An Anatomy of Fundamental Indexing

Lieven De Moor, Fang Liu, Piet Sercu, Tom Vinaimont

DEPARTMENT OF ACCOUNTANCY, FINANCE AND INSURANCE (AFI)

An Anatomy of Fundamental Indexing*

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

First draft: October 2009; this version Monday 6th February, 2012
Preliminary—comments very welcome

*We thank Gaël Imad Eddine, Martina Vandebroek, for useful comments. Any remaining errors are the authors' sole responsibility.

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

[‡]Central University for Finance and Economics, Beijing; fliu@cufe.edu.cn

[§]KU Leuven, Faculty of Business and Economics, and UCLouvain Louvain School of Management; piet.sercu@econ.kuleuven.be

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

Abstract

Adherents 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 a la Roll (1984), it is argued, underpriced stocks overperform but are underweighted and vice versa, implying a ‘drag’ which FI claims to avoid. Mixed into the debate is the question whether mispricing is partly identifiable or not, i.e. whether a policy of actively increasing the small-cap weights and *vv* helps. Carhart style regressions are unable to explain the extra return, but that conclusion is not robust across variant models, and there are substantial doubts about the constantness of factor sensitivities. We investigate the latter issue via cross-sectional regression of weight shifts, and find that not only the weight shifts are much larger than necessary to avoid drag, but the cross-sectional patterns are also quite variable over time. In short, there are style shifts and they are unstable.

To estimate the benefits from drag avoidance, purged of style shifts without having to rely on generalized FF regressions, our procedure is to sort stocks on size into vigintiles, and compare within each vigintile the performance of FI-weighted returns to equally-weighted (EW) returns, which should be immune to drag too without much style shift. We find that within-vigintile EW portfolios are style neutral w.r.t. market and value, and do not meaningfully outperform VW portfolios. 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.

JEL classification: G11, G12, G14

Key words: fund management, drag, noisy prices.

Introduction

The main idea behind Fundamental Indexing (henceforth FI) (Arnold, Hsu and Moore, 2005) is that passive investors should weight stocks on the basis of proxies for market value rather than market value itself, the traditional rule. The argument is that, if market prices are noisy indicators of true value, value-weighting a portfolio creates a ‘drag’ on return: tautologically, one over-weights stocks that happen to be overvalued—the very stocks that tend to underperform—and *vv*. To construct its alternative weights, the FI literature works with book values, cash flows, dividends, and sales as proxies for true value, or a combination thereof. FI’s main claim is that such alternative weights can and do avoid this drag, as long as the new valuation errors are independent of, or at least less correlated with, the errors in market prices. But drag avoidance might not be the sole channel through which extra returns are generated. As sceptics stress, part of the extra return is caused by style shifts: FI puts more weight on small and ‘value’ stocks. Any residual not explained by the above, lastly, would mean that fundamental analysis is simply useful, as analysts traditionally hold, and adds an alpha return.

When the objective is to avoid drag, there are alternatives to FI. Treynor (2005) already notes that equal weighting should dilute away almost all mispricing problems, thus efficiently avoiding any interaction with under- or overpricing. One reply to this proposal is that equal weighting (EW) is infeasible in practice unless it is done on an absolutely tiny scale. Second, EW creates a gigantic style shift. Style shifts are conventionally controlled for by looking at risk-adjusted returns (alphas) rather than at total returns, but it is probably true that FI creates less of a style shift than EW, and is far more implementable since it still invests far less in smaller stocks. A second alternative to FI has been proposed by Basset, Chen and Chen (2007): use lagged weights. This is simpler than FI, and probably largely preserves the style chosen by conventional value weighting (VW) as weights tend not to change drastically over a few months. Using lags also allows a nice diagnostic test: if pricing errors tend to disappear only slowly, then we expect the average extra return from using lagged weights to rise the longer the lag.

All of the proposed alternative weighting schemes do increase performance. Using U.S. data from 1990 until 2009—the data are described in subsection 1.1—we first show in Table 1 the *prima facie* benefits from FI when weights are based on relative Book Value, Sales and the absolute value of Free Cash Flow, respectively.¹ We also show the returns from EW, and

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

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

	Avg monthly return (%)	Standard deviation	Sharpe ratio	Extra return vis-a-vis VW average (%) t-statistic	
VW	0.733	4.59	0.089		
BV	1.045	4.90	0.147	0.312	3,59
SL	0.977	4.84	0.135	0.244	2,02
CF	0.961	4.87	0.131	0.228	2,09
EW	1.187	5.99	0.144	0.454	2.20
LW1	0.760	4.69	0.093	0.026	1.09
LW3	0.753	4.83	0.089	0.019	0.43
LW6	0.740	4.98	0.083	0.007	0.10

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.

from lagged weights (LW1, LW3 and LW6). 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 Absolute Free Cash Flow, respectively. Annualized, this means 2.5-4% extra. In our experiment, this comes with a small increase in risk but, unlike the rise in return, an increased volatility is not a general observation 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 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, work less well than in the Basset *et al.* (2007) test: they barely (and insignificantly) beat VW in terms of mean return and Sharpe ratio.² Actually, against the FI prediction, the extra return falls with the lag length, but since each gain is insignificant, the meaningfulness of this pattern across lags is rather unclear. Anyway, 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.

²There are important design differences. Basset *et al.* focus on the 1000 largest stocks; their sample period is 1962-2003; and they weigh using the 12-month median of the cap-weights rather than weights with fixed-length lags. We prefer three different lags because it allows to test whether the presumed drag drops the longer the lag.

One main objective in this paper in fact is to assess what part of FI's extra returns comes from drag avoidance (FI's official contribution), as opposed to style shift (the critics' view). There are two main findings. First, the amount of shifting of weights towards the mean totally overshoots the degree of shrinkage one expects if the objective just is to avoid drag; in addition, the shrinking waxes and wanes over time and, most of the time, is highly non-linear across the size spectrum. Second, drag indeed is insignificant, possibly because too much of the pricing errors are market-wide (an issue demurely glossed over in most of the FI literature). In passing, we also provide a simpler math treatment; we discover a potential problem with the Basset *et al.* test; and we sort out a theoretical debate between Perold (2007) and the FI team.

The structure in the remaining of the paper is as follows. In Section 1 we provide a standard performance analysis. The results are not very clear, so in Section 2 we have a closer look at the math. This provides the ammunition for an exploratory estimate of the properties of noise (Section 3.1), as well as two tests of the hypothesis that FI just avoids drag: cross-sectional analysis of weight shifts (Section 3.2), and size-controlled comparisons of FI, LE and EW returns (Section 4). Finally, we conclude.

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 begin-of-month market caps from Thomson Reuters Datastream (TRD) for the period January 1990 till May 2009. 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. On 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 \$1.³ Each month we also exclude stocks with negative or zero book value or sales.⁴ 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. This has the advantage that differences in

³This 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 have also limited liquidity, can be subject to high price pressure or price manipulation, and often represent too little value to warrant attention.

⁴Negative 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.

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 found in Panel A of Table 2. In view of the definition of the left-hand variable, 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 shift in the portfolio's sensitivity to the market, somewhat more in its exposure to SMB, and a lot in 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 (no big surprise here), but hardly more to market and HML than the VW portfolio. Using lagged weights affects style far less than does FI, as expected, except for exposure to SMB (LW3 and LW6). We conclude there are style shifts. But the alphas tell us that these increased exposures do not 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 exposures 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

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

Panel A: FF factors plus Momentum										
	constant (α)		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										
	constant (α)		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										
	constant (α)		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, monthly data. 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 Panels 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.

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 n 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. For FI strategies there is a link too, albeit weaker: the FI weights are insensitive to recent changes in market value, so the differences with the VW return are driven by changes in the value weights plus, this time, the difference between the lagged VW weights and the

Figure 1: Average extra 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.

FI ones. This 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 $t - 2$.⁵ For each performance category, we calculate returns for Basset *et al.*-style portfolios, weighted by market caps lagged over 1 month, 3 months and 6 months, and we compare them to a standard VW portfolio 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. 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, which is consistent with momentum: winners go on

⁵We limit the evaluation period to $t - 2$ in order to avoid a common bid-ask bounce affecting both the sorting procedure and the subsequent portfolio return.

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 pays off handsomely when these stocks recover. Reversal may just reflect a single outlier event, the rebound after the 2008 crash. Later we show that at least part is also due to an ongoing fallen-angel effect.⁶ But whatever the explanation, the reversal identified by the regressions now seems to stem from losers that rebound, not from the proverbial ‘technical corrections’ in the case of winners.

The finding of strong reversal prompted us to replace, in the performance analysis regressions, the momentum factor by French’s short-term reversal factor. Panel B of Table 2 shows the new results. We find clear evidence of positive extra exposure to reversal in all strategies, and mostly clearly so in the LW portfolios. In fact, the t-ratios for reversal 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 model, 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.⁷ In our analysis of the weight adjustments (Section 3.2) 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, based on XX, 200Z) as a crude way to separate this unusual period, see Panel C of Table 2, all alphas become insignificant.

This is not a comfortable situation. Performance analysis is supposed to start from a standard model, like Carhart’s, that was validated in the asset pricing literature and found to be adequate. To FI adherents, the above tests may look like unscientifically go fishing for alternative models until results do conform to orthodoxy’s priors. For this reason we now propose alternative ways to answer the questions (i) whether drag is economically important and, if so, whether FI is good at avoiding drag; and (ii) if the return largely reflects style

⁶This 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 résumé, 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.

⁷For example, a simple strategy to beat the market is to increase beta when (risk and) expected market returns are high and vice versa; the positive interaction between beta and risk premium then seems to show up as alpha, at least if the regression is estimating beta as a constant.

shift, how come the generalized FF regressions fail to explain the returns. To do that, we first critically review the math behind FI.

2 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

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

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:

$$\begin{aligned} 1 + R_{j,t} &:= \frac{V_{j,t}}{V_{j,t-1}} = \frac{v_{j,t}}{v_{j,t-1}} \frac{1 + \epsilon_{j,t}}{1 + \epsilon_{j,t-1}}, \\ &=: (1 + r_{j,t}) \frac{1 + \epsilon_{j,t}}{1 + \epsilon_{j,t-1}}, \end{aligned} \quad (2)$$

with

$$\epsilon_{j,t} = \rho_j \epsilon_{j,t-1} + \nu_{j,t-1}. \quad (3)$$

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.⁸ Adopting a quadratic approximation, $1/(1 + \epsilon) \approx 1 - \epsilon + \epsilon^2$, and ignoring third moments like $E(\epsilon_{j,t}\epsilon_{j,t-1}^2)$, we can assess the likely order of magnitude:

$$\begin{aligned} 1 + R_{j,t} &\approx (1 + r_{j,t})(1 + \epsilon_{j,t} - \epsilon_{j,t-1} + \epsilon_{j,t-1}^2 - \epsilon_{j,t}\epsilon_{j,t-1} + \epsilon_{j,t}\epsilon_{j,t-1}^2), \\ \Rightarrow E(1 + R_{j,t}) &\approx [1 + E(r_{j,t})][1 + \text{var}(\epsilon)(1 - \rho_j)], \\ E(R_{j,t}) &\approx E(r_{j,t}) + \text{var}(\epsilon)(1 - \rho_j). \end{aligned} \quad (4)$$

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,

⁸Note to students. Suppose the current market value is 100, the error either +25% or -25%, and the true expected return zero (for simplicity). The stock may be overvalued, in which case its true value is 80 and the expected return (on the price paid, 100) equal to -20%. Equally, the stock may be an undervalued firm really worth 133.33, in which case it should go up, on average, by 33.33%. Therefore the expected return is +6.67% even though the true expected return was assumed to be zero.

scaled by the market's value-weighted average error, $\epsilon_{m,t-1} := \sum_j w_{j,t-1} \epsilon_{j,t-1}$:

$$\begin{aligned}
W_{j,t-1} &:= \frac{v_{j,t-1}(1 + \epsilon_{j,t-1})}{\sum_k v_{k,t-1}(1 + \epsilon_{k,t-1})}, \\
&:= \frac{v_{j,t-1}(1 + \epsilon_{j,t-1})}{(\sum_k v_{k,t-1})(\sum_k w_{k,t-1}(1 + \epsilon_{k,t-1}))}, \\
&=: w_{j,t-1} \frac{1 + \epsilon_{j,t-1}}{1 + \epsilon_{m,t-1}}.
\end{aligned} \tag{5}$$

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 error. In-between, one could envisage a scenario where spurious price changes have both market-wide and idiosyncratic components. Our view is that the purely idiosyncratic-noise scenario is unconvincing, and some of our empirical results bear this out.

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, (2) and (5):

$$\begin{aligned}
(1 + R_{j,t})W_{j,t-1} &= \left[(1 + r_{j,t}) \frac{1 + \epsilon_{j,t}}{1 + \epsilon_{j,t-1}} \right] \left[w_{j,t-1} \frac{1 + \epsilon_{j,t-1}}{1 + \epsilon_{m,t-1}} \right], \\
&= (1 + r_{j,t}) w_{j,t-1} \frac{1 + \epsilon_{j,t}}{1 + \epsilon_{m,t-1}}.
\end{aligned} \tag{6}$$

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.⁹ Only if the market noise is trivial, is the errors-induced Jensen's Inequality effect totally wiped out.

⁹If $\epsilon_j \approx \epsilon_m$, Equation (6) is close to a noisy return (2) times a noise-free weight $w_{j,t-1}$. Unfortunately, in the bubble scenario also autocorrelation is high, undoing some of the effect of high error variance.

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:¹⁰

$$\begin{aligned} \mathbb{E}[(1 + R_j)(w_{j,t-1} - W_{j,t-1})] &= \mathbb{E}\left((1 + r_{j,t})w_{j,t-1}(1 + \epsilon_{j,t})\left[\frac{1}{1 + \epsilon_{j,t-1}} - \frac{1}{1 + \epsilon_{m,t-1}}\right]\right); \\ \frac{\mathbb{E}[(1 + R_j)(w_{j,t-1} - W_{j,t-1})]}{w_{j,t-1}} &\approx \sum_j [\mathbb{E}(1 + r_{j,t})] [\text{var}(\epsilon_j)(1 - \rho_j) - \text{var}(\epsilon_m)(1 - \gamma_j)], \end{aligned} \quad (7)$$

where $\gamma_j := \text{cov}(\epsilon_{j,t}, \epsilon_{m,t-1})/\text{var}(\epsilon_m)$, the regression coefficient of the stock's error vis-à-vis the 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 (7) shows that the drag would be avoided if we could use the true weights, but the latter are of course unobservable. Still, as FI ripostes, 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 be less correlated with the market's pricing error $\epsilon_{j,t}$. Let's denote the alternative weight by $w_{j,t-1}^* = w_{j,t-1}(1 + \phi_{j,t-1})$ where ϕ is the deviation between the FI weight and the true one. Then the extra return Δ_t on a portfolio from using the alternative weights is

$$\begin{aligned} \Delta_t &:= \sum_j (1 + r_{j,t}) \frac{1 + \epsilon_{j,t}}{1 + \epsilon_{j,t-1}} \left[w_{j,t-1}(1 + \phi_{j,t-1}) - w_{j,t-1} \frac{1 + \epsilon_{j,t-1}}{1 + \epsilon_{m,t-1}} \right], \\ &= \sum_j (1 + r_{j,t}) w_{j,t-1} \left[\frac{1 + \epsilon_{j,t}}{1 + \epsilon_{j,t-1}} (1 + \phi_{j,t-1}) - \frac{1 + \epsilon_{j,t}}{1 + \epsilon_{m,t-1}} \right]. \end{aligned} \quad (8)$$

In an ideal application of FI, the alternative weighting error $\phi_{j,t-1}$ 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:

$$\begin{aligned} \frac{\mathbb{E}(\Delta)}{w_{j,t-1}} &= \mathbb{E}_{t-1} \left((1 + r_{j,t}) \left[\frac{1 + \epsilon_{j,t}}{1 + \epsilon_{j,t-1}} (1 + \phi_{j,t-1}) - \frac{1 + \epsilon_{j,t}}{1 + \epsilon_{m,t-1}} \right] \right) \\ &\approx \sum_j [\mathbb{E}(1 + r_{j,t})] [\text{var}(\epsilon_j)(1 - \rho_j) - \text{var}(\epsilon_m)(1 - \gamma_j)]. \end{aligned} \quad (9)$$

where, as before, $\gamma_j := \text{cov}(\epsilon_{j,t}, \epsilon_{m,t-1})/\text{var}(\epsilon_m)$.

¹⁰ All expectations are conditional on time $t - 1$ information on j 's relevant characteristics, including its true value.

3 Two preliminary plausibility checks

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.

3.1 The time-series heavier of returns: Roll's (1984) test

How much potential return are we talking about? The term $\text{var}(\epsilon_j)(1 - \rho_j)$, the gross drag before subtracting the benefit from market-wide errors, has been studied by Brennan and Wang (2010). They estimate the standard deviations for mispricing to be 2-3%, and judge most pricing errors to be in the range $[-8\%, +8\%]$. So the gross effect of the noise on average returns could be 4 to 9 bps, at least if the horizon is long enough to make the end-of-period noise almost independent of the beginning-of-period pricing error (*i.e.* assuming $\rho_j \approx 0$).¹¹

Alternatively, 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:

$$\begin{aligned}
 \text{cov}(R_{j,t}, R_{j,t-1}) &\approx \text{cov}(r_{j,t} + \epsilon_{j,t} - \epsilon_{j,t-1}, r_{j,t-1} + \epsilon_{j,t-1} - \epsilon_{j,t-2}) \\
 &= \text{cov}(\epsilon_{j,t}, \epsilon_{j,t-1}) - \text{cov}(\epsilon_{j,t}, \epsilon_{j,t-2}) - \text{var}(\epsilon_{j,t-1}) + \text{cov}(\epsilon_{j,t-1}, \epsilon_{j,t-2}) \\
 &= \text{var}(\epsilon)(-1 + 2\rho_j - \rho_j^2) \\
 &= -\text{var}(\epsilon)(1 - \rho_j)^2; \\
 \Rightarrow \text{var}(\epsilon)(1 - \rho_j) &= \frac{-\text{cov}(R_{j,t}, R_{j,t-1})}{1 - \rho_j} \geq -\text{cov}(R_{j,t}, R_{j,t-1}),
 \end{aligned} \tag{10}$$

where the last inequality assumes a negative autocorrelation for the observed returns.

We take a sample 2000-2009, retaining all ca 4200 stocks with at least 56 observations out of a potential 112, and estimate autocovariances. When Roll estimated $\text{var}(\epsilon)$ as the negative of the return autocorrelation, he found the latter to be estimated with lots of noise, includ-

¹¹This is not easily annualized offhand. Over shorter horizons, where $\rho_j \neq 0$, the error does not shrink as much, but there is no easy way to estimate the autocorrelation for a particular holding period; and in the compounding of short-term returns one needs to think also of the negative autocorrelation of returns that noise induces.

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.

ing many positive estimates. The same happens in our data. But there are two encouraging features: the mean is negative, and smaller stocks have systematically more negative autocovariances than bigger companies, consistent with more noise for low-caps. We comment on these claims in turn.

The average autocovariance is -0.00047. If $\rho = 0$, this amounts to 4.7 bp/months, which is way below the 20-30 bp delivered by FI and the 45 bp from EW. So we conclude that ρ would have to be rather high, say 0.50 or 0.90: then $\text{var}(\epsilon)(1 - \rho_j)$ could add 9.4 to 47 bp to the return, nicely encompassing the observed extra returns. Actually, the 0.90 number is the more plausible one, as it relates an EW return to an EW autocovariance. A FI-weighted autocovariance would be much closer to zero, and therefore require a much higher ρ than 0.5, as explained below.

The second finding, in fact, is that big stocks have less negative autocovariances. This is consistent with the model combined with the intuition that smaller stocks tend to have noisier prices. To show the negative relation we first resort to a graph. As mentioned, individual autocovariance estimates are very noisy, to the extent that a plot of the estimates against log value becomes utterly uninformative. 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 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 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, 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

For FI, this is bad news, in a way: inevitably, any implementable portfolio is mostly big stocks, and since these have low implied noise, there would be no way for a portfolio to have a weighted average autocovariance as negative as -4.7 bp . This is the reason why, as mentioned, we need an even stronger boost from the $1/(1 - \rho)$ item in Equation (10), meaning ρ would have to be closer to unity than 0.5.

There is more FI-unfriendly news. If our autocovariance of -4.7 bp is the result of an error autocorrelation exceeding 0.9, then the implied standard deviation of the errors becomes quite large. Table 3 addresses the same issue as before, now starting from various possible levels of extra return, assuming all that those gains reflect gross drag, and then extracting error standard deviations for various levels of ρ . If ρ is above 0.9, then to explain all returns from FI as resulting from drag avoidance, the noise SE should be more like 20% percent, way above the Brennan-Wang estimate. The table also illustrates the meaning of these standard errors via a two-sigma confidence bound for v . What we show is the upper bound for a notional stock whose lower bound is standardized at 10 dollar. In terms of that, noise levels of 7-8% do not look implausible at all, but 16-18% (for $\rho = 0.9$) or 22-25 (for $\rho = 0.95$) would mean a confidence interval of $[10, 20]$ or $[10, 30]$, which may become more questionable for the average stock in a portfolio where big firms inevitable dominate.

A third spoilsport remark, for FI aficionados, is that all the above was about the gross drag, before correction for market-wide average errors. A Roll-type analysis for the market is

¹²The 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 can be confident that we got the sign right.

Table 3: Noise required to explain various levels of extra return when aggregate noise is zero

Δ p.a., %	1.000	2.000	3.000	4.000	5.000
Δ p.m., %	0.083	0.167	0.250	0.333	0.417
	required σ , %				
if $\rho = 0$	2.89	4.08	5.00	5.77	6.45
if $\rho = 0.5$	4.08	5.77	7.07	8.16	9.13
if $\rho = 0.9$	9.13	12.91	15.81	18.26	20.41
if $\rho = 0.95$	12.9	18.3	22.4	25.8	28.9
	2σ confidence bound for v when lower bound is 10				
if $\rho = 0$	10.0 – 11.2	10.0 – 11.8	10.0 – 12.2	10.0 – 12.6	10.0 – 13.0
if $\rho = 0.5$	10.0 – 11.8	10.0 – 12.6	10.0 – 13.3	10.0 – 13.9	10.0 – 14.5
if $\rho = 0.9$	10.0 – 14.5	10.0 – 17.0	10.0 – 19.2	10.0 – 21.5	10.0 – 23.8
if $\rho = 0.95$	10.0 – 17.0	10.0 – 21.5	10.0 – 26.2	10.0 – 31.4	10.0 – 37.3

Key The table first shows what standard errors for the price noise would be required to generate given levels of noise-induced extra return, assuming that noise is diversified away at the market level. The second part shows the upper bound in a 2σ confidence band for v when the lower bound is 10 dollar.

not an option, given that its autocovariance is clearly positive.¹³ But, obviously, even standard errors of 20 percent would not be enough to explain FI's returns if also the market as a whole is non-trivially noisy—a plausible view, to the authors.

So at this point we conclude that there might be room for substantial gains, but only when noise is high; whether much remains after market-wide effects are taken into account can be verified directly. But we first turn to the cross-sectional behavior of weight adjustments.

3.2 The cross-section of weight adjustments

3.2.1 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 in time, the aggregate market noise $e_{m,t-1}$ is a constant, so that only the noise in individual prices is active.

We study the weight adjustment $\ln(w^*/W)$, that is, the log distance between the the weights proposed by FI and the original market weights; 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

¹³probably reflecting thin trading.

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 noise, if any, is a constant. Then

$$\begin{aligned}
 \ln(w^*/W) &= [\ln w + \phi'] - [\ln w + \epsilon' - \epsilon_m], \\
 &= \phi' - \epsilon' + \epsilon_m; \\
 \Rightarrow \overline{\text{cov}}(w^*/W, \ln V) &= \overline{\text{cov}}(\phi' - \epsilon', \ln v + \epsilon'), \\
 &= \overline{\text{cov}}(\phi', \ln v) - \overline{\text{var}}(\epsilon').
 \end{aligned} \tag{11}$$

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 take FI's claim as our working hypothesis: the error in the alternative weight, being style neutral, is uncorrelated with size. The result is an implied level of noise, whose plausibility then sheds light on the credibility of FI orthodoxy. In a way, though, this is like flogging a dead horse: from the style regressions we already know that FI is not style neutral. So the merit in studying these covariances may be more in the alternative view: we can start from priors about the level of noise, and infer something about the first term on the right, the 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, easier to interpret as an estimate of the standard deviation of noise under the assumption of no style shift. The other one is the covariance scaled by the variance of size—that is, the slope coefficient in the regression

$$\ln(w_j^*/W_j) = a + b(\ln V_j - \overline{\ln V}) + \nu_j. \tag{12}$$

This is easier to interpret than a covariance, and provides an easy 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 just allows us to estimate the cross-sectional variance of the ϵ s, which is a mixture of stock-by-stock variances. Or, just taking the regression in itself, *i.e.* without the interpretation in terms of the Roll model, the above equation assumes that the weight adjustments are linear in size. We can generalize and allow the slope to vary linearly in size, implying that the total adjustment

becomes quadratic in size:

$$\text{If } b(V_j) = b_0 + b_1 (\ln V_j - \overline{\ln V}), \quad (13)$$

$$\text{then } \ln \frac{\widehat{W}_j}{W_j} = a + [b_0 + b_1 (\ln V_j - \overline{\ln V})] (\ln V_j - \overline{\ln V}), \quad (14)$$

We can then back out an implied noise level for stock j , or at least its projection on $\ln V$, via $\hat{\sigma}^2(\epsilon_j) = b(V_j) \overline{\text{var}}(\ln V)$.

Lastly, the regression version of the 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. The next subsection digresses on this.

3.2.2 Are market values informative about noise? A digression

Perold (2007) discusses this issue in a Bayesian framework and shows that, under a diffuse prior, the market value is uninformative. Yet Appendix A 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 log normality model. 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*.¹⁴

Whence the difference between our and Perold's claims? Perold's prior is a Bayesian one, while here we use the distribution of the true market values instead. Conceptually, the frameworks are different even though the math is similar. 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. Of course, as a distribution for $\ln v$, a uniform on $[-\infty, +\infty]$ does not make sense, neither conceptually nor empirically: observed histograms of $\ln V$ have a pronounced mode. But in practice Perold could still be correct to all practical purposes: if the noise is small relative to the cross-sectional variance, there is almost no useful news (b is small).¹⁵

A simple calculation gives a feel for what adjustments would be reasonable. In the U.S. histogram of $\ln V$ shown in the Appendix, for instance, $\ln V$ has a cross-sectional standard

¹⁴Note 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.

¹⁵There would still be one qualification to be added to Perold's work: he treats regression towards the mean as the key mechanism behind FI, while in reality the 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.

Table 4: Calculated value-correction factors $(E(v|V)/V - 1) * 100$ for the U.S. ($\sigma_V = 1.87$)

$\sigma_\epsilon, \%$	% bias, n stdevs from the mean, with $n=$					
	-3	-2	-1	1	2	3
2%	-0.06%	-0.04%	-0.02%	0.02%	0.04%	0.06%
8%	-1.02%	-0.68%	-0.34%	0.34%	0.69%	1.03%
18%	-5.07%	-3.41%	-1.72%	1.75%	3.53%	5.34%

Key In the samples shown in the graph, selected percentile observations are corrected following equation. The table shows values for $e^{-\frac{(\sigma_{\epsilon_j}^2 / \text{var}(\ln V))(\ln V_j - \overline{\ln V})}{2}} - 1$, in percent.

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

deviation of 1.87, which totally dwarfs the noise variances considered thus far. Table 4 illustrates what kind of required correction factors we get assuming errors standard deviations of 2% (Brennan and Wang's lower estimate) and 8 and 18% (the calibration results computed from the Roll autocorrelations), for prices whose logs are -3 to +3 standard deviations away from the mean. The rational price correction is typically less than five percent even for the largest noise level. For the Brennan-Wang estimate, it is absolutely tiny.

Contrast this to what actually happens. In the numbers underlying Figure 3, every month we sort stocks on size and 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 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. In fact, even at the 75th size percentile the FI weights are still 1.5 times their market counterparts, and only for the top vigintile FI recommends underinvestment. Obviously, bearing in mind the result from the above table, the amount of noise needed to explain such drastic corrections should be much larger than the values we just considered. To see how much larger, and how this varies over time, we turn to the regression tests.

3.2.3 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 telling us that these shifts are not random estimation errors. The slopes are, in addition, negative over most of our sample; only cash-flow-based weights turn mildly positive for a non-trivial length of time, namely the intercrisis lull of the noughties. A typical value 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 . All of this makes it questionable to talk about a mean value of their implied noise standard deviations. But for book value as the FI variable we have a stable and uninterrupted series.

BV is the most regressive weighting scheme, the one with the most pronounced shrinking towards the mean, as measured by b . (Recall that it is also the most successful FI strategy. This already suggests that the extra returns partly reflect extra risk.) A second immediate inference is that BV also needs generally higher implied levels of noise for the latter to be the sole reason for shrinking the weights. The time path of this implied risk potentially makes sense, with rising noisiness around the 2000/01 peak and the 2008-09 crisis, and uncommonly low noise in-between these two crisis episodes. But the levels of the required noise, mostly 60-80%, make it utterly implausible that the shrinkage entails no style shift. This is 20-30 times the Brennan-Wang estimates. Or, more concretely, 60-80% noise would mean extravagant confidence bounds for v between 10-110 dollar and 10-250 dollar.

Obviously, then, it is equally implausible that variation over time in the cross-sectional covariance mostly reflects changes in the noise level. If we take the Brennan and Wang estimate

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

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 $\ln \frac{\widehat{W}_j}{W_j} = a + b(\ln V_j - \overline{\ln V})$

Key This plot shows, for each of the three FI weights, $\sqrt{-b \text{var}(\ln V)}$, 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.

as our starting point, then the average noise variance is like 0.0004-0.0009, which is trivial to the average covariance, typically 0.3 to 0.4. Even if noise standard deviation is ten times the Brennan-Wang level, the noise variance would still be no more than 0.04-0.09, which remains small relative to the average covariance. According to Equation 11, then, most of the variability

must reflect variability in the style changes over time, which then casts doubt on the validity of style regressions.

And this is just the tip of an iceberg, as we discover when we turn to the quadratic model. 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 but not all of the estimated b_1 s are positive: 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 case. There is, in fact, a marked downward trend in all b_1 s 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.

Figure 7, next, plots the series of the size-varying slopes, $b(V, t) = b_{0,t} + b_{1,t} (\ln V_{j,t} - \overline{\ln V_t})$, for the BV weighting scheme and for the 1st, 5th, 10th, 15th and 19th vigintile values. The quadratically size-related corrections are quite strong in the early 1990s. They disappear in the late 1990s (that is, weight corrections become roughly linear, on average, consistent with noise being unrelated to size), only to return as of the 2008 crisis. Next to that graph we plot the noise volatility required to explain all this as style-neutral drag avoidance. The earlier estimate, 60-80%, seems to be valid for the bulk of the firms: while firms in vigintiles 1 to 15 do show lower and lower implied risks the larger they are, the differences are not very marked. Rather, the outlier, pre-1999 and post-2006, is the largest-cap group, whose noisiness would nevertheless still have to be in the implausible 40-60% range. Equally unconvincing, if we

Figure 6: Estimated b_0 and b_1 for BV weights in Equation (14)

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: $\ln(\widehat{W}_j/W_j) = a + [b_0 + b_1 (\ln V_j - \overline{\ln V})] (\ln V_j - \overline{\ln V})$. The graph shows the time series of estimated b_0 and b_1 .

Figure 7: Total shrinkage factor for vigintiles 1, 5, 10, 15 and 20; noise standard deviation required to make such shrinkage a normal Bayesian update for BV weights in Equation (14)

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: $\ln(\widehat{W}_j/W_j) = a + [b_0 + b_1 (\ln V_j - \ln \bar{V})] (\ln V_j - \ln \bar{V})$. The top graph shows the time series of estimated $b(t, V) = b_0 + b_1 (\ln V_j - \ln \bar{V})$ for the 1st, 5th, 10th, 15th and 20th vigintile firm. The bottom graph shows, for each of the three FI weights, $\sqrt{-b(t, V) \text{var}(\ln V)}$, which is the level of noise that would make the observed degree of shrinking a rational Bayesian response to noise.

would believe style neutrality, then in 2000-2006 big stocks would have been as hard to value as tiny startups. Even in the sleepy year 2005, all stocks would still need noise levels of 40%, if the BV investment strategy were just drag avoidance. In short, then, there must be substantial style shifts, unstable both across the size spectrum and over time. This raises questions about the FF regressions, with their assumedly constant exposures to all risks. So we now turn to alternative direct estimates of the benefit from avoiding drag.

4 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. As mentioned, Basset *et al.* consider lagged markets weights while Treynor proposed equal weights. One attraction of FI, in addition to avoiding drag, definitely is that it keeps the strategy feasible, steering most money away from micro firms and into big stocks, even if this is not done in a style-neutral way. 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. It is hard to believe, for instance, that a naive, mechanical strategy like EW creates real value, as its alpha in Section 1 would suggest.

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.

4.1 Test procedure: size-bucket-constrained portfolios

Our alternative test starts from Treynor's (2005) EW proposal. To avoid the concomitant style shift, we adopt a compromise between 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. Within the bucket, however, we weight equally, Treynor-style. In other words, each company's value weight is replaced by the average value weight of all stocks in the company's size vigintile—hence our label, $\bar{w}_v$ weighting, with the notation standing for average weight in vigintile v .

Studying these 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 $1/N_v$.¹⁶ Second, this drag reduction is achieved with far less scope for style changes than under a regular EW strategy, especially for size. One may object that this procedure just conditions the buckets on size, 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 EW 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 where the gains are coming from while largely controlling for size.

We also apply variants of the above design: we still put stocks into buckets per size vigintile, but within each of these size buckets we assign, in step two, weights based on fundamental variables or lagged relative market cap. In terms of drag, bucket per bucket, these strategies should not do better than the $\bar{w}_v$ variant; so if they do better, we have direct evidence of style shifts, even at the within-bucket level. Stated differently, we can compare the FI and EW strategies to $\bar{w}_v$ rather than the full VW one, to see, bucket by bucket, how much comes from drag and how much from extra style shift. Shifts in exposure to HML, for instance, are especially likely in BV-weighted portfolios, even when bucket-constrained: since within each size bucket the market values are rather similar, sorting on book value is close to sorting on book/market, *i.e.* HML exposure. If that analysis is valid, it should show up as an extra return above the bucket's EW return.

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

Table 5 shows how the mean extra returns (relative to VW) and standard deviations are shifting after bucket-constraining the weights. EW is hugely affected, unsurprisingly: it loses almost nine tenths of its extra return. For the LW strategies the effect is minor, which was again predictable. For FI, lastly, the effect is smaller than we had expected: much of the extra return does not come from massive size-style shifts, apparently. In fact, BV loses just one-third of its extra return, SL one-fourth, and CF virtually nothing. This is in line with the style regressions from Section 2, where we found that, relative to VW, FI portfolios load much more on HML than on SMB. Still, even these relatively small drops in return are sufficient to make all mean extra returns bar BV's insignificant. Since drag avoidance should not be affected by the new weights, this observation must mean that this simple way of controlling for size already makes

¹⁶This analysis is too optimistic: sorting by size V means some bunching of valuation errors ϵ , as argued in section 3.2; so the average error is not quite as good as the $1/N$ rule would suggest. But the discrepancy is small, as we see later.

Table 5: Mean extra return and standard deviations: unconstrained *v* bucket-constrained portfolios

	BV	SL	CF	EW	LW1	LW3	LW3
	extra average monthly returns, %						
unconstrained	0.312	0.244	0.228	0.454	0.027	0.020	0.007
constrained	0.237	0.162	0.200	0.048	0.027	0.006	-0.006
t versus VW	2.980	1.430	1.880	0.960	1.230	0.140	-0.100
	extra stdev of monthly returns, %						
unconstrained	0.310	0.250	0.280	1.400	0.100	0.240	0.390
constrained	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 size-bucket-constrained versions, and t-statistics for the latter return. In all these constrained 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.

the FI case less convincing.

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

4.3 Results (2): returns for equally-weighted vigintile portfolios

Figure 8 summarizes the performance of each alternative weighting schemes via its average monthly return in excess of the bucket’s market-value weighted index. For the key strategies, in addition, we consider the FF alphas for each bucket’s return in excess of VW, see Table 6 and Figure 9.

First, for EW within buckets the extra returns realized in each vigintile relative to a value-weighted counterpart are essentially zero, at least for size classes 3 to 20, suggesting that drag-avoidance benefits are economically meaningless. The same conclusion holds for conventionally risk-adjusted returns: all alphas except the first two are insignificant. All this is consistent with net drag being minute, except possibly for the tiniest 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 this (which we question, below), they are too small to meaningfully affect any implementable portfolio—just look at $\bar{w}_v$ ’s measly 0.0048 percent overall performance, for instance (Table 5). Optimistically, the near-zero drag in most stocks could mean markets are quite efficient; realistically, it could also mean that there is a large market-wide component in the pricing errors (market-wide bubbles, for instance), in which case 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

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.

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 past six months for all stocks 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 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

Table 6: 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.

months underperforms by about 2 and 1 percent, respectively, i.e. effectively 13 and 6 percent over six months. Part of that 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 6, 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. Note, indeed, that the variable under study is the return difference between the bucket's EW portfolio and the bucket's VW portfolio. So this refers to a portfolio that heavily loads on the smallest stocks in the bucket (the losers), quite unexposed to the middle stocks and heavily short in the largest stocks in the bucket. This is not drag avoidance; if anything, it sounds more like a style even though it is not a FF-canonized one. Also, we repeat, the low-cap stocks are too small anyway to have a meaningful impact on any realistic portfolio, so even if there is some net drag at the lower end of the size spectrum it does not explain FI's extra returns.

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).

4.4 Results (3): returns for non-EW vigintile portfolios

Returning to Figure 8, we now consider portfolios weighted on lagged market values. We see that those 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, 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 FI portfolios, we note a substantially positive extra return relative to their VW and EW benchmarks for each size bucket up to vigintile 17 or 18. This 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, within buckets where market values are quite similar, means loading on the book-to-market factor. Yet when we correct for FF risks (Table 6), the alphas remain significant for

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 (effective) 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.

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 3.2) is hard to say. But we can put the rest the idea that avoiding drag is a worthwhile strategy.

5 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-istas stress that extra returns can be reaped from avoidance of drag, by using weights that are less correlated with mispricing. But also style shifts can be at work, which would mean that the extra returns are not adding real value.

Using lagged returns, one alternative scheme suggested in the literature if one wants to avoid drag, has the drawback that it brings in momentum (for the winners, especially large ones) and reversal (for the losers, especially small ones). Equal weighting within size vigintiles combined with value weighting across vigintiles provides a better way of avoiding drag. When we implement this, we find that drag is an economically insignificant source of return, though.

It may still play some role for the smallest stocks, but the high returns at the extreme lower end of the scale are surely also reflecting a profitable reversal effect that fits in with the Fallen Angel phenomenon.

Testing whether FI shrinks the weights towards the mean by no more than what noisy prices justify, we find massive upward corrections for small stocks, which could reflect pure drag avoidance only if noise were implausibly gigantic. 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 quadratically related to size, while at other times the link is more linear, or disappears, or even reverts. This makes it implausible that style shifts, reflected in exposures of FI strategies to generalised FF factors, are constant over time.

Subject to the above caveat, FF regressions that do assume constant factor loadings show that only part of the extra returns can be traced to changes in the conventional styles (Book-to-market and Size). Adding French's short-term reversal factor does eliminate all alphas but one, though. Controlling for the dot.com period (a period with both an unusual size mix and unusual factor returns) does so too. In fact, most of the FI return seems to come from that episode; otherwise there is no clear extra return.

Appendix: Bayesian updating of market values

In the Roll (1984) model, mispricing is to some extent identifiable, on average: if $E(\epsilon_{j,t}|v_{j,t}) = 0$, in general we no longer have a zero expectation for ϵ conditional on V : $E(\epsilon_{j,t}|V_{j,t}) = E(\epsilon_{j,t}|v_{j,t} + \epsilon_{j,t}) \neq 0$, so that $E(v_{j,t}|V_{j,t}) \neq V_{j,t}$. 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 problem¹⁷ 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. For v , the choice of a fully diffuse uniform density would imply that, say, any young

¹⁷Perold 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 .

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.

upstart might ‘really’ be worth more than Apple with 50% probability. Additionally, a uniform on $[-\infty, +\infty]$ 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.

We first explain the logic of identifiable errors generated by a non-diffuse v via an example. For ease of exposition and interpretation, in the first example (Table 7, the panel labeled “additive”) we work with a symmetric distribution of true prices in a grid $v = 2, 3, \dots, 10$. 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 $E(V|v) = v$. 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 $E(v|V) \neq V$. 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

Table 7: Attenuation in a noisy-market model: two simple examples

additive														
		$N_V(V)$										$E(v V)$ $N_V(V)$	bias, \$ $E(v V)$	bias, %
12											5	5.0	10.00	2.00
11										10	10	20.0	9.50	1.50
10										20	20	110.0	9.45	0.05
9										35	40	225.0	8.56	0.04
8										70	280	415.0	7.71	0.04
7										35	80	655.0	6.86	0.02
6										70	560	740.0	6.00	0.00
5										490	70	655.0	5.14	-0.03
4										35	20	415.0	4.29	-0.07
3										70	40	225.0	3.44	-0.15
2										20	20	110.0	2.55	-0.27
1										10	10	20.0	2.50	-1.50
0										5	5	5.0	2.00	-infy
multiplicative														
	$N_v(v)$	100	200	400	400	700	700	400	200	100	$N_V(V)$	$E(v V)$	bias, \$	bias, %
14.930											5.9	10.37	4.56	0.31
12.442											21.8	33.62	9.76	0.22
10.368											50.0	117.24	9.09	0.12
8.640											18.2	246.77	7.82	0.10
7.200											4.1	440.40	6.70	0.07
6.000												646.78	5.78	0.04
5.000											720.00	5.00	0.00	0.00
4.167											623.22	4.32	-0.16	-0.04
3.472											409.60	3.73	-0.26	-0.08
2.894											223.23	3.20	-0.31	-0.11
2.411											102.76	2.75	-0.34	-0.14
2.009											26.38	2.56	-0.55	-0.27
1.674											4.10	2.41	-0.74	-0.44
$N_v(v)$		2.411	2.894	3.472	3.472	4.167	5.000	6.000	7.200	8.640	10.368			

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 $e = \{-2 \text{ ticks}, -1 \text{ tick}, 0, 1 \text{ tick}, 2 \text{ ticks}\}$ 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 .

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 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 $\mu_{v_j} := E(\ln v_j)$ and variance $\sigma_{v_j}^2 := \text{var}(\ln v)$; by experience, she knows there is an independent lognormal noise term, resulting in a lognormal V with *ex ante* mean $\mu_{V_j} := E(\ln V_j)$ and variance $\sigma_{V_j}^2 := \text{var}(\ln V)$. 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

$$\ln V = \ln v + \epsilon', \quad (15)$$

$$\text{where } \epsilon' := \ln(1 + \epsilon) \sim N(-(1/2)\sigma_{\epsilon'}^2, \sigma_{\epsilon'}^2), \quad (16)$$

so that we still have $E(V|v) = v$.¹⁸ The mean of the conditional distribution of v given V is identified via the regression of $\ln v$ on $\ln V$. Equation (15) immediately implies three properties that are used in the math below: $\text{cov}(\ln V, \ln v) = \sigma_v^2$, $\sigma_V^2 = \sigma_v^2 + \sigma_{\epsilon'}^2$, and $\mu_V = \mu_v - (1/2)\sigma_{\epsilon'}^2$. 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

¹⁸Note for students: If $\ln(1 + \epsilon)$ is normal with mean μ , then $E(1 + \epsilon) \equiv E(e^{\epsilon'}) = e^{\mu + (1/2)\sigma_{\epsilon'}^2} > e^{\mu}$, (a Jensen's Inequality effect). So to ensure unbiasedness of ϵ , $\ln(1 + \epsilon)$ must have a below-zero mean.

the error variance:

$$\begin{aligned}
 E(v|V) &= e^{E(\ln v|V)+1/2\text{var}(v|V)} \\
 &= e^{\mu_v+\beta(\ln V-\mu_V)+1/2\sigma_{\epsilon'}^2}, \\
 &\quad \text{where } \beta = \frac{\text{cov}(\ln V, \ln v)}{\text{var}(\ln V)} = \frac{\sigma_v^2}{\sigma_v^2 + \sigma_{\epsilon'}^2} < 1; \tag{17}
 \end{aligned}$$

$$\begin{aligned}
 &= e^{\mu_v+\beta(\ln V-\mu_V)}, \\
 &= V e^{(\beta-1)(\ln V-\mu_V)}; \\
 &= V e^{-\frac{\sigma_{\epsilon'}^2}{\sigma_v^2}(\ln V-\mu_V)}. \tag{18}
 \end{aligned}$$

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, $-\frac{\sigma_{\epsilon'}^2}{\sigma_v^2}$ is the same as the scaled covariance used in the Section 3.2, at least in the econometrician's view alluded to above.

References

- 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.
- Basset, G.W., C. Chen and R. Chen. 2007. Fundamental Indexation via Smoothed Cap Weights. *Journal of Banking and Finance* 31(11), 3486-3506
- 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
- 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), 2373-2397
- 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.
- 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
- Perold A. F., 2007, Fundamentally Flawed Indexing, *Financial Analysts Journal* **63(6)**, 31-37.
- 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.
- 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