On the Importance of Evaluating Mutual Fund's Alpha
via Alternative Frameworks

Sharon Garyn-Tal[1]

Abstract

We examine over 1000 U.S non-specialized mutual funds in 2001-2009 and offer a comprehensive report on benchmark model impact on fund's estimated alpha (risk-adjusted excess return). Consistent with previous literature, we find that estimated alpha's magnitudes strongly depend on the benchmark system. Furthermore, there exists substantial disagreement among our 7 benchmark systems regarding fund's classification as good or poor, and fund's ranking relative to other funds. Notably, the Jensen (CAPM based) alphas are most deviant from other models' alphas, and the best funds recommended by it (the CAPM) score lower Sharpe ratios in out of sample tests. This suggests that asset management and investment advice should also and perhaps mainly rely on performance scores of multi-factor benchmark models.

1 Introduction

Previous research suggests that alternative risk-adjusted benchmarks lead to substantive differences in performance measures. For example, [1] find that funds' rankings are very sensitive to the asset-pricing model chosen to measure performance. [2], [3], [4] also show that estimated performance is sensitive to the choice of the benchmarks.

This paper compares different benchmark models and evaluates their impact on estimating the fund manager's contribution, alpha, to the return of her investors. We estimate the alphas via 7 benchmark models. We then compare the fund's classification as good or poor, and the fund's ranking relative to other funds across these 7 benchmark models. We also examine the overlap of the 20 best performing funds and the 20 worst performing funds

[1]The Economics and Management Department, The Max Stern Academic College of Emek Yezreel, Emek Yezreel 19300, Israel..

across the 7 benchmark models. Using a sample of nine years (2001-2009) and over 1000 non-specialized open-end equity funds' returns, we offer a relatively comprehensive comparison of benchmark impact on fund performance.

We find that alpha's magnitude strongly depends on the benchmark model applied for funds' performance evaluation. There also exist substantial differences in funds rankings and classifications especially when comparing the CAPM with multi-factor models. In fact, the CAPM generates the most deviant and least consistent performance measures. Moreover, in an out of sample test, the funds recommended by the CAPM-based Jensen alphas, yield lower Sharpe ratios than those recommended by the multi-factor models.

We conclude that in asset management and investment advice, multi-factor risk-return models should assume a major role.

2   Data and Benchmark Models

2.1 Sample and Variables

We collect monthly data on mutual fund returns in 2001-2009 from CRSP (The Center for Research in Security Prices). The CRSP data uses the Lipper classification that divides the world of non-specialized open-end equity funds into 12 groups based on funds' style: LCCE (Large-Cap Core Funds), LCGE (Large-Cap Growth Funds), LCVE (Large-Cap Value Funds), MCCE (Mid-Cap Core Funds), MCGE (Mid-Cap Growth Funds), MCVE (Mid-Cap Value Funds), SCCE (Small-Cap Core Funds), SCGE (Small-Cap Growth Funds), SCVE (Small-Cap Value Funds), MLCE (Multi-Cap Core Funds), MLGE (Multi-Cap Growth Funds), and MLVE (Multi-Cap Value Funds). We prefer the Lipper classification because it divides the universe of non-specialized open-end equity funds into 12 categories, more than other fund classifications we considered. However, the Lipper classification of our funds starts only in 2001; thus our data starts in 2001.

For the empirical analysis we need data on several other variables. The Fama-French-Carhart four factors returns are collected from French website:

http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

The liquidity factor data is obtained from Pastor's website:

http://faculty.chicagobooth.edu/lubos.pastor/research/liq_data_1962_2011.txt.

In our empirical work we also classify funds by fund's style, where fund's style is defined as the benchmark that yields the highest correlation with respect to the fund's return. We review the internet information sheet of Fidelity and Vanguard non-specialized equity funds, and find that the main benchmarks used for evaluating these funds' performance are the S&P500 or an appropriate Russell index (Russell 1000, Russell 1000 growth, Russell 1000 value, Russell midcap, Russell midcap growth, Russell midcap value, Russell 2000, Russell 2000 growth, Russell 2000 value). Thus, we also construct groups of funds based on the 10 above benchmark indices that we consider as possible styles. The Russell indices data and returns are extracted from the Russell investments website (www.russell.com). S&P500 returns are collected from the CRSP database.

2.2 Data Characteristics

The data comprises nine years (2001-2009) of funds' returns and is divided into three subperiods: 2001-2003, 2004-2006, and 2007-2009.

There are 1037 funds with full return data throughout 2001-2003. Of the 1037 full-data funds, 188 are classified by Lipper as Large-Cap Core Funds (LCCE), 123 are Large-Cap Growth Funds (LCGE), 66 are Large-Cap Value Funds(LCVE), 52 are Mid-Cap Core Funds (MCCE), 85 are Mid-Cap Growth Funds (MCGE), 34 are Mid-Cap Value Funds (MCVE), 68 are Multi-Cap Core Funds (MLCE), 68 are Multi-Cap Growth Funds (MLGE), 98 are Multi-Cap Value Funds (MLVE), 89 are Small-Cap Core Funds (SCCE), 100 are Small-Cap Growth Funds (SCGE), and 66 are Small-Cap Value Funds (SCVE). The 2001-2003 average Total Net Assets (TNA) of all funds is 667.9 million dollars. The average expense ratio (management fee) charged by fund managers is 1.31% (0.74%) and the weighted (by TNA) average expense ratio (management fee) is 1.24% (0.74% as well). [5] study fees charged by mutual funds in 2002. They document (in Table 2, page 1288) an average management fee of 0.62% per year and an average expense ratio of 1.11% per year for U.S equity funds. These fees are slightly lower than the fees we report above.

There are 1547 funds with full return data in 2004-2006. The number of funds in each Lipper classification is available from the authors. However, there are at least 64 funds in each of the above-specified 12 Lipper groups. The 2004-2006 average TNA of all funds is 971.2 million dollars. The average expense ratio (management fee) charged by fund managers is 1.37% (0.78%).

There are 1312 funds with full return data throughout 2007-2009. In 2007-2009 we have at least 48 funds in each of the Lipper groups. The 2007-2009 average TNA of all funds is 759.6 million dollars. The average expense ratio (management fee) charged by fund managers is 1.27% (0.73%).

2.3 The Benchmark Models

We employ 7 different benchmark models to evaluate a fund's alpha. The models differ in the benchmark-system used in the regression. The 7 alternative benchmark models examined are:

1. M - the CAPM. The regression equation for fund P is:

$$R_{P,t} - R_{f,t} = \alpha_P + \beta_P(R_{M,t} - R_{f,t}) + \varepsilon_{P,t}.$$

The excess market return is based on the value-weight return on all NYSE, AMEX, and NASDAQ stocks (from CRSP) and on the one-month Treasury bill rate (from Ibbotson Associates).

2. S - the fund's style. The style benchmark is determined by computing the correlation between a fund's return and the return on the following ten indices - Russell 1000, Russell 1000 growth, Russell 1000 value, Russell midcap, Russell midcap growth, Russell midcap value, Russell 2000, Russell 2000 growth, Russell 2000 value, and S&P500. The index that has the highest correlation with the fund's return is defined as the fund's style. The regression equation for fund P is:

$$R_{P,t} - R_{f,t} = \alpha_P + \beta_P(R_{S,t} - R_{f,t}) + \varepsilon_{P,t}.$$

3. G(Class) - a group factor. Define $R_{G(Class)}$ as the average return of all funds belonging to the fund's (Lipper) classification group. The regression equation for fund P is:

$$R_{P,t} - R_{f,t} = \alpha_P + \beta_P(R_{G(Class),t} - R_{f,t}) + \varepsilon_{P,t}.$$

4. G(Style) - an alternative group factor. We define $R_{G(Style)}$ as the average return of all funds belonging to the fund's style group. The regression equation for fund P is:

$$R_{P,t} - R_{f,t} = \alpha_P + \beta_P(R_{G(Style),t} - R_{f,t}) + \varepsilon_{P,t}.$$

5. 3F - the Fama and French three-factor model. The regression equation for for fund P is:

$$R_{P,t} - R_{f,t} = \alpha_P + \beta_{P,1}(R_{M,t} - R_{f,t}) + \beta_{P,2}HML_t + \beta_{P,3}SMB_t + \varepsilon_{P,t}.$$

The three factors above are: 1. the excess market returns is based on the value-weight return on all NYSE, AMEX, and NASDAQ stocks (from CRSP) and on the one-month Treasury bill rate (from Ibbotson Associates), 2. the performance of value stocks relative to growth stocks (HML, High Minus Low), and 3. the performance of small stocks relative to big stocks (SMB, Small Minus Big), based upon the Fama-French Portfolios.

6. 4F - The four-factor model of [6]. The regression equation for fund P is:

$$R_{P,t} - R_{f,t} = \alpha_P + \beta_{P,1}(R_{M,t} - R_{f,t}) + \beta_{P,2}HML_t + \beta_{P,3}SMB_t + \beta_{P,4}MOM_t + \varepsilon_{P,t}.$$

This model is based on the Fama and French three-factor model and the additional momentum factor (MOM) of [6], constructed as the average return on the two high prior return portfolios minus the average return on the two low prior return portfolios.

7. 4F-Liquidity - a five factor model that adds the liquidity factor (see [7]) to the Fama-French-Carhart model. The regression equation for fund P is:

$$R_{P,t} - R_{f,t} = \alpha_P + \beta_{P,1}(R_{M,t} - R_{f,t}) + \beta_{P,2}HML_t + \beta_{P,3}SMB_t$$
$$+ \beta_{P,4}MOM_t + \beta_{(Liquidity)}Liquidity_t + \varepsilon_{P,t}.$$

3   Empirical Results

3.1 Funds' Alpha Estimates

We fit our benchmark models over three subperiods: 2001-2003, 2004-2006 and 2007-2009. In each subperiod we evaluate all funds with returns throughout all 36 months of that subperiod. On average, we examine 1,299 funds per subperiod. Table 1 reports regression results for 2001-2003, 2004-2006, 2007-2009 and the mean over these three subperiods.

Table 1: The alphas generated by 7 benchmark-models of funds return

Model	2001-2003 (1,037 funds) Annual α	2004-2006 (1,547 funds) Annual α	2007-2009 (1,312 funds) Annual α	Mean across the 3 sub-periods Annual α
M	0.88%**	-2.50%**	0.23%*	-0.46%
S	-1.36%**	-0.98%**	-0.13%	-0.82%
G(Class)	N.R	N.R	N.R	N.R
G(Style)	N.R	N.R	N.R	N.R
3F	-3.79%**	-0.70%**	-0.71%**	-1.73%
4F	-3.68%**	-0.30%**	-0.71%**	-1.56%
4F-Liquidity	-3.84%**	-0.30%**	-0.80%**	-1.65%

N.R - not relevant. In models including the group factor (G), the mean alpha is zero (see Appendix 1)

The mean overall period alphas vary between -1.73% for the three-factor model and -0.46% for the CAPM. Consistent with previous evidence – see [6], [8] and[9], for example, mutual funds tend to deliver disappointing net returns. The negative excess returns can be explained, on average, by the funds' expense ratios, suggesting that the before-fees average performance of the funds is on average fair and adequate.

The average alpha of benchmark models that include the group factor equals zero by definition (see Appendix 1 for a proof). For that reason, the average alpha of models that include the group factor are denoted as N.R (Not Relevant). The group-factor methodology allows only within a group classification and ranking.

An interesting observation is that the CAPM yields the highest alpha (positive and statistically significant) in 2001-2003 and 2007-2009 and the lowest alpha (negative and statistically significant) in 2004-2006. In contrast, all our other models yield, in all three periods, negative alphas that are mostly statistically significant. Evidently, the CAPM alphas are most (and perhaps too) volatile.

Table 2 reports the correlations between the alphas generated by the 7 benchmark models. The correlations between the alphas vary from 0.67, between the CAPM and the five-factor model, to 0.99, between the four-factor model and the five-factor model. The relative high correlations documented in Table 2 imply that there exist some wide agreement between the benchmark models. However, again, the CAPM alphas appear most deviant, i.e., have the lowest correlation with other model alphas.

Table 2: The correlations between the alphas generated by seven benchmark models

Benchmark Model	M	S	G(Class)	G(Style)	3F	4F	4F-Liquidity
				Correlation of Alphas			
M	1	0.78	0.77	0.73	0.73	0.69	0.67
S	0.78	1	0.93	0.81	0.84	0.79	0.79
G(Class)	0.77	0.93	1	0.85	0.85	0.82	0.82
G(Style)	0.73	0.81	0.85	1	0.79	0.76	0.78
3F	0.73	0.84	0.85	0.79	1	0.97	0.97
4F	0.69	0.79	0.82	0.76	0.97	1	0.99
4F-Liquidity	0.67	0.79	0.82	0.78	0.97	0.99	1

3.2 Differences in Funds Classification and Ranking

First, each fund is classified by each benchmark model as either a good fund (positive alpha) or a poor fund (negative alpha). Then, we compare funds classifications. If benchmark model i and benchmark model j agree on fund classification, then there is no classification difference between models i and j. On the other hand, if a fund is classified as a good (poor) fund according to benchmark model i and is classified as a poor (good) fund according to benchmark model j, then there is a classification difference between models i and j. Table 3 summarizes the results.

Table 3: Differences in fund classification and ranking across seven possible benchmark models

Panel A: Classification differences among the benchmark models

	M	S	G(Class)	G(Style)	3F	4F	4F-Liquidity
M	0.0%	24.9%	31.3%	27.7%	28.8%	30.8%	32.1%
S	24.9%	0.0%	25.7%	17.0%	21.9%	25.0%	25.5%
G(Class)	31.3%	25.7%	0.0%	17.5%	27.1%	27.9%	28.0%
G(Style)	27.7%	17.0%	17.5%	0.0%	24.7%	27.0%	27.7%
3F	28.8%	21.9%	27.1%	24.7%	0.0%	6.0%	8.2%
4F	30.8%	25.0%	27.9%	27.0%	6.0%	0.0%	3.5%
4F-Liquidity	32.1%	25.5%	28.0%	27.7%	8.2%	3.5%	0.0%
Average*:	29.3%	23.3%	26.3%	23.6%	19.5%	20.0%	20.8%

*The average calculations exclude signs differences between model i and itself, i.e., the zeros above.

Panel B: Rank differences (of more than a decile) among The benchmark models

	M	S	G(Class)	G(Style)	3F	4F	4F-Liquidity
M	0.0%	62.9%	57.1%	58.3%	56.5%	59.8%	62.2%
S	62.9%	0.0%	49.3%	29.2%	51.0%	55.1%	54.5%
G(Class)	57.1%	49.3%	0.0%	37.9%	54.6%	56.8%	55.3%
G(Style)	58.3%	29.2%	37.9%	0.0%	49.7%	54.4%	53.5%
3F	56.5%	51.0%	54.6%	49.7%	0.0%	12.9%	18.0%
4F	59.8%	55.1%	56.8%	54.4%	12.9%	0.0%	4.1%
4F-Liquidity	62.2%	54.5%	55.3%	53.5%	18.0%	4.1%	0.0%
Average*:	59.5%	50.3%	51.8%	47.2%	40.5%	40.5%	41.3%

*The average calculations exclude differences between model i and itself, i.e., exclude the zeros above.

Panel C: The overlap of the 20 best performing funds across the benchmark models

	M	S	G(Class)	G(Style)	3F	4F	4F-Liquidity
M	20.0	7.7	9.3	9.3	9.3	7.7	7.7
S	7.7	20.0	8.7	13.3	9.7	8.3	8.3
G(Class)	9.3	8.7	20.0	10.7	9.3	8.0	8.3
G(Style)	9.3	13.3	10.7	20.0	10.7	9.7	9.7
3F	9.3	9.7	9.3	10.7	20.0	16.0	15.3
4F	7.7	8.3	8.0	9.7	16.0	20.0	17.7
4F-Liquidity	7.7	8.3	8.3	9.7	15.3	17.7	20.0
Average*:	8.5	9.3	9.1	10.6	11.7	11.2	11.2

*The average calculations exclude numbers on the diagonal (all of which are 20)

Panel D: The overlap of the 20 worst performing funds across the benchmark models

	M	S	G(Class)	G(Style)	3F	4F	4F-Liquidity
M	20.0	13.0	11.0	11.0	11.3	11.7	11.0
S	13.0	20.0	12.3	13.3	10.3	9.0	9.0
G(Class)	11.0	12.3	20.0	12.3	12.3	11.7	11.0
G(Style)	11.0	13.3	12.3	20.0	11.3	10.3	10.7
3F	11.3	10.3	12.3	11.3	20.0	17.3	16.0
4F	11.7	9.0	11.7	10.3	17.3	20.0	18.0
4F-Liquidity	11.0	9.0	11.0	10.7	16.0	18.0	20.0
Average*:	11.5	11.2	11.8	11.5	13.1	13.0	12.6

*The average calculations exclude numbers on the diagonal (all of which are 20)

As reported in Panel A, the frequency of fund classification differences varies between from 3.5% and 32.1%. Minor classification disagreements exist between the four-factor model and the four-factor plus liquidity five-factor model (3.5%). Substantial classification differences emerge when comparing the CAPM and the three-, four- and five-factor models (28.8%-32.1%). Other one-factor models yield also substantial classification differences relative to the three- to five-factor models.

Panel B presents ranking differences. Assume N funds are available. For each benchmark model, the best performing fund with the highest alpha is ranked at the 1st place, the second best performing fund with the second highest alpha is ranked at the 2nd place, and so on. Given that in each subperiod a fund has 7 different alphas, in each subperiod funds are ranked 7 times. Assume fund P is ranked in the k-th place according to benchmark model i (i.e., according to $\alpha_{i,P}$) and is ranked in the l-th place according to benchmark model j (i.e., according to $\alpha_{j,P}$). If $|k - l| < 0.1N$, i.e., the ranking difference is less than 10% of the N existing funds, then we, somewhat arbitrarily, decide that there is no ranking difference between models i and j. On the other hand, If $|k - l| \geq 0.1N$, then we mark it as if there is a ranking difference between models i and j. Panel B reports the 2001-2003, 2004-2006 and 2007-2009 average frequency of ranking differences between the benchmark models.

The ranking-difference frequency varies between 4.1% and 62.9%. Small ranking differences emerge when comparing the four-factor model and the five-factor model (4.1%). Substantial ranking differences exist when comparing all other models, especially when comparing the CAPM with all other models (56.5%-62.9%).

For Panels C and D analysis, we narrow the sample and keep only funds ranked as 20 best performing funds or as 20 worst performing funds. Each subperiod, each benchmark model has its own list of funds included in the 20 best and in the 20 worst performing funds. Then, we count how many funds appear on both benchmark method i and benchmark method j lists of best (worst) performing funds, and report the numbers in Panel C (D).

In Panel C, the average ranking overlap of the 20 best performing funds varies between 5 and 17.7 funds. The highest overlap is obtained when comparing the four-factor model and the five-factor model. The lowest overlap of best performing funds is obtained when comparing the CAPM with other models.

In Panel D, the average ranking overlap of the 20 worst performing funds vary from 8.7 to 18. The highest overlap is obtained when comparing the four-factor model and the five-factor model. The lowest overlap occurs when comparing the one-factor models with other models.

In sum, the evidence in this subsection supports the view that a fund's classification and ranking depends significantly on the benchmark-system used for its evaluation. Notably, the CAPM appears most extreme - it generates the highest disagreement proportions in Table 3. Basing a performance evaluation solely on the CAPM benchmark appears erroneous.

3.3 A Glimpse at out of Sample Performance

This subsection takes a glimpse over the out of sample performance of our 7 benchmark models. We focus on the top two deciles of performing funds recommended by each benchmark model. Specifically, we build a portfolio of the funds belonging to the best two

deciles of performing funds according to benchmark model i in a specific subperiod (say 2001-2003) and examine its Sharpe ratio in the next subperiod (2004-2006 in our example). Then, we repeat the analysis with the top 2 deciles of performing funds in 2004-2006 according to model i, using 2007-2009 as the out of sample test subperiod.

Table 4 summarize our results. In the first column we report the Sharpe ratios in 2004-2006 of the portfolio of the top two deciles of performing funds in 2001-2003. The portfolio of best funds recommended by the CAPM (four-factor model) scores a Sharpe ratio of 0.257 (0.276 respectively) in 2004-2006. Reviewing the first column it becomes apparent that the funds recommended by the multi-factor models score higher Sharpe ratios in the out-of-sample period.

Table 4: The out of sample Sharpe ratios of "best funds".

Model	Formation Period: 2001-2003 Performance Period: 2004-2006	Formation Period: 2004-2006 Performance Period: 2007-2009
	Sharpe Ratio for:	
M	0.257	-0.071
S	0.250	-0.046
G(Class)	0.235	-0.047
G(Style)	0.218	-0.045
3F	0.277	-0.035
4F	0.276	-0.030
4F-Liquidity	0.259	-0.030

In the second column, documenting Sharpe ratios in 2007-2009, a similar picture emerges. The portfolios of the best funds in 2004-2006 according to the multi-factor models score higher Sharpe ratios in 2007-2009 than their counterparts based on the CAPM and other one-factor models.

The glimpse over performance in this section suggests that the CAPM is not only deviant relative to alternative models in its performance evaluation; it is also less successful in predicting future "winner" funds.

4   Summary and Conclusion

In this paper we compare 7 different benchmark models, some of which are novel or only recently developed, and evaluate their impact on estimating the fund manager's contribution, alpha, to the return of her investors. We use a sample of over 1000 non-specialized open-end equity funds' returns extracted from CRSP. The sample period is 2001-2009.

We begin by estimating funds' alpha over three-year subperiods using 7 benchmark models. Consistent with previous literature, we find that funds, on average, do not beat their benchmarks; Rather funds underperform the benchmark. Alpha's magnitude strongly depends on the benchmark model used for funds' performance evaluation, with CAPM alphas most deviant and least correlated with other benchmark method alphas.

When we compare the ranking and classification of funds based on the alphas generated by

the 7 benchmark models, significant classification and ranking differences emerge, with the traditional CAPM benchmark demonstrating the strongest differences relative to all other methods.

Last, we assessed and compared the out of sample performances of the "best funds" recommended by each of our 7 benchmark models. Consistent with our previous results, the out of sample Sharpe ratios of the funds recommended by multi-factor models are superior to that of the CAPM and the rest of the one-factor benchmarks.

We conclude that Jensen's alpha tends to be extreme in fund's performance evaluation. Multi-factor benchmark systems appear superior in performance evaluation not only because of their more elaborated theoretical basis, but also because of the less extreme and less volatile performance measures they provide. Also supporting this conclusion is our finding that in the out of sample period, the Sharpe ratios of the best funds recommended by the multi-factor models exceed the Sharpe ratios of the best funds recommended by the CAPM. Evidently, performance measurement, asset management and investment advice should also and perhaps mainly rely on the multi-factor models.

In light of the considerable diversion in alphas, fund classification and ranking exposed in this study, a second possible conclusion is that it is important for the investor consultants to attempt to identify and specify their client's benchmark model preferences before recommending any fund or portfolio.

References

[1]   B.N. Lehmann and D.M. Modest, Mutual fund performance evaluation: A comparison of benchmarks and benchmark comparisons, *Journal of Finance*, **42**, (1987), 233-265.

[2]   E.J. Elton, M.J. Gruber, D. Sanjiv and M. Hlavka, Efficiency with costly information: A reinterpretation of evidence from managed portfolios, *Review of Financial Studies*, **6**, (1993), 1-22.

[3]   M. Grinblatt and S. Titman, A study of monthly mutual fund returns and performance evaluation techniques,*Journal of Financial and Quantitative Analysis*, **29**, (1994), 419-444.

[4]   L.K.C. Chan, S.G. Dimmock and J. Lakonishok, Benchmarking money manager performance: Issues and evidence, *Review of Financial Studies*, **22**, (2009), 4553-4599.

[5]   A. Khorana, H. Servaes and P. Tufano, Mutual fund fees around the world,*Review of Financial Studies*, **22**, (2009), 1279-1310.

[6]   M.M. Carhart, On persistence in mutual funds performance,*Journal of Finance*, **52**, (1997), 57-82.

[7]   L. Pastor and R.F. Stambaugh, Liquidity risk and expected stock returns,*Journal of Political Economy*, **111**, (2003), 642-685.

[8] M.C. Jensen, The performance of mutual funds in the period 1945-1964,*Journal of Finance*, **23**, (1968), 389-416.

[9] M. Grinblatt and S. Titman, Mutual fund performance: An analysis of quarterly portfolio holdings,*Journal of Business*, **62**, (1989), 393-416.

Appendix

Limitations of the group-factor
A methodological issue occurs whenever using the group factor G in a regression, whether as the only factor or as an additional factor in a multi-factor framework.
First, assume that factor G is the only factor in the regression. Factor G's return is an equally weighted average of all N funds with the same classification C or with the same style S:

$$\widetilde{R}_G = \sum_{i=1}^{N} X_i \widetilde{R}_i = \frac{\sum_{i=1}^{N} \widetilde{R}_i}{N}.$$

For fund i, the regression equation on factor G is:

$$\widetilde{R}_i - R_f = \alpha_i + \beta_i(\widetilde{R}_G - R_f) + \tilde{\varepsilon}_i. \tag{1}$$

Summing up the N regression equations of the N funds with the same classification C (or style S) gives:

$$\sum_{i=1}^{N} \widetilde{R}_i - \sum_{i=1}^{N} R_f = \sum_{i=1}^{N} \alpha_i + \sum_{i=1}^{N} \beta_i(\widetilde{R}_G - R_f) + \sum_{i=1}^{N} \tilde{\varepsilon}_i.$$

Dividing the equation above by N gives:

$$\frac{\sum_{i=1}^{N} \widetilde{R}_i}{N} - R_f = \frac{\sum_{i=1}^{N} \alpha_i}{N} + \frac{\sum_{i=1}^{N} \beta_i}{N}(\widetilde{R}_G - R_f) + \frac{\sum_{i=1}^{N} \tilde{\varepsilon}_i}{N},$$

so
$$\widetilde{R}_G - R_f = \bar{\alpha} + \bar{\beta}(\widetilde{R}_G - R_f). \tag{2}$$

For fund i, $\beta_i = \frac{\text{cov}(\widetilde{R}_G, \widetilde{R}_i)}{\text{var}(\widetilde{R}_G)} = \frac{\text{cov}(\frac{\sum_{j=1}^{N} \widetilde{R}_j}{N}, \widetilde{R}_i)}{\text{var}(\widetilde{R}_G)}.$

Thus, the average beta is $\bar{\beta} = \frac{\sum_{i=1}^{N} \beta_i}{N} = \frac{\sum_{i=1}^{N} \text{cov}\left(\frac{\sum_{j=1}^{N} \widetilde{R}_j}{N}, \widetilde{R}_i\right)}{N\text{var}(\widetilde{R}_G)} = \frac{\text{cov}\left(\frac{\sum_{j=1}^{N} \widetilde{R}_j}{N}, \frac{\sum_{i=1}^{N} \widetilde{R}_i}{N}\right)}{\text{var}(\widetilde{R}_G)} = \frac{\text{cov}(\widetilde{R}_G, \widetilde{R}_G)}{\text{var}(\widetilde{R}_G)} = 1.$

Note that in equation (2) if $\bar{\beta} = 1$ then $\bar{\alpha} = 0$.