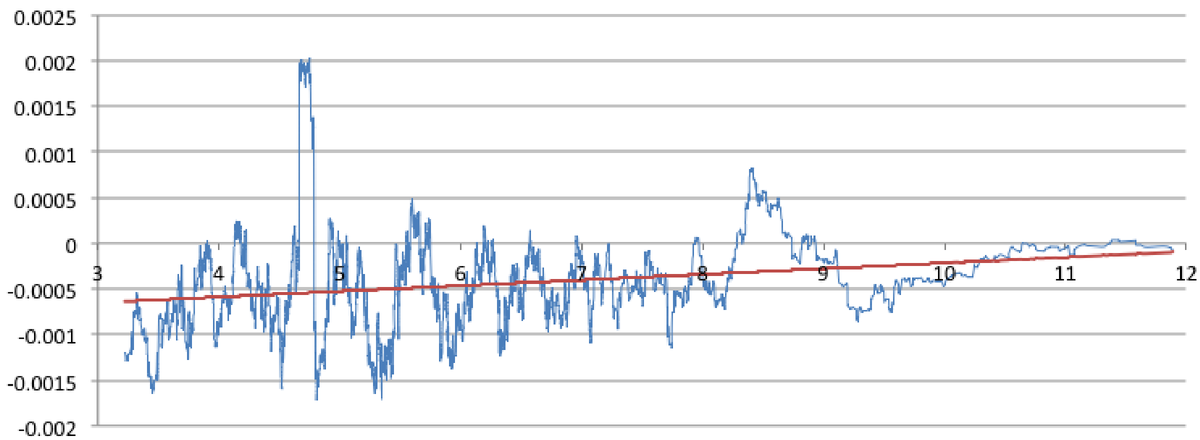
$$\left.\begin{aligned}
\sigma^2(1-\rho_j) &\geq \Delta_j, \\
\sigma^2(1-\rho_j)^2 &= -\mathrm{cov}(R_{j,t}, R_{j,t-1}),
\end{aligned}\right\} \Rightarrow \left\{\begin{aligned}
\sigma_j &\geq \frac{\Delta_j}{\sqrt{-\mathrm{cov}(R_{j,t},R_{j,t-1})}}, \\
\rho_j &\geq 1 + \frac{\mathrm{cov}(R_{j,t},R_{j,t-1})}{\Delta}.
\end{aligned}\right. \qquad (7)$$

The Δ_j-related bound is an inequality because market-wide noise, if any, would lower the expected return below the gross drag $\sigma_j(1-\rho_j)$). The bound also assumes no style shift; this is in fact the working hypothesis that we want to reject.

We take a sample of 2000-2009 returns, retaining all ca 4200 stocks with at least 56 observations out of a potential 112, and estimate autocovariances. The average of our stock-by-stock

[11] Brennan and Wang study mispricing by sorting stocks on the basis of idiosyncrasy (the variance of Fama-French residual noise), and analyzing the unexplained returns via a Kalman filter.

Figure 2: Autocovariances in stock returns plotted against log market cap



Key We take a sample 2000-2009, retaining all 4200 stocks with at least 56 observations out of a potential 112, and estimate autocovariances. These are plotted against the log of the average market cap, after smoothing. The irregular path is obtained as follows. We rank the stocks by average size, compute the mean autocovariance for the 100 smallest stocks, and plot this point against the log average market cap for stock 100. We then move the sample to stocks 2-101, and so on and so forth, until we reach the largest 100. The second way of filtering consists of a simple OLS regression of autocovariances on log market values, the result of which is the line in the figure.

return autocovariances is -0.00047, which does have the right sign. However, many individual estimates are positive, as prior studies have found starting with Roll (1984). In fact, autocovariance estimates appear so noisy that a plot of the estimates against log value becomes utterly uniformative.

But a clearer picture emerges when we filter the numbers. Loosely assuming that similar-sized stocks have similar autocovariances, we first look at averages across 100 similar-sized stocks. Specifically, we rank the stocks by average size, then compute the mean autocovariance for the 100 smallest stocks, and plot this point against the log average market cap for stock 100. We then move the sample to stocks 2-101, and so on and so forth, until we reach the largest 100. The plot of smoothed autocovariances against log market caps shows up as the irregular path in Figure 2. Clearly, the averaging works well at the high-value end, but still leaves us with lots of noise at the small-stock side of the graph. Encouragingly, however, most smoothed values are negative, and there is a clear positive link with log size without any evidence of non-linearity. In light of this, we then run a simple OLS regression of autocovariances on log market values, the result of which is the line in Figure 2.[12] All the fitted values are negative,

[12]The regression is just meant to be exploratory. A more precise estimate may be obtained by accounting for heteroscedasticity. Also, if autocovariance depends on size, then our use of average size implies a large errors-in-the-regressor bias towards zero. Also windsorising might have been useful: the spike at log$V$=4.8 must be quite influential in flattening the line. But if despite of these shortcomings we still find a negative slope, we

Table 5: **Implied noise and autocorrelation assuming no style shifts**

			lower bounds on	
weights	$\overline{\Delta}$	$\overline{\text{cov}(R_t, R_{t-1})}$	σ	ρ
BV	0.00312	-0.000113	0.29	0.96
SL	0.00244	-0.000115	0.23	0.95
CF	0.00228	-0.000103	0.22	0.95
EW	0.00454	-0.000470	0.21	0.90

including those for big corporations; but those for small stocks amount to − 6.4bp while for the top-100 the projected autocovariance amounts to barely −1 bp. We use the fitted values to compute weighted average autocovariances per investment strategy.

Table 5 shows the inferred bounds for the average σ_j and its autocorrelation.[13] The idea that these investment rules are style neutral looks quite implausible unless one accepts very high levels of noise, namely standard deviations of 20-30%, which is about five times the Brennan-Wang estimate. Noise like this would mean a long-run return at least equal to the squared volatility, i.e. 4-9 percent. In monthly returns, the impact of this high noise would then be attenuated because 90 to 96 percent of the mispricing is still present at the end of the month. Also this autocorrelation is very different from Brennan and Wang's, where ρ is at most 0.33 and falls the less idiosyncratic the stock is; their safer stocks even have negative error autocorrelations. The picture here is incompatible with theirs, in short, which suggests that the gains from FI are larger than can be explained by drag avoidance. Our next tests, which consider the cross-sectional behavior of weight adjustments, comes to very similar conclusions.

2.3 The cross-section of weight adjustments

2.3.1 Set-up of the test

In this section we consider cross-sections of the FI-administered weight adjustments and ask the question whether these weight adjustments might indeed be style-neutral and, if not, how stable these style shifts might be. The interesting feature of cross-sections is that, at any point

can be confident that we got the sign right.

[13] To compute these we worked with weighted average extra returns and weighted average covariances, using either BV, SL, CF, and EW weights.The weighted extra returns are available in Table 1. For the autocovariances we ranked stocks by size, every month, and put them into twenty size buckets ('vigintiles'), and computed value weights per vigintile under each of the investment strategies. We then computed average weights over time for every vigintile. These average weights were then combined with the fitted auto covariance read off at every vigintile midpoint in Figure 2.

in time, the aggregate market noise $e_{m,t-1}$ is a constant, so that only the noise in individual prices is active in such a cross-section. It also gives us a figure per period, instead of the unconditional numbers we get from the Roll (1984)-like test. The drawback is that we cannot infer anything about the autocorrelation.

We study the weight adjustment $\ln(w^*/W)$, that is, the log distance between the weights w^* proposed by FI and the original market weights W; and we simply relate them cross-sectionally to size, $\ln V$. For ease of reading we temporarily drop time subscripts: all terms are observed at the same moment anyway. Log percentages are denoted by primes, like in $x' := \ln(1+x)$, and cross-sectional (co)variances, unconditional on any information about the firm, are denoted by $\overline{\text{cov}}$ and $\overline{\text{var}}$. Below, we first write out the weight adjustments, and then look at the covariance with size, bearing in mind that within a given cross-section the market-wide mispricing, if any, is a constant. Then

$$\begin{aligned} \ln(w^*/W) &= [\ln w + \epsilon^{*'}] - [\ln w + \epsilon' - \epsilon_m], \\ &= \epsilon^{*'} - \epsilon' + \epsilon_m; \\ \Rightarrow \overline{\text{cov}}(w^*/W, \ln V) &= \overline{\text{cov}}(\epsilon^{*'} - \epsilon', \ln v + \epsilon'), \\ &= \overline{\text{cov}}(\epsilon^{*'}, \ln v) - \overline{\text{var}}(\epsilon'). \end{aligned} \tag{8}$$

The first term on the right is a measure of size style change, the second is a noise variance unconditional on j, i.e. mixing all ϵ_j together.

At the conceptual level there are two ways to proceed. First, we can extract an upper bound on the noise. This is because the style-change-related covariance on the right is, if anything, negative: small firms get a boost and vv.[14] It follows that

$$\text{if } \overline{\text{var}}(\epsilon') \leq 0 \text{ then } \overline{\text{var}}(\epsilon') \leq -\overline{\text{cov}}(\epsilon^{*'}, \ln v). \tag{9}$$

This differs from the Roll (1984) test not just because it is an upper bound rather than a floor, but also because it can be calculated period by period and, if one finesses the regression, even per size class. These calculations can also be used to test, if that is still necessary, the Null of no style shift: then the bound becomes an estimate, clean of market-wide noise and time-varying, and the plausibility of the estimate reflects on the plausibility of the Null.

The alternative use of our Equation (8) to is use priors about the level of noise, notably Brennan-Wang, and infer something about the first term on the right, the size and time-series

[14] Actually, all we know at this stage is that the covariance with the nominal value is negative. If prices are noisy, however, it is theoretically possible that the covariance with the true values would still be zero. But our tests below defuse this caveat because it would require gigantic levels of noise.

behavior of systematic style shifts related to size.

In terms of estimation, we process each of the 233 cross-sections and show the results as time series plots of parameters. The parameters shown are actually transforms of the covariance. One is its square root, an estimated standard deviation of noise under the Null of no style shift. The other one is the covariance scaled by the variance of size—that is, the slope coefficient in the regression

$$\ln(w_j^*/W_j) = a + b\,(\ln V_j - \overline{\ln V}) + \nu_j. \tag{10}$$

This is easier to interpret than a covariance, and comes with a ready-made significance test.

The regression, after tweaking, also allows us to get some insight into how the noise variance changes across the size spectrum. As it stands, under FI assumptions the above regression just allows us to estimate the cross-sectional variance of the ϵs, which is a mixture of stock-by-stock variances. If one would confine the calculations to a particular size segment, one would get size-class-specific estimates. We can get a similar result by conditioning b on size. If we allow the slope to vary linearly in size, for instance, the total adjustment becomes quadratic in size:

$$\begin{aligned}
\text{If } b(V_j) &= b_0 + b_1\,(\ln V_j - \overline{\ln V})), && (11)\\
\text{then } \ln\frac{\widehat{W}_j}{W_j} &= a + [b_0 + b_1\,(\ln V_j - \overline{\ln V})]\,(lnV_j - \overline{\ln V}), && (12)
\end{aligned}$$

We then back out a time- and size-specific covariance as $\hat{\sigma}_t^2(\epsilon_j) = b_t(V_j)\,\overline{\text{var}}_t(\ln V)$.

2.3.2 Are market values informative about noise? A digression

The regression version of the above test is also interesting as it is related to the question, raised by Perold (2007), to what extent one can correct for noise by just considering the stock's position in the histogram of log market values. Perold (2007) discusses this issue in a Bayesian framework and shows that, under a diffuse prior, the market value is uninformative. Yet Appendix B shows that the above regression, with b set at $\sigma^2(\epsilon_j)/\overline{\text{var}}(\ln V)$, is exactly the correct Bayesian correction in a lognormality model.[15] That is, if FI managers do their job well—and style-neutrally—they would on average correct following a rule like our regression, typically increasing the weight if (and to the extent that) the stock's size is below the average and *vice versa*.

Of course, our claim and Perold's are made in different contexts. Perold's prior is a Bayesian

[15]Note also that this b is exactly the attenuation bias expected if v, the ideal regressor, is replaced by a noisy version V. So outside a joint lognormality framework, the above correction would still be 'best' among all loglinear correction models.

Table 6: Calculated value-correction factors $(\mathrm{E}(v|V)/V-1)*100$ for the U.S. ($\sigma_V = 1.87$)

	% bias, n stdevs from the mean, with n=							
σ_ϵ, %	-4	-3	-2	-1	1	2	3	4
0.01	0.0000	0.0000	0.0000	0.0000	-0.0000	-0.0000	-0.0000	-0.0000
0.06	0.0001	0.0001	0.0000	0.0000	-0.0000	-0.0000	-0.0001	-0.0001
20.00	8.1877	6.2059	4.1813	2.1130	-2.1586	-4.3638	-6.6165	-8.9179
30.00	17.4862	13.4245	9.1629	4.6915	-4.9224	-10.0872	-15.5062	-21.1919

Key In the samples shown in the graph, selected percentile observations are corrected following equation. The table shows values for $\mathrm{EXP}(-(\sigma^2_{\epsilon'_j}/\overline{\mathrm{var}}(\ln V))(lnV_j - \overline{\ln V})) - 1$, in percent.

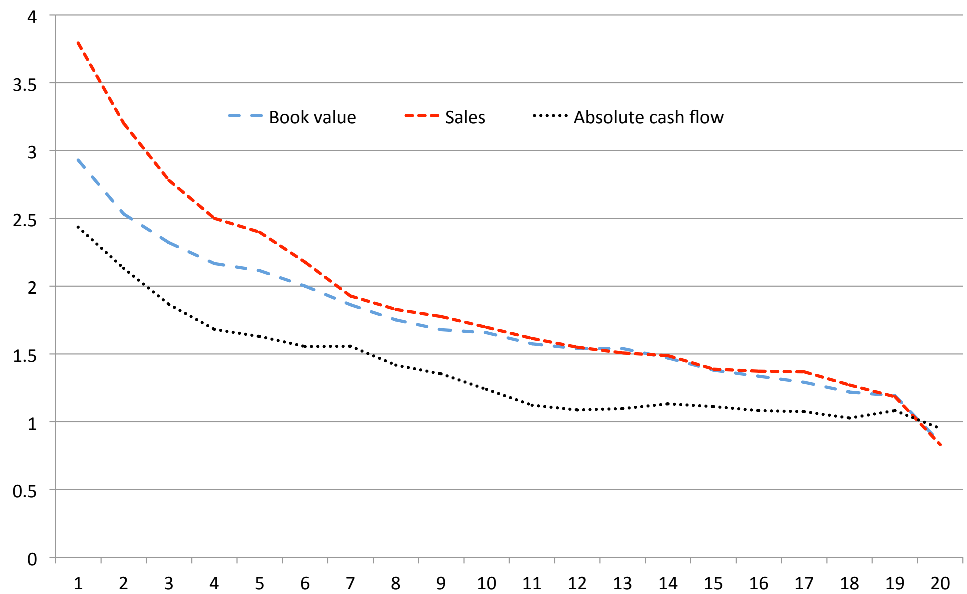
one and refers to two stocks, while here we have in mind the distribution of all true market values. One can easily be agnostic about which of two similar-sized companies really is worth more, while accepting that the entire distribution for thousands of stocks is not uniform. But if the distribution of $\ln v$ were diffuse (uniform on $[-\infty, +\infty]$), then $\ln V$ would have an infinite variance, and b would be zero indeed—Perold's result. And also in actual fact Perold could be approximately right, namely if noise is small relative to the cross-sectional variance of true values.[16]

A simple calculation gives a feel for what adjustments would be reasonable. In the U.S. histogram of $\ln V$ shown in Appendix B, for instance, $\ln V$ has a cross-sectional standard deviation of 1.87, which totally dwarfs the noise variances considered thus far. Table 6 illustrates what kind of Bayesian update factors we get assuming errors standard deviations of 1 and 6% (Brennan and Wang's lower and mean estimates) and 20 and 30% (the calibration bounds computed from the Roll autocorrelations), for prices whose logs are -4 to +4 standard deviations away from the mean. The rational style-neutral price correction is typically less than 20 percent even for the largest noise level and for very small or large stocks. For the Brennan-Wang estimated values it is absolutely tiny.

Contrast this to what actually happens. In the numbers underlying Figure 3, every month we sort stocks on size, and we put each of them into one of twenty size buckets (0-5%, 5-10%, etc) or vigintiles. The Figure shows for every vigintile and FI strategy the ratio of average FI-recommended weight over average market weight. We see that for the smallest stocks the FI weights are, on average, 2.5 to 3.75 times the market weights rather than 1.2 times or less. In fact, even at the 75th size percentile the FI weights are still 1.5 times their market counterparts.

[16]Still, regression towards the mean is nor the key mechanism behind FI. FI's idea is to replace one set of errors by another set, independent of the first. If FI does pull this off, a small amount of regression towards the mean would be observed, but that is not a conscious policy and surely not the main objective. In that sense Perold's objection fo FI is moot.

Figure 3: Average relative fundamental size divided by the average relative market value of the US size buckets



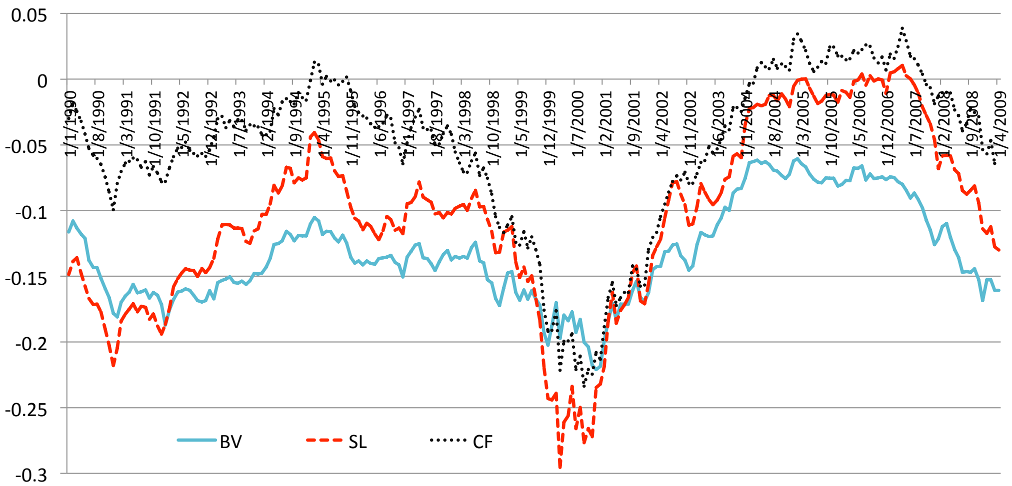
Obviously, bearing in mind the result from the above table, the amount of noise needed to explain such drastic corrections should be vastly larger than the values we just considered. To see how much larger, and how this varies over time, we turn to the regression tests.

2.3.3 Empirical results

We start with the linear regression. Figure 4 shows time-series plots of the 232 month-by-month slope coefficients for each of the weighting schemes. The slopes are very variable over time, with a substantial autocorrelation. The slopes are, in addition, negative over most of our sample; only for cash-flow-based weights the bs turn mildy positive for a non-trivial length of time, namely the intercrisis lull of late 2003 to early 2007. A typical value for b is -0.10 to -0.15.

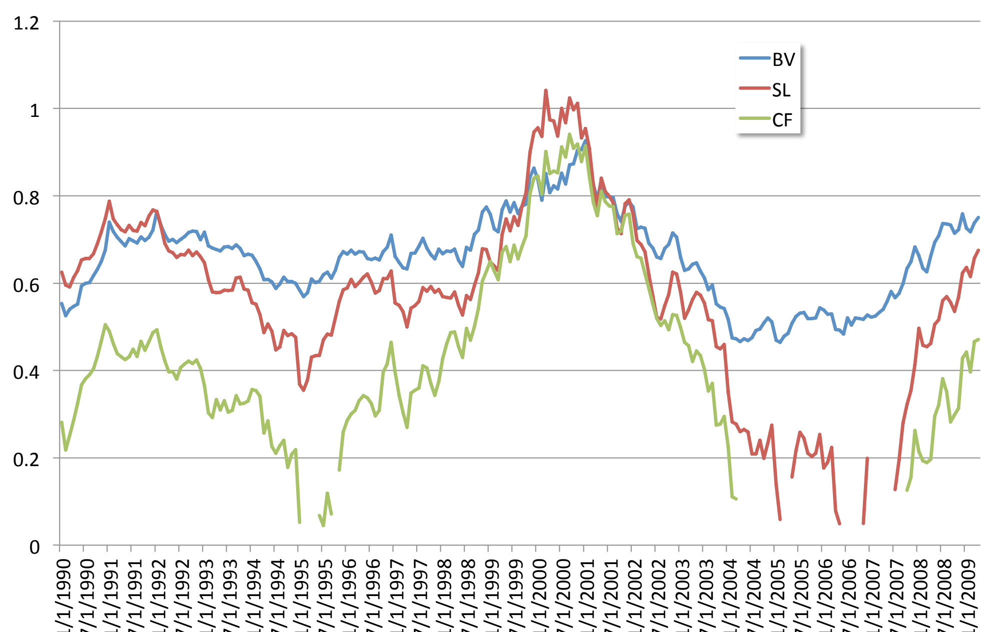
Assuming, initially, there is no style shift, this should be an estimate of $-\overline{\text{var}}(\epsilon')/\overline{\text{var}}(\ln V)$. Figure 5 shows this implied noise as a volatility, $\sqrt{-b\ \overline{\text{var}}(\ln V)}$, for each of the monthly cross sections, using the month's b and cross-sectional variance of $\ln V$. The estimates based on sales or cash flow show the highest variability over time and contain episodes where the procedure breaks down because of a positive b, so that computing a mean for the implied noise would be questionable. But for book value as the FI variable we have a stable and uninterrupted series of negative bs.

Figure 4: Plot of b over time from the regression $\ln \frac{\widehat{W}_j}{W_j} = a + b\,(lnV_j - \overline{\ln V})$



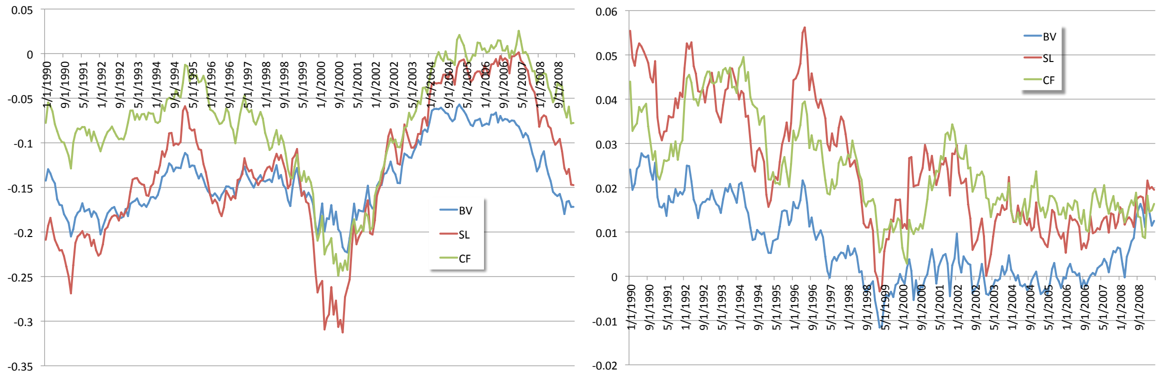
Key In monthly cross-sections, log percentage differences between FI and VW are regressed on log market values. This plot shows, for each of the three FI weights, the time series of slopes. A negative slope means regression towards the center of the distribution of $\ln V$.

Figure 5: Plot of critical noise standard deviation over time implied by the regression $\ln \frac{\widehat{W}_j}{W_j} = a + b\,(lnV_j - \overline{\ln V})$



Key This plot shows, for each of the three FI weights, $\sqrt{-b\,\overline{\mathrm{var}}(\ln V)}$, which is the level of noise that would be needed if FI just avoids drag make the observed degree of shrinking a rational Bayesian response to noise.

Figure 6: Estimated b_0 and b_1 for BV weights in Equation (12)



Key Every month, logs of FI-weights over market weights are regressed cross-sectionally on demeaned log market values, with a slope that is linear in the demeaned log market value: $\ln(\widehat{W}_j/W_j) = a + [b_0 + b_1 (\ln V_j - \overline{\ln V})] (lnV_j - \overline{\ln V})$. The graph shows the time series of estimated b_0 and b_1.

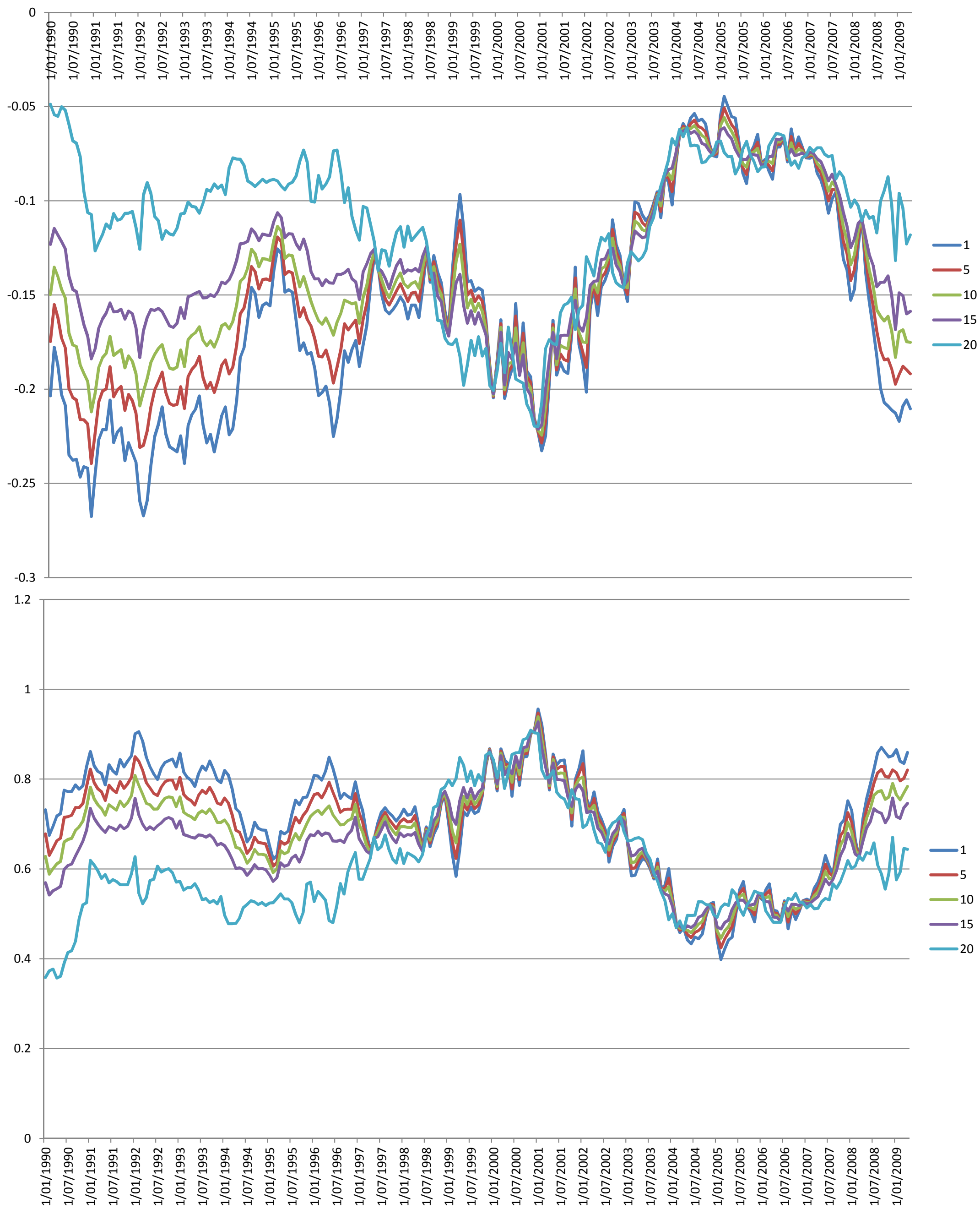
As an upper bound on noise, the numbers pictured in Figure 5 potentially makes sense in a qualitative sense, rising around the 2000/01 bubble-and-crash and the 2008-09 crisis, and uncommonly low in-between these two crisis episodes. But as an actual estimate of noise, (i.e. under the Null of style neutrality) the levels of the required noise, mostly 60-80% but with peaks of up to 100%, are really egregious, making it utterly implausible that the shrinkage entails no style shift.[17] This is 10-12 times the Brennan-Wang estimates.

While taking the cross-sectional covariance $\overline{\text{cov}}(w^*/W, \ln V)$ as an estimate of the noise variance does not make sense, the alternative extreme view works better: $\overline{\text{cov}}(w^*/W, \ln V)$ must be closely reflecting the style shift, $\overline{\text{cov}}(w^*/W, \ln v)$. If we take the Brennan and Wang estimate as our starting point, then the average noise variance is like 0.0036, which is about one percent of the actual average covariance, typically 0.3 to 0.4. Thus, 99% of b (and of its changes) reflects style shift, giving us direct evidence that the style shift must be very ustable.

And this is just the tip of an iceberg, as we discover when we turn to the quadratic model where, it will be recalled, we let b depend on size: $b(V,t) = b_{0,t} + b_{1,t} (\ln V_{j,t} - \overline{\ln V_t})$. Figure 6 shows time series plots of b_0 and b_1 for each of the three FI weighting schemes. As expected, b_0 remains below unity—that is, there is shrinkage for the average stock. More interestingly, most of the estimated b_1s are positive (*i.e.* small stocks are often shrunk towards the mean to a greater extent than average-sized or big ones) but this is not always and everywhere the

[17]The levels are noticeably higher than the ones inferred from the autocorrelations and FI returns. But recall that this new figure is a gross drag rather than a drag in excess of the drag created by market-wide valuation errors. Also, the two tests rely on very different parameters (cross-sectional covariances versus time-series means and autocorrelations).

Figure 7: Total shrinkage factor (top) and noise standard deviation required to make such shrinkage a normal Bayesian update (bottom) for vigintiles 1, 5, 10, 15 and 20, using BV weights in Equation (12)



Key Every month, logs of FI-weights over market weights are regressed cross-sectionally on demeaned log market values, with a slope that is linear in the demeaned log market value: $\ln(\widehat{W_j}/W_j) = a + [b_0 + b_1\,(\ln V_j - \overline{\ln V})]\,(lnV_j - \overline{\ln V})$. The top graph shows the time series of estimated $b(t,V) = b_0 + b_1\,(\ln V_j - \overline{\ln V})$ for the 1st, 5th, 10th,15th and 20th vigintile firm. The bottom graph shows, for each of the three FI weights, $\sqrt{-b(t,V)\,\overline{\mathrm{var}}(\ln V)}$, which is the level of noise that would make the observed degree of shrinking a rational Bayesian response to noise.

case. There is, in fact, a marked downward trend in all b_1s during all of the 1990s, and the one for BV actually falls to about zero in 1997; b_1 even dips into negative values around the time of the dot.com bubble (1999). Afterwards b_1 floats around zero (meaning that weight adjustments are roughly linear in log size), only to re-emerge when the 2008- crisis starts unfolding. In short, then, there must have been substantial style shifts, unstable both across the size spectrum and over time. This confirms our earlier questions about the FF regressions, notably their assumedly constant exposures to all risks.

We now turn to alternative direct estimates of the benefit from avoiding drag.

3 Direct estimates of the likely gains from drag avoidance

If FI is just about drag avoidance, any set of weights whose errors are uncorrelated with the market's will do. In general, the problem is how to make sure the alternative weights do not pick up an unexpected style shift (like reversal, in LW) or, in the case of expected style shifts, how to find out whether a conventional regression-based risk correction is adequate. In this paper we propose to nevertheless use VW, FI, EW and LW—not for the population as a whole, though, but only within size buckets, thus substantially curtailing the room for style shifts.

3.1 Test procedure: mixed portfolio strategies

Our alternative test starts from Treynor's (2005) EW proposal. To avoid the concomitant style shift, we adopt a mixture of VW and EW, as follows. Portfolio capital is allocated to stocks in two steps. At the beginning of every month, we first sort all stocks into 20 equally-populated size buckets (vigintiles). We assign capital to each size bucket as a whole on the basis of the bucket's aggregate market cap, like in VW. Within the bucket, however, we weight equally, Treynor-style. In other words, each company' own value weight is replaced by the average market-value weight of all stocks in the company's size vigintile. The strategy is referred to as the VW/EW mixture.

Studying the VW/EW returns is useful for the following reasons. First, like FI and its competitors, this portfolio rule should avoid drag. In fact, using averaged weights means that the pricing error for an individual stock is replaced by the average pricing error for all stocks in the size bucket; so if a vigintile has N_v stocks, the variance of the stock's pricing error relative to the market's could be reduced by a factor of up to $1/N_v$.[18] Second, this drag reduction is

[18] This analysis is too optimistic: sorting by size V means some bunching of valuation errors ϵ, as argued

achieved with far less scope for style changes than under a regular EW strategy, especially as far as size is concerned. One may object that this procedure just constrains the buckets on market cap, not on *e.g.* book-to-market and momentum. But there is some reassurance from our finding, in the style regressions from Section 1, that an EW portfolio is not exposed to HML. A third advantage is that the results can be studied not just at the total-portfolio level, but also bucket per bucket, giving us some clues as to from what size groups the gains, if any, are coming from.

In addition, we apply variants of the above design: in step 1 we still put stocks into buckets per size vigintile and weight the latter by market cap, but within each of these size buckets we assign, in step 2, weights based on fundamental variables or lagged relative market cap. These mixed styles are referred to as VW/FI, etc, with VW again indicating the weighting across vigintiles and FI, specified as either BV, SL, or CF, the rule for weighting within them. In terms of drag, bucket per bucket, these strategies should not do better or worse than the VW/EW variant; so if they do generate higher returns, we have direct evidence of style shifts, even at the within-bucket level. Thus, we can decompose the extra return of a FI-mixed strategy relative a full VW strategy (denoted VW/VW) into two components:

$$\mathrm{E}(r^{vw/fi} - r^{vw/vw}) = \underbrace{\mathrm{E}(r^{vw/fi} - r^{vw/ew}}_{\text{residual style shift}} + \underbrace{\mathrm{E}(r^{vw/ew} - r^{vw/vw})}_{\text{drag avoided}}, \tag{13}$$

This shows, bucket by bucket, how much comes from drag and how much from within-vigintile style shift. Shifts in exposure to HML, for instance, are especially likely in BV-weighted portfolios, even when overall bucket weights are set by market value: since within each size bucket the market values are rather similar, sorting on book value is close to sorting on book/market, creating HML exposure. If that analysis is valid, the style shift should show up as an extra return above the bucket's EW return.

3.2 Results

Results (1): total and risk-adjusted returns at the portfolio level

Table 7 shows how the mean extra returns (relative to VW) and standard deviations are shifting after mixing the strategy with VW. There is a general drop in mean return and standard deviation, and although the effect unsurprisingly differs across the strategies, almost all mean extra returns relative to VW become insignificant. The exception is the VW/BV

---

in section 2.3; so the average error is not quite as good as the $1/N$ rule would suggest. Recall, however, that $1 + \epsilon_{j,t-1}$ is scaled by $1 + \epsilon_{m,t-1}$; so any market-wide errors are already neutralized.

Table 7: Mean extra return and standard deviations: pure v mixed style portfolios

weighting across vigintiles	weighting scheme within vigintile EW	LW1	LW3	LW3	BV	SL	CF
	extra average monthly returns, %						
based on chosen style (i.e. pure strategy)	1.400	0.100	0.240	0.390	0.310	0.250	0.280
based on market weights (mixed strategy)	0.170	0.090	0.210	0.350	0.220	0.060	0.200
t versus VW	0.960	1.230	0.140	-0.100	2.980	1.430	1.880
	extra stdev of monthly returns, %						
based on chosen style (i.e. pure strategy)	0.310	0.250	0.280	1.400	0.100	0.240	0.390
based on market weights (mixed strategy)	0.220	0.060	0.200	0.170	0.090	0.210	0.350

Key The table shows mean returns in excess of the VW return for the original strategies and their mixed versions, and t-statistics for the latter return. In all the mixed versions, the non-VW weighting rules are applied within vigintiles only; across vigintiles, the weights are value-based. The last two lines show standard deviations in excess of those from VW. All returns are in percent per month.

mixture. Since drag avoidance should be similar across the mixed strategies, the performance of VW/BV relative to VE/EW must be style.

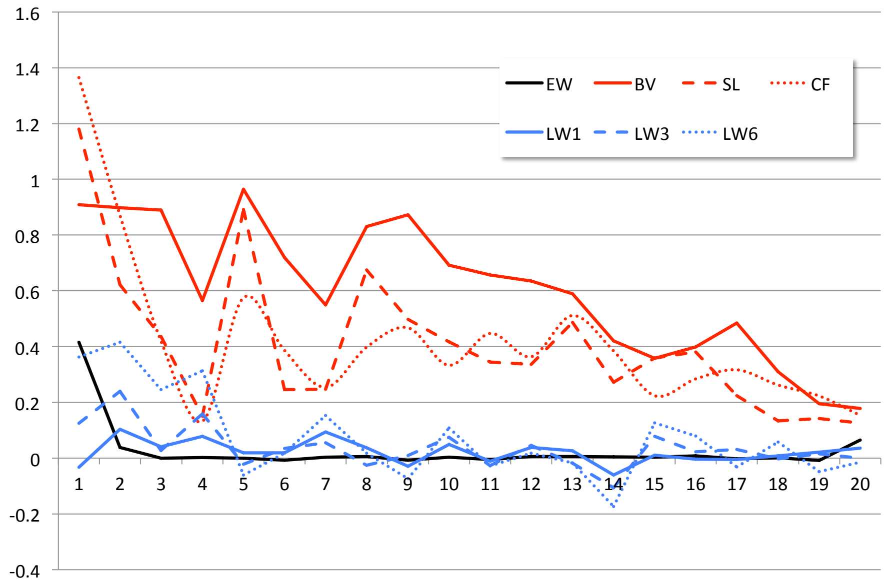
We next turn to the bucket-by-bucket results, starting with EW.

Results (2): performance of the VW/EW mixture, per vigintile

Figure 8 summarizes the performance of each of the mixed weighting schemes via its average monthly return in excess of the bucket's market-value weighted index. For the VW/FI and VW/EW strategies, in addition, we consider the FF alphas for each bucket's return in excess of standard VW, see Table 8 and Figure 9. This subsection just discuses the key strategy, VW/EW.

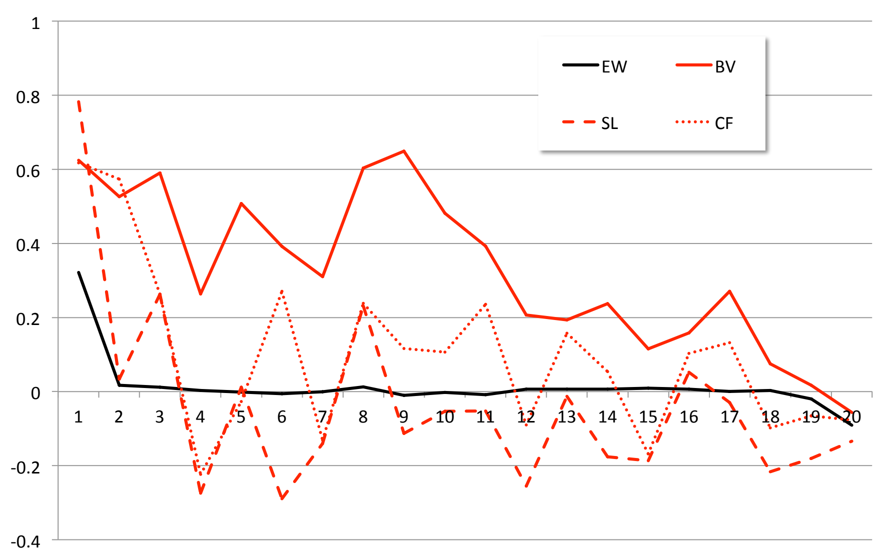
First, the extra returns realized by EW within each vigintile relative to its value-weighted counterpart are essentially zero, at least for size classes 3 to 20, suggesting that drag-avoidance benefits are economically and statistically meaningless. The same conclusion holds for conventionally risk-adjusted returns: for all buckets except the first two, the alphas are economically and statistically insignificant. All this is consistent with net drag being minute, except possibly for the smallest stocks. It also means the FI's return cannot come from drag avoidance: even if the two smallest buckets would do a good job at avoiding drag (about which there are doubts—see below), they represent too little market value to meaningfully affect any implementable portfolio—just look at VW/EW measly 0.0048 percent overall performance, for instance (Table 7). Optimistically, the near-zero drag in most stocks could mean markets are quite efficient; realistically, it could also mean that there is a high persistence and/or a large market-wide component in the pricing errors (market-wide bubbles, for instance, in which case

Figure 8: Extra return from alternative indices, calculated per size bucket (1990-2009)



Key: Each month stocks are sorted per market cap in 20 equally-populated buckets (vigintiles). For every vigintile, monthly returns are calculated using, as intra-bucket weights, either equal weights, or weights based on book value, sales, absolute free cash flow and the one-, three- and six-month lagged market value weights. The graphs show the average excess returns per bucket relative to the VW return for the same bucket.

Figure 9: Alphas from Fama-French regressions per size bucket.



Key Every month, stocks are sorted into size vigintiles, and within that smaller universe a FI investment style is applied. Returns from these FI portfolios in excess of the VW return are then regressed on three FF factors. The figure plots the 20 intercepts (alphas).

Table 8: Alphas from Fama-French regressions per size bucket. The dependent variables are the extra returns on the FI-weighted portfolios relative to VW, calculated per bucket

Size bucket		EW	(t)	SL	(t)	CF	(t)	BV	(t)
(small)	Bucket1	0.420	19.52	1.142	4.151	1.309	4.180	0.893	4.158
	Bucket2	0.040	3.500	0.297	0.759	0.524	1.205	0.752	4.158
	Bucket3	-0.002	-0.238	0.315	1.373	0.320	1.608	0.827	7.302
	Bucket4	0.004	0.531	-0.002	-0.009	0.067	0.381	0.511	4.484
	Bucket5	0.003	0.486	0.717	2.171	0.427	1.526	0.879	5.255
	Bucket6	-0.006	-0.995	0.095	0.456	0.253	1.093	0.699	5.281
	Bucket7	0.003	0.552	0.065	0.346	0.127	0.696	0.486	5.957
	Bucket8	0.008	1.479	0.495	2.741	0.222	1.025	0.765	7.104
	Bucket9	-0.005	-0.878	0.380	2.232	0.359	1.759	0.781	5.900
	Bucket10	0.004	0.646	0.223	1.302	0.189	1.163	0.614	6.899
	Bucket11	-0.003	-0.620	0.225	1.288	0.278	1.415	0.582	5.173
	Bucket12	0.003	0.543	0.206	1.296	0.194	1.272	0.560	5.672
	Bucket13	0.006	1.326	0.339	2.167	0.394	2.026	0.523	3.853
	Bucket14	0.006	0.993	0.125	0.735	0.258	1.749	0.338	3.269
	Bucket15	0.003	0.500	0.254	1.607	0.099	0.769	0.296	3.719
	Bucket16	0.008	1.253	0.291	2.165	0.166	1.413	0.353	4.641
	Bucket17	-0.004	-0.546	0.125	0.943	0.243	2.112	0.425	5.024
	Bucket18	-0.001	-0.076	0.043	0.353	0.186	1.719	0.249	3.080
	Bucket19	-0.022	-1.995	0.082	0.786	0.134	1.560	0.130	1.893
(large)	Bucket20	-0.020	-0.348	0.022	0.280	0.060	0.787	0.064	0.992

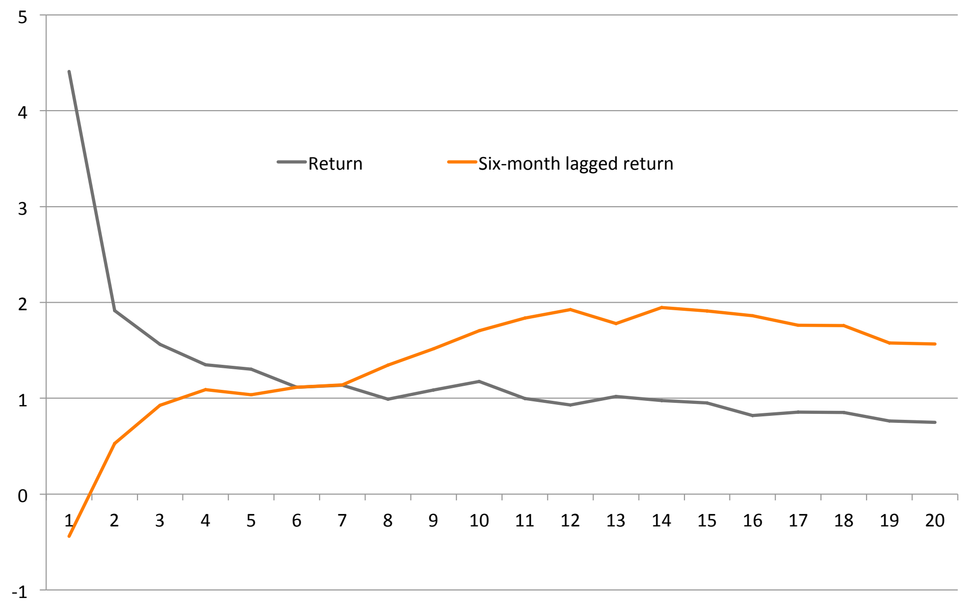
Key Every month, stocks are sorted into size vigintiles, and within that smaller universe a FI investment style is applied. Returns from these FI mini-portfolios in excess of the VW return are then regressed on three FF factors. The intercepts (alpha) are shown, with their Newey-West t; they are in per cent per month.

market weights remain closer to ideal weights anyway).

Let's now consider the smallest stocks, FI's last hope. While the idea that those may be far noisier is quite plausible, it seems that this certainly is not the total story. The extra return for vigintile 1 is over 40 bp. For this to reflect drag, we would need a very negative autocovariance, but in Figure 2 we see numbers like –6bp (on the regression line) or perhaps –10bp (the moving-average plot) rather than –40bp. The only way out, in the noisy-market math, would be a large ρ, as argued before; this would then require gigantic noise levels, and even more so if there is market-wide noise.

But potential sources of return other than drag avoidance are readily at hand. At the very least, the return is to some extent due to the strong reversal phenomenon documented in Section 1. In fact, it turns out that vigintiles 2 and especially 1 contain many fallen angels. In Figure 10 we plot for each of the size buckets (i) the equally-weighted portfolio monthly return (the total return, this time, rather than the extra return relative to a cap-weighted investment) and (ii) the equally-weighted monthly average return over the preceding six months for all stocks

Figure 10: Mean return and mean return over the preceding six months,



Key At the start of each month t stocks are allocated to 20 size buckets. "Return" is the average EW return per bucket. "Six-month lagged return" is a monthly average—not the total (cumulated) return—over the six months $t-2$ to $t-7$, on the stocks that, in month t, will be in vigintile v. This is a hindsight result, not a feasible investment strategy.

that are in the vigintile at t. (Months $t-7$ to $t-2$, to be precise.) The difference between the two curves is, of course, due to migration between deciles: if all stocks always remained in a given bucket, both averages of monthly returns would essentially coincide. Unsurprisingly, stocks below the size median tend to be stocks that did somewhat poorly in the preceding six months. This is especially true for the two smallest-stock vigintiles, whose average lagged return over six months underperforms by about 2 and 1 percent, respectively, i.e. effectively 13 and 6 percent over six months. Some of that value loss is recuperated subsequently: the 5% smallest stocks overperform by over 3 percent, and the second vigintile still by 1 percent.

Returning to Figures 8-9 and Table 8, the diagnosis is that drag is surely not the total answer: the smaller EW bucket portfolios have stumbled upon a strategy that happens to exploit the fallen-angel anomaly. Also, we repeat, the low-cap stocks are too small anyway to have a meaningful impact on any realistic portfolio, so even if there actually is some net drag at the lower end of the size spectrum it does not explain FI's extra returns.

Results (3): performance of the VW/FI mixtures, per vigintile

Returning to Figures 8 and 9, let us consider portfolios weighted on lagged market values. We see that, within a size bucket, LW strategies do not provide any meaningful benefits either, except in the first four vigintiles for lags 3 and 6. This may again be due to drag, to some extent. Actually, this time we do observe that the longer the lag, the higher the return, which is consistent with a noisy-market model where errors (and therefore also drag) take time to disappear. But it is more than plausible that fallen angels helped here, too: relative to VW, LW actively loads on underperformers and *vv.* Regardless, LW's total extra returns at the portfolio level, which are insignificant, also tell us again that the bit of drag that might have been avoided among the small stocks by no means adds anything economically material to the overall performance.

Turning at long last to the VW/FI portfolios, we note a substantially positive extra return relative to their VW/VW and VW/EW benchmarks for each size bucket all the way up to vigintile 17 or 18. The difference relative to VW/EW tells us that there must have been extra sources of return over and above drag avoidance, like style changes. Note, for instance, that the best extra returns are reaped by BV-weighting, which implies a bet on HML. Yet when we correct for FF risks (Table 8), via regression, the alphas for the BV strategy remain significant for all vigintiles but the largest ones. Whether this means that fundamental analysis really adds value (even in the crude way applied by FI), or that the FF model misses something, or that the exposures are too unstable over time (Section 2.3) is hard to say. But we can definitely put the rest to idea that purely avoiding drag is a strategy that pays off well.

4 Conclusion

Building portfolios based on weights from fundamental corporate data instead of weights equal to market cap should increase portfolio returns, and our *prima facie* results are promising, like in many prior studies. FI believers stress that extra returns can be reaped from avoidance of drag, by using weights that are less correlated with mispricing. But also style shifts or market timing can be at work, which would mean that the extra returns are not adding real value.

In our time series, generalized Jensen-FF regressions do not provide reliable measures of pure extra return (alpha), as FI's factor exposures are shown to be unstable. This instability may be one of the reasons why the results also differ so much depending on the factors and the sample period. So in most of the paper we try to establish that there are style shifts, that they are unstable, and that drag itself is a negligible phenomenon, without resorting to style regressions.

First, taking the absence of style shift as out Null, we can obtain lower bounds on the standard deviation and autocorrelation by combining the FI returns with autocovariances. The implied numbers are impossibly high, meaning we reject the null of no style shift.

Second, taking again the absence of style shift as out Null, we test whether FI is shrinking the weights towards the mean by no more than what noisy prices justify. We find massive upward corrections for small stocks, and these could reflect pure drag only if noise were at implausibly high levels. More likely, therefore, the boosting of investments in small stocks is a style shift. We also note that the correction patterns across the weight spectrum are quite variable over time: sometimes small stocks' overweights are al least quadratically related to size, while at other times the link is more linear, or disappears, or even reverts. Hence the shifts in exposure.

Third, we introduce mixed strategies, focusing primarily on 'VW/EW', i.e. value weighting across size vigintiles and equal weighting within vigintiles. This should still avoid drag and largely sidestep any style shifts. The extra return is economically and statistically insignificant. Even FI-based weights within vigintiles no longer pay off (that is, style shifts within vigintiles are much diluted) except for book-value-based investing. These negative results also fit in with those from lagged weights à la Basset and Chen, where no extra return is found even though these strategies should avoid drag just the same as VW/EW and FI or VW/FI. For the smallest stocks there seems to be some extra return but these markets are too small to mean much, in practice, and most of the returns seem to come from reversal. In short, despite its intellectual plausibility, the idea that drag is something to worry about in practice should be put to rest.

Appendix A: FI and portfolio performance: reviewing the math

Let $V_{j,t}$ denote the price of asset j observed at time t. In a Roll (1984) noisy-markets model, $V_{j,t}$ is a combination of noise ϵ_t overimposed on the true price $v_{j,t}$. The noise could be additive or, more plausibly, multiplicative. In the latter case, the relation is

$$V_{j,t} = v_{j,t}(1+\epsilon_{j,t}) \text{ where } \mathrm{E}(\epsilon_{j,t}|v_{j,t}) = 0. \tag{14}$$

One familiar implication is that the return observed between $t-1$ and t, $R_{j,t}$, can be composed into a true return, here denoted as $r_{j,t}$, and the noise or pricing error, which we allow to exhibit autocorrelation:

$$\begin{aligned} 1+R_{j,t} := \frac{V_{j,t}}{V_{j,t-1}} &= \frac{v_{j,t}}{v_{j,t-1}}\frac{1+\epsilon_{j,t}}{1+\epsilon_{j,t-1}}, \\ &=: (1+r_{j,t})\frac{1+\epsilon_{j,t}}{1+\epsilon_{j,t-1}}, \end{aligned} \tag{15}$$

with

$$\epsilon_{j,t} = \rho_j \epsilon_{j,t-1} + \nu_{j,t-1}. \tag{16}$$

The ratio of the two errors terms is mean-reverting; that is, we have $0 \leq \rho_j < 1$. A familiar implication of this mean-reversion is that undervalued stocks tend to overperform and *vv.* Note that the net effect on the expected one-period return is positive, as a result of the convexity of the above return in $\epsilon_{j,t-1}$ and Jensen's Inequality. Adopting a quadratic approximation, $1/(1+\epsilon) \approx 1-\epsilon+\epsilon^2$, and ignoring third moments like $\mathrm{E}(\epsilon_{j,t}\epsilon^2_{j,t-1})$, we can assess the likely order of magnitude:

$$\begin{aligned} 1+R_{j,t} &\approx (1+r_{j,t})(1+\epsilon_{j,t}-\epsilon_{j,t-1}+\epsilon^2_{j,t-1}-\epsilon_{j,t}\epsilon_{j,t-1}+\epsilon_{j,t}\epsilon^2_{j,t-1}), \\ \Rightarrow \mathrm{E}(1+R_{j,t}) &\approx [1+\mathrm{E}(r_{j,t})][1+\mathrm{var}(\epsilon)(1-\rho_j)]. \\ \mathrm{E}(R_{j,t}) &\approx \mathrm{E}(r_{j,t})+\mathrm{var}(\epsilon)(1-\rho_j). \end{aligned} \tag{17}$$

FI adherents point out that the initial mispricing also shows up in the value weight: the observed weight $W_{j,t-1}$ differs from the true (noise-free) weight $w_{j,t-1}$ by the same error term, scaled by the market's value-weighted average error, $\epsilon_{m,t-1} := \sum_j w_{j,t-1}\epsilon_{j,t-1}$:

$$\begin{aligned} W_{j,t-1} &:= \frac{v_{j,t-1}(1+\epsilon_{j,t-1})}{\sum_k v_{k,t-1}(1+\epsilon_{k,t-1})}, \\ &:= \frac{v_{j,t-1}(1+\epsilon_{j,t-1})}{\left(\sum_k v_{k,t-1}\right)\left(\sum_k w_{k,t-1}(1+\epsilon_{k,t-1})\right)}, \\ &=: w_{j,t-1}\frac{1+\epsilon_{j,t-1}}{1+\epsilon_{m,t-1}}. \end{aligned} \tag{18}$$

In qualitative discussions of FI (as opposed to the formal treatment in *e.g.* Hsu, 2009), the market-wide average error gets little attention. The correctness of such a view depends on the underlying view on the individual noise terms. If the latter would be small and independent across stocks and over time, like bid-ask bounce, the market average would have a minuscule variance and no autocorrelation. At the other extreme one could think of market-wide bubbles and busts, where mispricing tends to be persistent, large, and quite similar across stocks, implying a large and slow-to-disappear market-wide error. In-between, one could envisage a scenario where spurious price changes have both market-wide and idiosyncratic components.

Returning to FI, the core argument is that by value weighting a portfolio one creates a systematic interaction between the two effects of initial mispricing. This follows from the expressions for returns and weights, (15) and (18):

$$\begin{aligned}(1+R_{j,t})W_{j,t-1} &= \left[(1+r_{j,t})\frac{1+\epsilon_{j,t}}{1+\epsilon_{j,t-1}}\right]\left[w_{j,t-1}\frac{1+\epsilon_{j,t-1}}{1+\epsilon_{m,t-1}}\right], \\ &= (1+r_{j,t})\,w_{j,t-1}\frac{1+\epsilon_{j,t}}{1+\epsilon_{m,t-1}}. \qquad (19)\end{aligned}$$

In the first line we see that $\epsilon_{j,t-1}$'s effects on return and on weight offset each other: underpriced stocks (with $\epsilon_{j,t-1} < 0$), for instance, have higher expected returns but this mispricing is offset in the weight, and similarly for overpriced stocks. If the market-wide error has trivial variance, the result is that the positive Jensen's Inequality effect of noise on average returns is wiped out entirely. This is the 'drag' that the FI literature refers to. But that is not the end of the story: as the second line shows, the general result is that the individual initial pricing error is being replaced by the market's average error, which brings its own Jensen's-Inequality benefit. In a scenario of strong and pervasive bubbles and busts, for instance, individual and market-wide errors could be so similar that value-weighting hardly lowers returns.[19] Only if the market noise is trivial, is the errors-induced Jensen's Inequality effect totally wiped out.

In general, then, the expected cost of weighting by value is proportional to the expected cost of replacing the individual error by the market-wide error:[20]

$$\begin{aligned}\mathrm{E}[(1+R_j)(w_{j,t-1}-W_{j,t-1})] &= \mathrm{E}\left((1+r_{j,t})w_{j,t-1}(1+\epsilon_{j,t})\left[\frac{1}{1+\epsilon_{j,t-1}}-\frac{1}{1+\epsilon_{m,t-1}}\right]\right); \\ \frac{\mathrm{E}[(1+R_j)(w_{j,t-1}-W_{j,t-1})]}{w_{j,t-1}} &\approx [\mathrm{E}(1+r_{j,t})]\,[\mathrm{var}(\epsilon_j)(1-\rho_j)-\mathrm{var}(\epsilon_m)(1-\gamma_j)], \qquad (20)\end{aligned}$$

where $\gamma_j := \mathrm{cov}(\epsilon_{j,t},\epsilon_{m,t-1})/\mathrm{var}(\epsilon_m)$, the regression coefficient of the stock's error vis-*à*-vis the

[19] If $\epsilon_j \approx \epsilon_m$, Equation (19) is close to a noisy return (15) times a noise-free weight $w_{j,t-1}$.

[20] All expectations are conditional on time $t-1$ information on j's relevant characteristics, including its true value.

lagged market-wide pricing error. So the degree of drag reduction induced by noise at the market level depends on the variance of this initial market-wide noise and its correlation with the final stock-specific error, but it is safe to say that value weighting will usually induce at least some net drag (i.e. miss some potential extra return).

Read differently, Equation (20) shows that the drag would be avoided if we could use the true weights, but the latter are of course unobservable. Still, as FI riposts, we can still reduce the drag (and ideally even avoid it), on average, if we replace the noisy market weights by other proxies for the true weight, here based on 'fundamental' (i.e. accounting-based) variables. There is no claim that the new weights are more precise; in fact, they may even noisier than the market weight. The only requirement is that the alternative weight have a lower covariance with the market's pricing error $\epsilon_{j,t}$. Let's denote the alternative weight by $w^*_{j,t-1} = w_{j,t-1}(1+\epsilon^*_{j,t-1})$ where ϵ^* is the deviation between the FI weight and the true one. Then the extra contribution Δ_t to the portfolio return from using the alternative weights is

$$\begin{aligned} \Delta_t &:= (1+r_{j,t})\frac{1+\epsilon_{j,t}}{1+\epsilon_{j,t-1}}\left[w_{j,t-1}(1+\epsilon^*_{j,t-1}) - w_{j,t-1}\frac{1+\epsilon_{j,t-1}}{1+\epsilon_{m,t-1}}\right], \\ &= (1+r_{j,t})\,w_{j,t-1}\left[\frac{1+\epsilon_{j,t}}{1+\epsilon_{j,t-1}}(1+\epsilon^*_{j,t-1}) - \frac{1+\epsilon_{j,t}}{1+\epsilon_{m,t-1}}\right]. \end{aligned} \tag{21}$$

In an ideal application of FI, the alternative weighting error $\epsilon^*_{j,t-1}$ has zero conditional expectation and is not correlated with either the market's pricing error $\epsilon_{j,t-1}$, the expected true return, or the true weight. Then, on average, all of the net drag induced by value weighting could be avoided. To see this, compute the expected gain per unit of weight,

$$\begin{aligned} \frac{\mathrm{E}(\Delta)}{w_{j,t-1}} &= \mathrm{E}_{t-1}\left((1+r_{j,t})\left[\frac{1+\epsilon_{j,t}}{1+\epsilon_{j,t-1}}(1+\epsilon^*\epsilon^*_{j,t-1}) - \frac{1+\epsilon_{j,t}}{1+\epsilon_{m,t-1}}\right]\right) \\ &\approx \mathrm{E}(1+r_{j,t})\left[\mathrm{var}(\epsilon_j)(1-\rho_j) - \mathrm{var}(\epsilon_m)(1-\gamma_j)\right]. \end{aligned} \tag{22}$$

where, as before, $\gamma_j := \mathrm{cov}(\epsilon_{j,t}, \epsilon_{m,t-1})/\mathrm{var}(\epsilon_m)$. The expression exactly matches the expected drag, per unit of weight, from using value weight instead of the ideal true weights (see Equation (20)).

Appendix B: Bayesian updating of market values

In the Roll (1984) model, mispricing is to some extent identifiable, on average: if $\mathrm{E}(\epsilon_{j,t}|v_{j,t}) = 0$, in general we no longer have a zero expectation for ϵ conditional on V: $\mathrm{E}(\epsilon_{j,t}|V_{j,t}) = \mathrm{E}(\epsilon_{j,t}|v_{j,t}+\epsilon_{j,t}) \neq 0$, so that $\mathrm{E}(v_{j,t}|V_{j,t}) \neq V_{j,t}$. In a group of large stocks, for instance, there should typically be some bunching of overvaluations (positive errors) and *vice versa*.

Perold's (2007) analysis of a related problem[21] shows that it is possible to have independence of ϵ relative to both v and V simultaneously, notably when returns are measured as log changes and when the distribution of $\ln v$ is uniform. He demonstrates this via an example; his appendix shows that the claim is true everywhere on the scale of log market caps if the uniform distribution is on $[-\infty, +\infty]$ (*i.e.* a fully diffuse prior). While diffuse priors are quite respectable for Bayesian problems, our distribution of true values v has different desiderata. Perold just considers two stocks and asks whether observing their market value is helpful if the investors has no idea which is really more valuable. We talk about a distribution of all stocks. For v, the choice of a fully diffuse uniform density for all stocks would would imply that, say, any young upstart might 'really' be worth more than Apple with 50% probability. Additionally, a uniform on $[-\infty, +\infty]$ for all stocks would also be hard to square with the observed distribution of $\ln V$, which exhibits a pronounced mode and asymmetric tails—see the examples for U.S. and Chinese stocks in Figure 11—different from a diffuse variable plus random noise. Thus, while we may very well be ignorant about how to rank a particular pair of stocks, it becomes quite hard to remain utterly uninformed about the entire universe in the sense that the distribution of v across all stocks would be log-uniform on $[-\infty, \infty]$.

We first explain the logic of identifiable errors generated by a non-diffuse v via and example. For ease of exposition and interpretation, in the first example (Table 9, the panel labeled "additive") we work with a symmetric distribution of true prices in a grid $v = 2, 3, ..., 10$. We then add IID noise between –2 and +2 and show in each v-column the expected number of firms for each possible level of V. Tautologically, the table confirms that $\mathrm{E}(V|v) = v$. But if we look at the distribution not column by column (*i.e.* V conditional on v) but, instead, row by row (*i.e.* v conditional in V), we see that $\mathrm{E}(v|V) \neq V$. For instance, a stock priced at $V = 4$ may in reality be an overvalued "3" or an overvalued "5", but in the example the latter scenario is more likely as there are more stocks really worth 5, and likewise for the true values 6 (more probable) and 2 (less). In this example, the value $V = 4$ is below the mode, and, in light of the above, the odds that this stock is underpriced exceed 50%. For observations above the mode, the opposite logic applies. In the example there is one exception: at the mode itself, there is no identifiable error because in that neighborhood the distribution is symmetric: the numbers of firms with a true value one unit above and below the mode are assumedly equal to each other, and likewise for firms with true values two units above and below the mode. The

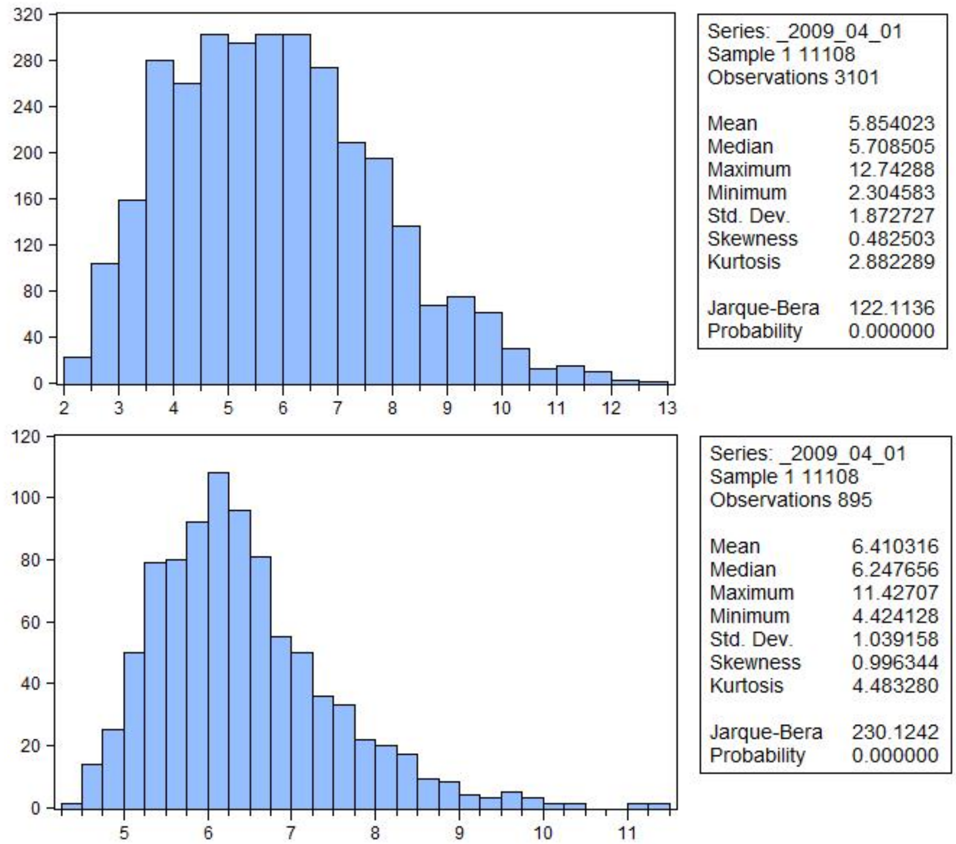
[21] Perold does not really discuss a distribution of true values v, and instead refers to some Bayesian prior about the true value of one stock. But his mathematical arguments do survive a transition from a subjective prior distribution to the distribution of true vs.

Table 9: Attenuation in a noisy-market model: two simple examples

additive										$N_V(V)$	E$(v\|V)$	bias, \$	bias, %
											$N_V(V)$	E$(v\|V)$	bias, \$	bias, %
12									5	5.0	10.00	2.00	0.17
11								10	10	20.0	9.50	1.50	0.14
10							20	20	70	110.0	9.45	0.55	0.05
9						35	40	140	10	225.0	8.56	0.44	0.05
8					40	70	280	20	5	415.0	7.71	0.29	0.04
7				35	80	490	40	10		655.0	6.86	0.14	0.02
6			20	70	560	70	20			740.0	6.00	0.00	0.00
5		10	40	490	80	35				655.0	5.14	-0.14	-0.03
4	5	20	280	70	40					415.0	4.29	-0.29	-0.07
3	10	140	40	35						225.0	3.44	-0.44	-0.15
2	70	20	20							110.0	2.55	-0.55	-0.27
1	10	10								20.0	2.50	-1.50	-1.50
0	5									5.0	2.00	-2.00	$-infty$
	2	3	4	5	6	7	8	9	10				
$N_v(v)$	100	200	400	700	800	700	400	200	100	3600			
multiplicative										$N_V(V)$	E$(v\|V)$	bias, \$	bias, %
14.930									5.9	5.90	10.37	4.56	0.31
12.442								11.8	21.8	33.62	9.76	2.68	0.22
10.368							23.6	43.6	50.0	117.24	9.09	1.28	0.12
8.640						41.3	87.3	100.0	18.2	246.77	7.82	0.82	0.10
7.200					47.2	152.7	200.0	36.4	4.1	440.40	6.70	0.50	0.07
6.000				41.3	174.5	350.0	72.7	8.2		646.78	5.78	0.22	0.04
5.000			23.6	152.7	400.0	127.3	16.4			720.00	5.00	0.00	0.00
4.167		11.8	87.3	350.0	145.5	28.7				623.22	4.32	-0.16	-0.04
3.472	5.9	43.6	200.0	127.3	32.8					409.60	3.73	-0.26	-0.08
2.894	21.8	100.0	72.7	28.7						223.23	3.20	-0.31	-0.11
2.411	50.0	36.4	16.4							102.76	2.75	-0.34	-0.14
2.009	18.2	8.2								26.38	2.56	-0.55	-0.27
1.674	4.1									4.10	2.41	-0.74	-0.44
	2.411	2.894	3.472	4.167	5.000	6.000	7.200	8.640	10.368				
$N_v(v)$	100	200	400	700	800	700	400	200	100	3600			

Key The table shows and example with nine possible prices, using expected numbers of stock instead of probabilities. In the top part the price grid is one dollar, and noise is homoscedastic in dollar terms (additive); in the bottom part the grid is by steps of 20% calibrated on a middle value of 5 dollars, and noise is homoscedastic in percentage terms (multiplicative). The line $N_v(x_i)$ shows the expected numbers of stocks with a true value equal to the column footer value v_i; rescaled by the total number of stocks (3600), this would give us the distribution of true values v. The lines labeled E$(N_V|v = x)$ show the expected numbers of stocks, per level of true value, that end up overpriced, correctly priced, or undervalued. The error has domain e = {-2 ticks, -1 tick, 0, 1 tick, 2 ticks} with probabilities {1/20, 1/10, 7/10, 1/10, 1/20} (additive) or {0.059, 0.218, 0.500, 0.182, 0.041} (multiplicative). This then produces a probability distribution of observed prices, which lastly implies a distribution of V and v given V.

Figure 11: Histograms of log market caps of all US listed stocks (top) and all Shanghai-Shenzen listed stocks (bottom), April 1, 2009



Key Market caps downloaded from Datastream (top) or Wind (bottom). The graphs show a histogram of logs of capitalisations (the latter in millions of USD or CNY) on April 1, 2009.

second part of the table modifies the example towards a multiplicative setting with a geometric grid for the prices, and finds a similar pattern.

For a more formal discussion we adopt the workhorse model for multiplicatives, the joint lognormal for V and v. In one view, these values refer to one particular company j about which the analyst has a prior with mean $\mu_{v_j} := \mathrm{E}(\ln v_j)$ and variance $\sigma^2_{v_j} := \mathrm{var}(\ln v)$; by experience, she knows there is an independent lognormal noise term, resulting in a lognormal V with *ex ante* mean $\mu_{V_j} := \mathrm{E}(\ln V_j)$ and variance $\sigma^2_{V_j} := \mathrm{var}(\ln V)$. Equally, the math below may reflect the econometrician's view who has no data on particular companies, but observes market values and infers a distribution of true values using exogenous distributional properties for the noise. To accommodate both views, we drop j subscripts to the moments. The link between the true and market values is

$$\ln V = \ln v + \epsilon', \tag{23}$$

$$\text{where } \epsilon' := \ln(1+\epsilon) \sim N(-(1/2)\sigma^2_{\epsilon'}, \sigma^2_{\epsilon'}), \tag{24}$$

so that we still have $\mathrm{E}(V|v) = v$. The mean of the conditional distribution of v given V is identified via the regression of $\ln v$ on $\ln V$. Equation (23) immediately implies three properties that are used in the math below: $\mathrm{cov}(\ln V, \ln v) = \sigma^2_v$, $\sigma^2_V = \sigma^2_v + \sigma^2_{\epsilon'}$, and $\mu_V = \mu_v - (1/2)\sigma^2_{\epsilon'}$. Below, we start from the familiar equation linking the expectation of a lognormal variate to the mean and variance of the variate's log. Given joint lognormality, the conditional mean of $\ln v$ given $\ln V$ is described by the regression of $\ln v$ on $\ln V$, and the conditional variance is the error variance:

$$\begin{aligned}
\mathrm{E}(v|V) &= e^{\mathrm{E}(\ln v|V) + 1/2\mathrm{var}(v|V)} \\
&= e^{\mu_v + \beta\,(lnV - \mu_V) + 1/2\sigma^2_{\epsilon'}}, \\
&\quad \text{where } \beta = \frac{\mathrm{cov}(\ln V, \ln v)}{\mathrm{var}(\ln V)} = \frac{\sigma^2_v}{\sigma^2_v + \sigma^2_{\epsilon'}} < 1; \qquad (25) \\
&= e^{\mu_V + \beta\,(lnV - \mu_V)}, \\
&= V\, e^{(\beta-1)\,(lnV-\mu_V)}; \\
&= V e^{-\frac{\sigma^2_{\epsilon'}}{\sigma^2_V}(lnV - \mu_V)}. \qquad (26)
\end{aligned}$$

We conclude that, like in the examples, the correction relative to V is upward for below-average market caps and vice versa, and that the percentage bias is approximately proportional to the deviation of $\ln V$ from its mean. We also note that the adjustment coefficient, $-\frac{\sigma^2_{\epsilon'}}{\sigma^2_V}$ is the same as the scaled covariance used in the Section 2.3, at least in the econometrician's view alluded to above.

References

Amenc, N., F. Goltz and V. Le Sourd, 2008, The Performance of Characteristics-based Indices. *European Financial Management* **15**(2), 241-278

Arnott, R. D., J. Hsu and P. Moore, 2005, Fundamental Indexation, *Financial Analysts Journal* **61**(2), 83-89

Arnott, R. D., J. Hsu, J. Liu and H. M. Markowitz, 2007, Does Noise Create the Size and Value Effects?, working paper, http://ssrn.com/abstract=936272

Arnott, R. D. and H. M. Markowitz, 2008. Fundamentally Flawed Indexing: Comment. *Financial Analysts Journal* **64**(2), 12-14.

Asness, Clifford. 2006. The Value of Fundamental Indexing. *Institutional Investor*. (19 October):67-71.

Basset, G.W., C. Chen and R. Chen. 2007. Fundamental Indexation via Smoothed Cap Weights. *Journal of Banking and Finance* 31(11), 3486-3506

Blitz, D., B. Van der Grient, and P. Van Vliet,, 2010. Fundamental Indexation: Rebalancing Assumptions and Performance. *Journal of Index Investing,* 1(2), 82-88, 2010.

Bogle J C and B G Malkiel (2006), Run on a paradigm, *Wall Street Journal*, June 27 2006, A14

Brennan and Wang. 2010. The Mispricing Return Premium. *Review of Financial Studies* 23(9), 3437-3468.

Bhattacharya U., Galpin N. (2005), Is stock picking declining around the world?, AFA Chicago Meetings Paper, http://ssrn.com/abstract=849627.

Blitz D., Swinkels L.A.P. (2008), Fundamental Indexation: An active value strategy in disguise, *Journal of Asset Management*, 9(4); 264-269.

Blume, Marshall E. and Stambaugh, Robert F. (1983), Biases in computed returns : An application to the size effect, *Journal of Financial Economics*, Elsevier, 12(3); 387-404.

Carhart, M., 1997. On Persistence in Mutual Fund Performance. *Journal of Finance* 52(1), 57-82

Chan, K. C., Chen, N., 1991, Structural and return characteristics of small and large firms, *Journal of Finance* 46, 1467-1484.

Chen, N. and F. Zhang. 1998. Risk and Return of Value Stocks, *The Journal of Business* 71(4), 501-535

Dopfel, F.E., 2008. The arithmetic of Fundamental Indexing. *The Journal of Investing*, 17(2), 114-118

Estrada, J. 2008, Fundamental indexation and international diversification, Journal of Portfolio Management, (Spring), Vol.34, No.3, 93-109

Fama and French, 2007. Nouveau indexes, noise and the nonsense of active management. *Journal of Indexes*, 8

Ferreira, R. and J. D. Krige, 2011. The application of fundamental indexing to the South African equity market for the period 1996 to 2009, *Investment Analysts Journal* 73, 1-12.

Forbes,B. M. and A. K. Basu, 2011. Does Fundamental Indexation Lead to Better Risk Adjusted Returns? New Evidence from Australian Securities Exchange. WP, Queensland UT

Graham, J., 2009. Comment on the Theoretical and Empirical Evidence of Fundamental Indexing. University of California, Davis 2009

Hemminki, J and V. Puttonen, 2008, Fundamental indexation in Europe, *Journal of Asset Managemen*t, Vol.8, No.6 (February), P.401-405

Henriksson, R.D. and R. C. Merton, 1981. On Market Timing and Investment Performance, *Journal of Business* **54**(4), 513-533

Hsu, J.C. (2006), Cap-weighted portfolios are sub-optimal portfolios, *Journal of investment*, 4(3); 1-10.

Ikenberry, D., J. Lakonishok and T. Vermaelen, 2002. Stock Repurchases in Canada: Performance and Strategic Trading. *The Journal of Finance* 55(5), 23732397

Ince, O., S., Porter, R., B., 2006. Individual equity return data from Thomson Datastream: handle with care!. *Journal of Financial Research*, Vol. XXIX, No. 4, 463-479.

Jun, D., B. G. Malkiel, 2008. New Paradigms in Stock Market Indexing. *European Financial Management* 14(1), 118126

Kaplan P.D. (2008), Why Fundamental Indexation Might—or Might Not—Work, *Financial Analysts Journal*, 64(1); 32-39.

Lakonishok, J. and T. Vermaelen, 1990. Anomalous Price Behavior Around Repurchase Tender Offers. *The Journal of Finance* 45(2), 455 - 477

McQuarrie, E. F., 2008. Fundamentally Indexed or Fundamentally Misconceived: Locating the Source of RAFI Outperformance. *The Journal of Investing*, 17(4), 29-37

Mihm, M. and H. Locarek-Junge, 2010. Empirical Examination of Fundamental Indexation in the German Market. In *Advances in data analysis, data handling, and business intelligence. Studies in Classification, Data Analysis, and Knowledge Organization*, Springer, 2010, Part 7, 471-479.

Neukirch, T., 2008. Alternative Indexing with the MSCI World Index. http://ssrn.com/abstract=1106109 or http://dx.doi.org/10.2139/ssrn.1106109.

Perold A. F., 2007, Fundamentally Flawed Indexing, *Financial Analysts Journal* **63(6)**, 31-37.

Peltomki, J., 2010. Nonlinear Exposures of Fundamental Index Returns. *The Journal of Wealth Management* 13(3), 96-106

Peyer, U. and T. Vermaelen, 2009. The Nature and Persistence of Buyback Anomalies. *Review of Financial Studies* 22(4), 1693-1745

Roll R., 1984, A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market, *Journal of Finance* **39 (4)**, 1127-1139.

Siegel , J., 2006. The Noisy Market Hypothesis, *Wall Street Journal*, June 14

Stotz O., Dhnert K., and Wanzenried G., 2007. Do fundamental indexes produce higher risk-adjusted returns than market cap indexes? Evidence for European stock markets. Published online: 16 July 2010, Swiss Society for Financial Market Research 2010

Subramanian, M, P. Kulkarni, R. Kouzmenko, D. Melas, 2011. Capturing the Value Premium. MSCI Barra Research Paper No. 2011-12

Tabner, I.T. (2009), Benchmark concentration: capitalization weights versus equal weights in the FTSE 100 Index, *Multinational finance journal*, 13(1/2); 161-180.

Treynor, J., 2005. Why Market-Valuation-Independent Indexing Works. *Financial Analysts Journal* 61(5), 65-69

Walkshusl, C. and S Lobe, 2010. Fundamental indexing around the world. *Review of Financial Economics*, 19(3), 117127.