www.joim.com

Journal Of Investment Management, Vol. 23, No. 1, (2025), pp. 22–38

CAN UNDER-DIVERSIFICATION EXPLAIN THE SIZE EFFECT?

Moshe Levy[a]

None of the explanations suggested so far for the size anomaly seems to be consistent with the empirical evidence. This paper examines under-diversification as a possible explanation for the size effect. When the portfolio weight of a stock is non-negligible, its variance is priced. As small stocks are much more volatile than large stocks, this induces a size effect. We analytically derive the relation between under-diversification and the size premium, which allows us to estimate the magnitude of the under-diversification-induced size effect. We find it to be in close agreement with the empirically measured size effect.

1 Introduction

The size effect, also known also as the Small Firm Effect (SFE), is one of the earliest and most persistent anomalies in finance. Banz (1981) documents that the average returns of small US firms are substantially larger than those expected by the CAPM, given their betas. This finding has been extensively confirmed and extended (notably by Fama and French, 1992, 1995, 2012, 2015). The discovery of the SFE has attracted a great deal of interest by both academics trying to understand this phenomenon, as well as practitioners attempting to exploit it. Indeed, the discovery of the SFE has led to the creation of an entire new category of small-cap investment funds and indices (Reinganum, 1983).

[a]The Hebrew University, Jerusalem 91905, Israel. 972 25883219. E-mail: mslm@huji.ac.il

Additional research that accumulated over the following decades has made many scholars somewhat skeptical about the economic significance and magnitude of the SFE. Schwert (2003) suggests that the publication of the SFE, and the subsequent wave of new small-cap funds, have made the effect disappear. Indeed, the SFE has been found to be unstable over time, with decades where the SFE disappears or is even reversed. Moreover, the SFE has been found to be concentrated in January and in micro-cap stocks, making it difficult to exploit in practice. In addition, only little support for the SFE has been found in markets other than the US (Crain, 2011; Bryan, 2014).

This skeptical view of the SFE has changed rather dramatically with the publication of two influential studies. Asness *et al.* (2018) convincingly resurrect the SFE, showing that after controlling for "quality" (or its inverse, "junk"), the SFE is much larger and more robust than previously believed: it is stable over time, persisting after the publication of the effect in the early 1980s, it is not confined to January and micro-cap stocks, and it is observed in 24 different countries. Taking a somewhat different approach, Hou and Van Dijk (2019) reach the same conclusion about the robustness of the SFE. They show that after adjusting for the price impact of profitability shocks, the SFE remains robust and large even after the publication of the effect in the early 1980s.[1]

The remarkable persistence and magnitude of the SFE calls for a theoretical explanation. Asness *et al.* (2018) examine the various possible explanations that have been suggested in the literature, including those based on risk premiums, growth options, liquidity, infrequent trading, and behavioral biases. They conclude that none of these explanations provides a satisfactory answer to the SFE puzzle,[2] leading them to conclude that the SFE

"...*should be restored as one of the central cross-sectional empirical anomalies for asset pricing theory to explain*" (p. 508).

In this study we suggest under-diversification as a potential explanation for the SFE. There is vast empirical evidence indicating that most individual investors have significant holdings in a small number of stocks. Blume *et al.* (1974) find that 34.1% of investors in their sample directly held only one stock, 50% held two stocks or less, and only 10.7% held more than 10 stocks. Blume and Friend (1975) report an average number 3.41 stocks per portfolio. Barber and Odean (2000) report that individual investors hold on average 4.3 stocks, with a median number of 2.6 stocks. Goetzmann and Kumar (2008) find that 28% of investors directly hold only one stock, 60% of investors hold three stocks or less, and only 9.2% of investors hold more than 10 stocks. Kimball and Shumway (2010) find an average of 3.3 stocks held in the portfolio. Grinblatt *et al.* (2011) report a median of two stocks per portfolio in their sample. Phan *et al.* (2018) find that 60% of investors in their sample hold less than five stocks.[3] Possible reasons for under-diversification (or, more neutrally, "portfolio concentration") include transaction costs and mental costs (Levy, 1978), information asymmetry (Merton, 1987),[4] the cost of information acquisition (Van Nieuwerburgh and Veldkamp, 2009, 2010), and over-confidence and hubris (Daniel *et al.*, 1997; Phan *et al.*, 2018; Broekema and Kramer, 2021; Levy, 2023).

This paper shows that under-diversification induces a size premium, and quantifies this premium. The intuition for this effect is as based on the following four observations:

(1) A mean–variance investor measures risk by the variance (or standard deviation) of her entire portfolio.
(2) The riskiness of an individual stock held as part of the portfolio is measured by its marginal contribution to the portfolio's variance. In the standard CAPM, where all investors hold the market portfolio, this marginal contribution is measured by the stock's beta. In general, it is measure by the stock's generalized "beta" relative to the investor's specific portfolio.
(3) In well-diversified portfolios, where the stock's portfolio weight is small, the contribution of the stock's own variance to the portfolio variance is negligible. However, in under-diversified portfolios, the stock's variance may have a substantial contribution to the portfolio variance. Therefore, under-diversification implies that variance is priced, above and beyond the standard CAPM beta.
(4) Small stocks typically have much higher variances than large stocks: the average variance

of small stock (decile 10) is about four times larger than the variance of a large stock (decile 1, see Table 1).

Thus, under-diversification induces an SFE. The two main contributions of this paper are to formalize this intuition (especially point 3 above), and to estimate the magnitude of the induced SFE. Our focus is the size effect, and therefore we study this effect in isolation of other factors. But of course, the under-diversification-induced SFE may act in concert with other effects such as book-to-market, momentum, and quality.

There are two opposing views regarding the SFE. According to the first view, the SFE represents compensation for a risk factor that is not captured by the standard CAPM beta or other factors. According to the second view, the SFE is a mispricing anomaly – small stocks are underpriced, and investors can exploit this mispricing to achieve abnormal returns. We show that both views can be simultaneously correct. Investors holding under-diversified portfolios demand compensation for individual-stock variance risk. For them, the SFE represents risk compensation. At the same time, investors who are well-diversified can exploit the low pricing of small stocks: they tilt their portfolio weights toward these stocks and achieve risk-adjusted returns that are better than those of the market portfolio. In equilibrium, all investors hold optimal portfolio weights, given the set of stocks that they hold.

We should clarify at the outset that the under-diversification explanation is based on the variance of the stock's returns, *not* its idiosyncratic volatility, which is the variance of the residuals relative to a given asset pricing model, as discussed, for example, by Bali and Cakici (2008), Ang *et al.* (2009), Bekaert *et al.* (2012), and Stambaugh *et al.* (2015).[5] Also, the explanation suggested here is different than the "neglected stock" effect discussed by Levy (1978) and Merton (1987), which is based on the number of investors holding the stock (or the "investor base"). In our framework the SFE may arise even if *all* investors hold a given stock, but hold it in small portfolios (i.e. with only a few other stocks). Finally, the under-diversification explanation does not require all investors to hold small portfolios – clearly, many investors, both individual and institutional, hold well-diversified portfolios. It suffices that some of the investors have part of their holdings concentrated in a few stocks for the SFE to arise. Thus, our explanation is consistent with investments in well-diversified mutual funds, as long as some investors also invest directly in individual stocks.[6]

The next section describes the framework of analysis. Section 3 derives an analytical approximation for the under-diversification-induced SFE. In Section 4 we evaluate the size of the induced SFE with empirical data, and compare the predictions of the analytical approximation with the exact solution obtained via numerical analysis. We find that both yield similar results, and that the under-diversification-induced size effect is in close agreement with the empirically observed SFE. Section 5 concludes with a discussion of the theoretical and practical implications.

2 The Model

We employ the mean–variance framework. The departure of the present analysis from the standard CAPM is that we relax the assumption that all investors hold all available assets. Instead, for a variety of reasons (informational, psychological, transaction costs), some investors may hold only a subset of the available risky assets. The two main studies investigating this setup are the closely related General CAPM (GCAPM) of Levy (1978) and the Segmented-Market model of Merton (1987). Below we adopt the somewhat more general GCAPM framework.

There are n risky stocks and one risk-free asset. Investor k holds n_k stocks, where n_k may be smaller than n. Given the n_k stocks in his portfolio, the investor holds these stocks in the proportions that maximize his portfolio's Sharpe ratio, exactly as in the CAPM, where all available stocks are held. As in the CAPM, investors agree on the return parameters. Following the standard analysis (see, for example, Lintner, 1965 and Roll, 1977), this leads to the following equilibrium relationship:

$$\mu_i = \beta_{ik}\mu_k, \tag{1}$$

where μ_i denotes the expected return of stock i in excess of r_f, the risk-free rate, μ_k is the expected excess return of investor k's portfolio, and β_{ik} is the "beta" of stock i relative to investor k's portfolio, given by:

$$\begin{aligned}\beta_{ik} &\equiv \frac{\operatorname{Cov}(\tilde{r}_i, \tilde{R}_k)}{\sigma_k^2} \\ &= \frac{x_{ik}\sigma_i^2 + \sum_{j\neq i}^{n_k} x_{jk}\sigma_{i,j}}{x_{ik}^2\sigma_i^2 + \sum_{j\neq i}^{n_k} x_{jk}\sigma_j^2 + 2\sum_{\substack{j=1\\ q>j}}^{n_k} x_{jk}x_{qk}\sigma_{j,q}},\end{aligned} \tag{2}$$

where x_{ik} denotes the proportion of stock i in investor k's equity portfolio, $\tilde{r}_i$ is the stock i's stochastic return and $\tilde{R}_k$ is the return on investor k's portfolio (see Levy, 1978, Eq. (6)).[7] Note that we separate the variance terms in the denominator of Equation (2) into the variance of stock i and another term which includes all other variances. Equation (1) holds for all investors, each one with his own personal portfolio, i.e. with his own μ_k and β_{ik}. Denote the wealth of investor k invested in stocks by T_k, and the investor's wealth relative to the total wealth of all investors holding stock i by: $w_k \equiv \frac{T_k}{\sum_{s=1}^{N_i} T_s}$, where N_i is the number of investors who hold stock i in their portfolios, and the summation is only over these investors. Notice that by definition $\sum_{k=1}^{N_i} w_k = 1$. Multiplying Equation (1) by w_k and summing over all N_i investors who hold stock i we have:

$$\mu_i = \sum_{k=1}^{N_i} w_k\mu_k\beta_{ik}. \tag{3}$$

Let us define the generalized beta of stock i by:

$$\beta_i^* \equiv \sum_{k=1}^{N_i} w_k \frac{\mu_k}{\mu_m}\beta_{ik}, \tag{4}$$

where μ_m is the expected excess return of the market portfolio. Thus, Equation (3) becomes:

$$\mu_i = \beta_i^*\mu_m.^{8} \tag{5}$$

This is very similar to the standard SML, with the generalized β_i^*, the GCAPM risk index of stock i, replacing the standard CAPM β_i (recall that the μ's are expected returns in *excess* of the risk-free rate).[9] Notice that in the special case where all investors hold all stocks, as in the CAPM, we have $\mu_k = \mu_m$, β_i^* becomes equal to β_i, and Equation (5) reduces to the standard SML. Thus, the CAPM is obtained as a special case of the GCAPM.

In the CAPM, where investors hold a large number of stocks in their portfolios, the variance risk of stock i is "washed out", and does not command a risk premium. This is because σ_i^2's contribution to the standard CAPM β_i is negligible. This is also generally true in the GCAPM setting, if the number of stocks held in the portfolio, n_k, is large: the σ_i^2 term is only one of n_k terms in the numerator of β_{ik}, and only one of n_k^2 terms in its denominator (see Equation (2)). However, when the number of stocks in the investor's portfolio is small (or more precisely, when the investment proportion in the stock is substantial), the contribution of the individual stock's variance to β_{ik} may be substantial. As β_i^* is a weighted average of all the β_{ik}'s, it too is affected by the stock's variance, and hence the stock's variance may affect its equilibrium expected return. If a stock's β_i^* increases in its variance (all else equal), stocks with higher variances, which are typically smaller stocks, will have higher expected returns than larger stocks with lower variances and the same CAPM beta, leading to an equilibrium SFE.

Under-diversification generally leads to equilibrium asset pricing different than the CAPM pricing. The GCAPM equilibrium implies that if risk is (inappropriately) measured by the standard CAPM β_i (rather than by β_i^*), returns may seem to be anomalous. The equilibrium expected return of stock i is $\beta_i^* \mu_m$, while by the CAPM one would expect $\beta_i \mu_m$. The "excess return" of a stock, as measured by its Jensen's alpha, is thus given by:

$$\alpha_i = (\beta_i^* - \beta_i)\mu_m. \tag{6}$$

The SFE intuition described above implies that under-diversification induces a positive relationship between a firm's volatility (and thus size) and its alpha. Our goal is to formalize this intuition, which will also allow us to estimate the magnitude of the effect.

Equations (2)–(6) reveal that calculating the under-diversification-induced SFE requires full information about the subset of assets held by each investor, as well as her wealth. Obviously, this information is typically not available. Thus, an analytical derivation of the SFE requires some simplifying approximations. Fortunately, these approximations turn out to be quite reasonable. This is revealed by comparison to the exact results calculated numerically, as shown in Section 3. Theorem 1 below provides our main result:

Theorem 1. *The GCAPM β_i^* can be approximated by*:

$$\beta_i^* = \bar{x}_i \left(\frac{1}{1 + \frac{1}{n^*}\left(\frac{1-\rho}{\rho}\right)} \frac{\sigma_i^2}{\sigma_m^2} \right) + (1 - \bar{x}_i)\beta_i, \tag{7}$$

where $\bar{x}_i$ is the portfolio weight of stock i, averaged (wealth-weighted) across all investors who hold the stock, i.e. $\bar{x}_i \equiv \sum_{k=1}^{N_i} w_k x_{ik}$, $\frac{1}{n^*}$ is the wealth-weighted average of the inverse number of stocks in investors' portfolios $\frac{1}{n^*} \equiv \sum_{k=1}^{N_i} w_k \frac{1}{n_k}$, and ρ is the average correlation between stocks.

Proof. see Appendix.

Equation (7) shows that the GCAM β_i^* can be viewed as a weighted average of the standard CAPM β_i and the variance-dependent term $\left(\frac{1}{1+\frac{1}{n^*}\left(\frac{1-\rho}{\rho}\right)} \frac{\sigma_i^2}{\sigma_m^2}\right)$, where the weights of these two factors are determined by the average portfolio allocation, $\bar{x}_i$. In a CAPM world where all investors are well-diversified, we have $\bar{x}_i \approx 0$, and β_i^* converges to the standard CAPM β_i. On the other extreme, if one concentrates all of her holdings in one stock ($\bar{x}_i \approx 1$), the CAPM β_i becomes irrelevant, and risk is captured by σ_i^2. Equation (7) implies that if the market is in the GCAPM equilibrium, but one (inappropriately) measures risk by the standard CAPM beta, an SFE is observed. Namely, small (i.e. more volatile) firms will appear to yield an anomalous excess returns. The size of this induced SFE is:

$$\begin{aligned}\alpha_i &= (\beta_i^* - \beta_i)\mu_m \\ &= \left[\left(\frac{1}{1 + \frac{1}{n^*}\left(\frac{1-\rho}{\rho}\right)} \right) \frac{\sigma_i^2}{\sigma_m^2} - \beta_i \right] \bar{x}_i \mu_m. \end{aligned} \tag{8}$$

Equation (8) implies that volatility commands a premium: stocks with high return variance yield higher alphas (all else equal). As small stocks on average are much more volatile than large stocks (see Table 1), this implies an SFE. The magnitude of the effect is proportional to the average portfolio weight, $\bar{x}_i$: in a CAPM world with $\bar{x}_i \approx 0$, the SFE vanishes. The more under-diversified investors are, i.e. the larger $\bar{x}_i$, the larger the SFE. Finally, note that alphas can be negative for stocks with low volatility. Indeed, some alphas *must* be negative, as the value-weighted average alpha is by definition zero. However, as this is a value-weighted average, it is certainly possible that most stocks have positive alphas, and only a few large stocks have negative alpha.

The key question is whether the magnitude of the induced SFE is comparable with the empirically measured effect. In other words, how much of the empirically documented effect can be explained by under-diversification? To this question we turn next.

3 The Magnitude of the Under-Diversification-Induced SFE

Equations (7) and (8) allow us to estimate the magnitude of induced SFE by employing the empirical estimates of the parameters. These equations are based on several (rather plausible) approximations, as detailed in the Appendix. To examine the accuracy of these equations we compare them with the exact results, based on the direct solution to Equations (2)–(6). We find that both approaches yield similar results. In both cases, the under-diversification-induced SFE is very similar to the empirically measured effect.

3.1 Analytical estimation

To estimate the return parameters, we take all stocks in the CRSP file (share codes 10 and 11), and employ monthly returns over the January 1926–June 2018 period. At month t, for each stock we measure its average monthly return, standard deviation, and CAPM β (calculated against the returns of the S&P500 index), based on the preceding 120 monthly returns.[10] We rank all stocks by their time t market capitalizations and sort them into 10 size deciles. Decile compositions are updated monthly. Table 1 reports the equal-weighted average stock parameters by size decile.

First, note that the average standard deviation is monotonic in size. Small firms are much more volatile than large firms. The average standard deviation in decile 10 (16.4%) is almost double the average standard deviation in decile 1 (8.4%). This is consistent with previous findings in the literature (see, for example, Perez-Quiros and Timmermann, 2000), and implies that the variance of small firms is about four times larger than that of large firms. This relationship between size and volatility plays a key role in explaining the under-diversification-induced SFE.

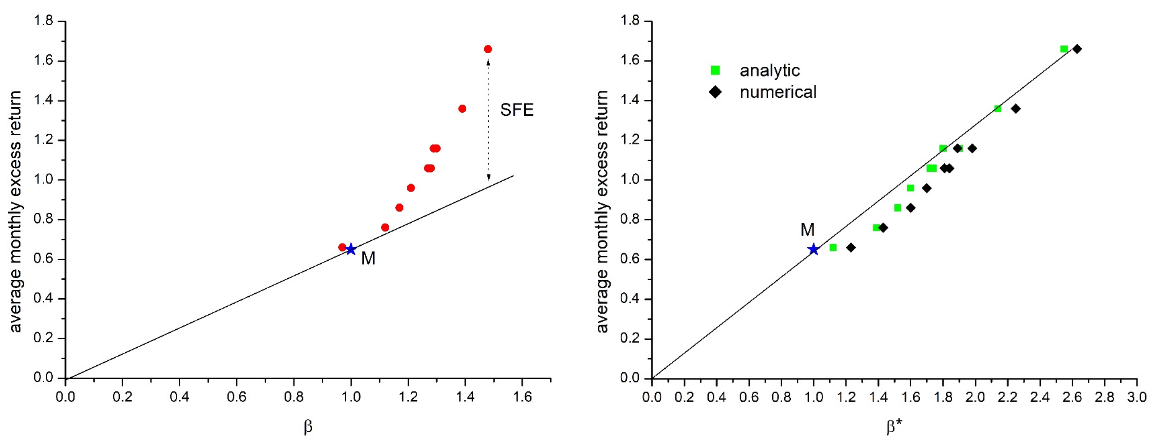
As expected, CAPM β's and average returns are also larger for small firms. However, the average returns of small firms are "too high" to be explained by their CAPM β's: this can be seen by the column in Table 1 reporting the empirical Jensen alphas, and by the top panel of Figure 1, depicting the average returns (in excess of the risk-free rate) as a function of the CAPM β. The magnitude of the SFE for decile 10 firms (about 0.7% per month) is comparable with the estimates reported by Asness *et al.* (2018).

To examine whether under-diversification can explain this large SFE, let us first calculate the size of the effect as reflected in the analytical approximations (7) and (8). μ_m and σ_m are taken as the average monthly return of the S&P500 index in excess of the risk-free rate, and its standard deviation, over our sample period. These are 0.65% and 5.32%, respectively. The average correlation ρ is taken as the sample average correlation between the monthly returns of all stock pairs in our sample, which is $\rho = 0.302$ (this is similar to the value reported by Levy, 2023). $\bar{x}_i$ is the wealth-weighted average portfolio weight of stock i, $\bar{x}_i \equiv \sum_{k=1}^{N_i} w_k x_{ik}$. Obviously, $\bar{x}_i$ may be different across stocks, as it depends on the other stocks included in the portfolios in which stock i is held. However, in our estimation we will make the approximation that these differences

Table 1 Risk and return parameters by size decile.

Size decile	Standard deviation (monthly, %)	Average excess return (monthly, %)	Beta: β CAPM	Beta: β^* Analytic	Beta: β^* Numerical	Alpha: Empirical	Alpha: GCAPM predicted analytic	Alpha: GCAPM predicted numerical
Largest 1	8.4	0.66	0.97	1.12	1.23	0.030	0.099	0.169
2	10.0	0.76	1.12	1.39	1.43	0.032	0.176	0.202
3	10.8	0.86	1.17	1.52	1.60	0.099	0.225	0.280
4	11.3	0.96	1.21	1.60	1.70	0.174	0.257	0.319
5	11.9	1.06	1.27	1.72	1.81	0.235	0.295	0.351
6	12.0	1.06	1.28	1.74	1.84	0.228	0.301	0.364
7	12.4	1.16	1.29	1.80	1.89	0.322	0.333	0.390
8	13.1	1.16	1.30	1.90	1.98	0.315	0.391	0.442
9	14.3	1.36	1.39	2.14	2.25	0.457	0.489	0.559
Smallest 10	16.4	1.66	1.48	2.55	2.63	0.698	0.693	0.748

Every month stocks are sorted into 10 size deciles according to their time t market capitalizations. The standard deviation and average return (in excess of the risk-free rate) are calculated as simple averages of all stocks in the size decile, over months t-1:t-120. β's are calculated in the same way: CAPM β's are calculated in the standard way, relative to the S&P500 index. GCAPM β^*'s are calculated both by the analytical approximation (7) and as the exact numerical solution to Equations (2) and (4). The numbers reported in the table are the averages across all 120-month periods in our 1926–2018 sample (1,002 periods).



(A) CAPM Risk-Return (B) GCAPM Risk-Return

Figure 1 The risk–return relationship for the 10 size deciles, when risk is measured by the standard CAPM β's (panel A) and by GCAPM β^*'s (panel B). Notice that returns are in excess of the risk-free rate, thus, the SML is expected to cross the origin. The market portfolio (M) is denoted by the star. When risk is measured by CAPM β's the well-known SFE is observed. Small stocks earn an average monthly return about 0.7% more than expected by the CAPM (see dashed arrow in panel A). However, when risk is measured by GCAPM β^*'s (panel B) the risk–return relationship is almost perfectly linear: $R^2 = 0.987$ for the β^*'s analytically approximated by Equation (7), and $R^2 = 0.992$ for the exact β^*'s calculated numerically. This conforms with the risk–return relationship of the GCAPM (Equation (5)).

tend to average out across portfolios, i.e. that if each portfolio contains n_k stocks, the average weight of each stock across all portfolios is $1/n_k$ (this assumption is relaxed in the exact numerical analysis described in the next sub-section). As stock i is held both in under-diversified portfolio and well-diversified portfolios (such as mutual funds and ETFs), the relative holdings of the stock by these two different types of portfolios must be considered, because the weight of each stock in well-diversified portfolios is close to zero. Gârleanu and Pedersen (2022) report that the direct holdings in US equities has declined from about 95% in the 1940s to about 37% today. We take the average value of 66%, and the typical empirically estimated median value of three stocks directly held in individual portfolios to obtain $\bar{x}_i = 0.66 \cdot \frac{1}{3} = 0.22$. Similarly, we take $\frac{1}{n^*} = 0.66 \cdot \frac{1}{3} + 0.34\frac{1}{\infty} = 0.22$.

Table 1 reports the values of the GCAPM β^* for each of the size deciles, analytically calculated by Equation (7). For large stocks, which have low volatility, β^* is similar to the standard CAPM β. However, for small stocks, which are much more volatile, β^* is much larger than β.

This implies that if the market is in a GCAPM equilibrium, but risk is (inappropriately) measured by the standard CAPM β, an SFE is observed. Table 1 shows the alphas generated by this effect, as calculated by Equation (8). These alphas are similar to those empirically measured. This can also be seen in panel B of Figure 1 (squares). This figure shows that if the empirical average excess returns are plotted against the GCAPM β^* calculated by Equation (7) an almost perfect linear fit is obtained ($R^2 = 0.987$).

The results in Table 1 are derived under the assumption that 66% of the portfolio is concentrated in $n = 3$ stocks, and the remaining 34% of the portfolio is held in well-diversified mutual funds. How does the magnitude of the under-diversification-induced SFE depend on the number of individual stocks held, n? Table 2 reports β^* and the alpha measured relative to the CAPM, as given in Equations (7) and (8), for three alternative values of n: 5, 20, and 100. All other parameters are the same as in Table 1. As expected, the induced SFE decreases with n. Table 1 shows that with $n = 3$ virtually all of the decile 10 alpha can be explained by under-diversification. Table 2 shows that with $n = 5$ about 72% of this alpha can be explained by under-diversification (the predicted alpha is 0.501, compared with the empirical value of 0.698). For $n = 20$ this is dramatically reduced to only 23% (0.159 compared with the empirical 0.698). For $n = 100$ the SFE almost completely vanishes. Thus, the number of stocks held in the concentrated part of the portfolio plays a central role in the analysis. The value of $n = 3$ employed in Table 1 is based on the empirical estimates. Table 2 tells us that if the average number of stocks held in the concentrated part would increase to $n = 20$, the size of the under-diversification-induced SFE would be only a quarter of the empirically observed SFE.

3.2 *Exact numerical estimation*

The analytical formula for the GCAPM β^*, Equation (7), is based on several simplifying approximations, as detailed in the appendix. To examine the accuracy of these approximations, we compare them with the exact results (given by Equations (2) and (4) and calculated numerically). To do so, we must specify a certain market structure, i.e. the wealth and subset of stocks held by each investor. We employ the simplest market structure possible: all investors are assumed to have the same wealth $T_k = T$. The value of T is chosen so that the total wealth invested in the market is equal to the empirical total market capitalization. Investors hold the same number of

Table 2 Risk and return parameters by size decile.

	Empirical		$n = 5$		$n = 20$		$n = 100$	
Size decile	β CAPM	Alpha	β^*	Predicted alpha	β^*	Predicted alpha	β^*	Predicted alpha
Largest 1	0.97	0.030	1.10	0.082	1.01	0.029	0.98	0.006
2	1.12	0.032	1.33	0.137	1.19	0.047	1.14	0.010
3	1.17	0.099	1.43	0.172	1.26	0.058	1.19	0.013
4	1.21	0.174	1.51	0.194	1.31	0.065	1.23	0.014
5	1.27	0.235	1.61	0.222	1.38	0.073	1.29	0.016
6	1.28	0.228	1.63	0.226	1.39	0.075	1.30	0.016
7	1.29	0.322	1.67	0.248	1.42	0.081	1.32	0.018
8	1.30	0.315	1.74	0.289	1.44	0.094	1.33	0.020
9	1.39	0.457	1.94	0.358	1.57	0.115	1.43	0.025
Smallest 10	1.48	0.698	2.25	0.501	1.72	0.159	1.53	0.034

Every month stocks are sorted into 10 size deciles according to their time t market capitalizations. The standard deviation and average return (in excess of the risk-free rate) are calculated as simple averages of all stocks in the size decile, over months t-1:t-120. β's are calculated in the same way: CAPM β's are calculated in the standard way, relative to the S&P500 index. GCAPM β^*'s are calculated both by the analytical approximation (7) and as the exact numerical solution to Equations (2) and (4). The numbers reported in the table are the averages across all 120-month periods in our 1926–2018 sample (1,002 periods).

stocks in their portfolios. For consistency with the analytical formula, where we had $\bar{x}_i = 0.22$, we take $n_k = 5$ (note that here we are modelling a market only with individual investors, without well-diversified mutual funds). The n_k stocks included in each investor's portfolio are drawn randomly; the probability of each stock being drawn is proportional to its market capitalization. Thus, large stocks are held by many investors, while small stocks are held by fewer investors (this is consistent with the empirical findings reported in Merton's (1987) Table 1). More details can be found in the online Appendix.

Table 1 and Figures 1–3 compare the GCAPM β^*'s and α's obtained in the exact numerical calculation with those obtained with the analytical approximations in Equations (7) and (8). They reveal close agreement between the two methods, which implies that the approximations used in deriving the analytical expressions (7) and (8) do not introduce a substantial error. The small firm β^*'s and α's are even a little larger in the exact calculation, meaning that the analytical expression may provide a slight underestimation of the induced SFE. When expected returns are regressed against the numerically calculated GCAPM β^*'s, again an almost perfect linear fit is obtained ($R^2 = 0.992$), consistent with the equilibrium prediction of the GCAPM (Equation (5)).

Figure 3 shows Jensen's α for each size decile relative to the predictions of the CAPM. The analytical approximation and the exact numerical solution are similar, and both are in good agreement with the empirically documented α's. The monthly "extra" return of small stocks relative to their CAPM β's is about 0.7%, consistent with previous estimates in the literature of the long-run SFE (see, for example, Fama and French, 1992, 1995, 2012; Asness *et al.*, 2018).

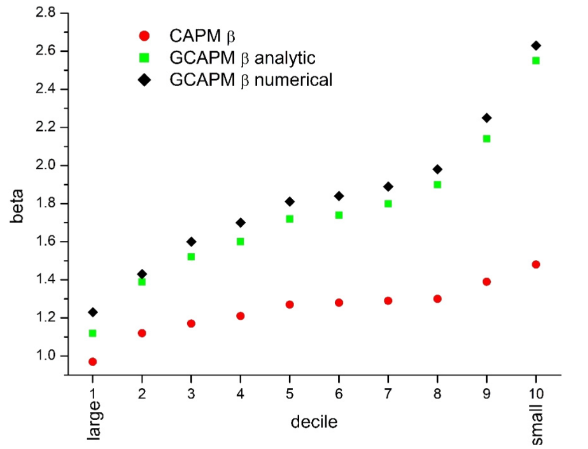


Figure 2 CAPM β's and GCAPM β^*'s for the 10 size deciles. There is close agreement between the GCAPM β^*'s obtained with the analytical approximation (7) and the exact values calculated numerically with Equations (2) and (4). The big difference between the GCAPM β^*'s and the CAPM β's for small firms induces the SFE.

We should emphasize that this numerical exercise, which is based on the empirical parameters, does not constitute an empirical test of the model. It only shows that the analytic expression for β^* and α in Equations (7) and (8) provide a good approximation for the exact values calculated numerically. Firm size, exposure to the SMB factor, return variance, and β^* are all closely related. However, the correlations between these factors are not perfect: there are small firms with low variance, and there are large firms with high variance. Similarly, not all firms with the same variance have the same β^*, as β^* also depends on the firm's CAPM β (see Equation (7)). To test the under-diversification explanation and compare it with other competing models, one could empirically examine which of the above factors best explains stock returns, for example by following the methodology of Daniel and Titman (1997) or Avramov and Chordia (2006). This analysis is beyond the scope of the present paper, but seems to be a promising path for further investigation.

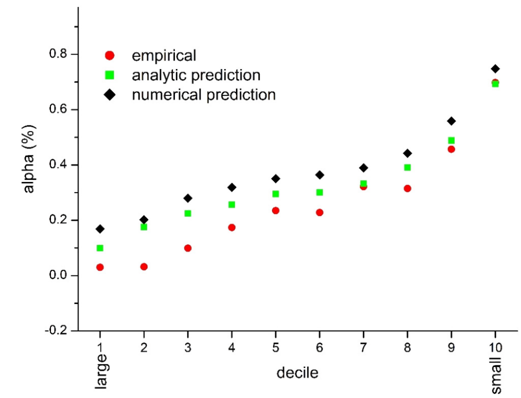


Figure 3 If investors are under-diversified and the market is in the GCAPM equilibrium, but risk is (inappropriately) measured by CAPM β's, small firms appear to yield anomalously high returns. Jensen α's theoretically predicted by this effect (squares and diamonds) are close to the empirically measured α's (circles).

4 Conclusions and Implications

The Small Firm Effect is considered to be one of the most central and most persistent anomalies which contradicts market efficiency. Some studies have argued that the SFE has vanished, or that it is limited to micro-cap stocks or only to January. However, the two comprehensive studies by Asness *et al.* (2018) and Hou and Van Dijk (2019) convincingly resurrect the SFE. Asness *et al.* conclude:

"*Our results revive the size anomaly, putting it on a more equal footing with other anomalies such as value and momentum in terms of its efficacy, and dismiss several previous explanations and challenges to the size effect*".

If the SFE represents an abnormal profit opportunity, we would expect it to be exploited, and to vanish. The robustness of the SFE, as empirically documented, suggests that the SFE represents a compensation for risk that conventional models such as the CAPM do not capture. The present paper shows that under-diversification offers an explanation for the SFE: when investors are under-diversified, return variance is priced (above and beyond CAPM β), and as small firms are much more volatile than large firms, this induces the SFE. We derive an analytical expression for the under-diversification-induced SFE. Estimates of the under-diversification-induced SFE reveal that the theoretically predicted effect is in close agreement with the empirically measured SFE.

Over the decades since the discovery of the SFE, scholars have debated whether the "extra" return to small firms is compensation for some form of non-CAPM-beta risk, or a mispricing anomaly that can be exploited by small-cap funds. According to the under-diversification explanation the answer is: *yes to both!* Under-diversified investors require a risk premium for small (and volatile) stocks because of their variance-risk, but for well-diversified investors these stocks represent an opportunity for abnormal (i.e. above-market) performance. Thus, in a market with both fully-diversified investors and under-diversified investors, equilibrium prices are simultaneously consistent with both views.

It has been well-documented that many investors are under-diversified, most of them holding only 1–5 stocks (Blume *et al.*, 1974; Blume and Friend, 1975; Levy, 1997; Barber and Odean, 2000; Goetzmann and Kumar, 2008). Several rational reasons have been suggested as explanations for this phenomenon, including asymmetric information (Merton, 1987), transaction costs, and the mental cost of tracking many firms (Levy, 1978). However, given the large economic cost of under-diversification (which Levy (2023) estimates at about 1% per annum), and the typical poor performance of individual investors (Barber and Odean, 2000), it is hard to justify under-diversification rationally. It seems more likely that the source of under-diversification lies in the realm of investor psychology, involving over-confidence and hubris. Over-confidence has been suggested to be advantageous in some evolutionary conditions (Charness *et al.*, 2018; Schwardmann and Van der Weele, 2019), and may therefore be genetically hard-wired (Johnson and Fowler, 2011). If this is the case, over-confidence may be hard to overcome.

Over the years, the role of mutual funds and ETFs in the market has dramatically increased. The proportion of equity directly held by individual investors in the US was about 95% in the 1940s. It has since decreased substantially, with the rising popularity of mutual funds, ETFs and hedge funds. However, the proportion of direct holdings by individuals is still substantial, at about 37% today (Gârleanu and Pedersen, 2022). Should the under-diversification-induced SFE decrease with the increasing popularity of diversified funds? The answer is not obvious. On the one hand, funds tend to be well-diversified. On the other hand, they tend to concentrate on large and medium stocks, and to avoid very small stocks. Thus, as the proportion of equities held by funds increases, the β^*'s of large and medium firms become closer to their CAPM β's. However, this is not so for small firms. Thus, the premium for small stocks *relative* to large stock may actually increase due to the existence of mutual funds. Therefore, it seems that as long as a substantial percentage of equities will continue to be held by under-diversified individual investors, the Small Firm Effect will remain large and robust.

Appendix. Proof of Theorem 1

In order to derive Equation (7) for β_i^* we employ the following three approximations:

- $$\sum_{k=1}^{N_i} w_k \left(\frac{\mu_k}{\sigma_k^2}\right)(x_{ik} - x_{ik}^2\beta_{ik}) \approx \left(\sum_{k=1}^{N_i} w_k \frac{\mu_k}{\sigma_k^2}\right)\left(\sum_{k=1}^{N_i} w_k \left(x_{ik} - x_{ik}^2\beta_{ik}\right)\right) \tag{A.1}$$

- $$\sum_{k=1}^{N_i} w_k x_{ik}^2 \beta_{ik} \ll \sum_{k=1}^{N_i} w_k x_{ik} \qquad \text{(A.2)}$$

- $$\frac{\sigma_m^2}{\mu_m}\left(\sum_{k=1}^{N_i} w_k \frac{\mu_k}{\sigma_k^2}\right) \approx \frac{1}{1+\frac{1}{n^*}\left(\frac{1-\rho}{\rho}\right)}, \qquad \text{(A.3)}$$

where $\frac{1}{n^*}$ is the wealth-weighted average of the inverse number of stocks in the portfolio, $\frac{1}{n^*} \equiv \sum_{k=1}^{N_i} w_k \frac{1}{n_k}$, and ρ is the average correlation between stock pairs. Let us first discuss the rationale for these approximations, and then employ them to derive Equation (7). Below we discuss the economic logic of each approximation. Our exact numerical analysis confirms that these approximations are reasonable.

Approximation (A.1) states that the term $\frac{\mu_k}{\sigma_k^2}$ is approximately uncorrelated with the term $(x_{ik} - x_{ik}^2\beta_{ik})$ across the portfolios of all investors holding stock i, and thus, the (wealth-weighted) average of their multiplication is approximately equal to the average of one times the average of the other. Why are these two expressions approximately uncorrelated? $\frac{\mu_k}{\sigma_k^2}$ and x_{ik} are very complicated and very different functions of the parameters of all stocks included in the portfolio.[11] Indeed, in the numerical study described in Section 3.2, which is conducted with the empirical stock return parameters, we find that the sample correlation between $\frac{\mu_k}{\sigma_k^2}$ and x_{ik} across investors is less than 0.001, and the sample correlation between $\frac{\mu_k}{\sigma_k^2}$ and $x_{ik}^2\beta_{ik}$ is only 0.03. Both sample correlations are not significantly different from zero, which is consistent with approximation (A.1).

Approximation (A.2) essentially states that the average portfolio weight of a stock is much smaller than 1 (even in small portfolios), and therefore it is much larger than the average of its square value (betas are in the order of 1, so they do not have a big impact on this inequality). For example, if there are five stocks in each portfolio, we expect the average holding of each stock (averaged across all portfolios) to be in the order of 0.20, and the average value of the square holding to be in the order of 0.04. These values are very close to those recorded in the exact numerical solution, where we find $\sum_{k=1}^{N_i} w_k x_{ik} = 0.20$ and $\sum_{k=1}^{N_i} w_k x_{ik}^2 \beta_{ik} = 0.043$.

Approximation (A.3) is based on two intuitions. The first is that the expected return of a subset of randomly selected stocks, averaged over many such subsets, is close to the market portfolio's expected return, i.e. $\sum_{k=1}^{N_i} w_k \mu_k \approx \mu_m$. The second is that the average ratio between the variance of a well-diversified portfolio and the variance of an under-diversified portfolio can be expressed as a function of the number of stocks held in the under-diversified portfolio, n_k, and the correlation between stocks, ρ. Let us elaborate.

Levy (2023) shows that a naïve symmetrical model provides a surprisingly good approximation for the average variance of a portfolio of n_k randomly-drawn stocks. Namely, the average variance of many such portfolios can be approximated by:

$$\sigma_P^2 \approx \frac{\sigma^2}{n_k} + \frac{n_k(n_k-1)}{n_k^2}\rho\sigma^2, \qquad \text{(A.4)}$$

where σ_P^2 is the average portfolio variance, σ^2 is the average variance of individual stocks, and ρ is the average correlation. Employing this approximation, the average ratio between the variance of a well-diversified portfolio ($n_k \to \infty, \sigma_P^2 \to \rho\sigma^2$) and an under-diversified n-stock portfolio can be approximated by:

$$\frac{\sigma_m^2}{\sigma_k^2} \approx \frac{\rho\sigma^2}{\frac{\sigma^2}{n_k} + \frac{n_k(n_k-1)}{n_k^2}\rho\sigma^2}$$

$$= \frac{\rho}{\rho + \frac{1}{n_k}(1-\rho)}$$

$$= \frac{1}{1 + \frac{1}{n_k}\left(\frac{1-\rho}{\rho}\right)}. \qquad \text{(A.5)}$$

If investors hold different numbers of stocks in their portfolio, the weighted average of $\frac{\sigma_m^2}{\sigma_k^2}$ can be approximated as:

$$\sum_{k=1}^{N_i} w_k \frac{\sigma_m^2}{\sigma_k^2} \approx \frac{1}{1 + \frac{1}{n^*}\left(\frac{1-\rho}{\rho}\right)}, \qquad \text{(A.6)}$$

where $\frac{1}{n^*} \equiv \sum_{k=1}^{N_i} w_k \frac{1}{n_k}$. This is obviously not mathematically precise, but can serve as a good first-order approximation. In the exact numerical solution the average value of $\frac{\sigma_m^2}{\mu_m}\left(\sum_{k=1}^{N_i} w_k \frac{\mu_k}{\sigma_k^2}\right)$ is 0.612, while the approximation $\frac{1}{1+\frac{1}{n^*}\left(\frac{1-\rho}{\rho}\right)}$ with $n^* = 5$ and the empirical average correlation of $\rho = 0.302$ yields 0.684.

Let us now prove that approximations (A.1)–(A.3) imply Equation (7). First, let us derive β_{ik} in Equation (2) with respect to σ_i^2:

$$\frac{\partial \beta_{ik}}{\partial \sigma_i^2} = \frac{x_{ik}\sigma_k^2 - x_{ik}^2(\beta_{ik}\sigma_k^2)}{\sigma_k^4}$$

$$= \frac{x_{ik}}{\sigma_k^2} - \frac{x_{ik}^2 \beta_{ik}}{\sigma_k^2}, \qquad \text{(A.7)}$$

where we employ the relation: $x_{ik}\sigma_i^2 + \sum_{j\neq i}^{N} x_{jk}\sigma_{i,j} = \beta_{ik}\sigma_k^2$ (see Equation (2)). β_i^* is the weighted-average of β_{ik} across all investors holding stock i (see Equation (4)), and therefore:

$$\frac{\partial \beta_i^*}{\partial \sigma_i^2} = \sum_{k=1}^{N_i} w_k \frac{\mu_k}{\mu_m} \frac{\partial \beta_{ik}}{\partial \sigma_i^2}$$

$$= \frac{1}{\mu_m} \sum_{k=1}^{N_i} \left(\frac{\mu_k}{\sigma_k^2}\right) [w_k(x_{ik} - x_{ik}^2 \beta_{ik})]. \qquad \text{(A.8)}$$

Employing Approximation (A.1) yields:

$$\frac{\partial \beta_i^*}{\partial \sigma_i^2} = \frac{1}{\mu_m}\left(\sum_{k=1}^{N_i} w_k \frac{\mu_k}{\sigma_k^2}\right) \times \sum_{k=1}^{N_i} [w_k(x_{ik} - x_{ik}^2 \beta_{ik})]. \qquad \text{(A.9)}$$

Approximation (A.2) allows us to neglect the term involving x_{ik}^2, yielding:

$$\frac{\partial \beta_i^*}{\partial \sigma_i^2} \approx \frac{1}{\mu_m}\left(\sum_{k=1}^{N_i} w_k \frac{\mu_k}{\sigma_k^2}\right) \sum_{k=1}^{N_i} w_k x_{ik}$$

$$= \frac{1}{\mu_m}\left(\sum_{k=1}^{N_i} w_k \frac{\mu_k}{\sigma_k^2}\right) \bar{x}_i \qquad \text{(A.10)}$$

(recall that by definition $\bar{x}_i \equiv \sum_{k=1}^{N_i} w_k x_{ik}$), which can be rewritten as:

$$\frac{\partial \beta_i^*}{\partial \sigma_i^2} \approx \frac{1}{\sigma_m^2} \frac{\sigma_m^2}{\mu_m} \left(\sum_{k=1}^{N_i} w_k \frac{\mu_k}{\sigma_k^2}\right) \bar{x}_i. \qquad \text{(A.11)}$$

Employing Approximation (A.3) yields:

$$\frac{\partial \beta_i^*}{\partial \sigma_i^2} \approx \frac{1}{\sigma_m^2} \bar{x}_i \left(\frac{1}{1 + \frac{1}{n^*}\left(\frac{1-\rho}{\rho}\right)}\right). \qquad \text{(A.12)}$$

As the derivative of β_i^* with respect to σ_i^2 is a constant, this implies that β_i^* is a linear function of σ_i^2 which can be expressed as:

$$\beta_i^* = \bar{x}_i \left(\frac{1}{1 + \frac{1}{n^*}\left(\frac{1-\rho}{\rho}\right)}\right) \frac{\sigma_i^2}{\sigma_m^2} + C, \qquad \text{(A.13)}$$

where C is an integration constant. C that can be determined form economic considerations: we know that in the CAPM equilibrium, where $\bar{x}_i \approx 0$, β_i^* should converge to the standard CAPM β_i. With $\bar{x}_i \approx 0$ the first term in Approximation (A.12) vanishes, and therefore C should converge to β_i when $\bar{x}_i \to 0$. On the other hand, when the investor hold only one stock, she cares about the stock's variance, and does not care at all about its CAPM beta. Thus, when $\bar{x}_i \rightarrow 1\, C$ must converge to 0. These two conditions dictate that $C = (1 - \bar{x}_i)\beta_i$, which finally yields:

$$\beta_i^* = \bar{x}_i \left(\frac{1}{1 + \frac{1}{n^*}\left(\frac{1-\rho}{\rho}\right)} \frac{\sigma_i^2}{\sigma_m^2} \right) + (1 - \bar{x}_i)\beta_i.$$

Notations

μ_k – the expected rate of return of investor k's portfolio, in excess of the risk-free rate

σ_k – the standard deviation of investor k's portfolio

x_{ik} – the proportion of stock i in investor k's equity

β_i – the standard CAPM beta of stock i

β_{ik} – the beta of stock i relative to investor k's portfolio (Equation (2))

β_i^* – the GCAPM beta of stock i (Equation (4))

N_i – the number of investors who hold stock i in their portfolios

T_k – investor k's wealth invested in the stock market

w_k – investor k's wealth divided by the total wealth of all investors who hold stock i

n – the total number of stocks in the market

n_k – the number of stocks held by investor k

n^* – the solution to $\frac{1}{n^*} \equiv \sum_{k=1}^{N_i} w_k \frac{1}{n_k}$

ρ – the average correlation between stocks

Endnotes

1 Guo (2023) discusses the size effect conditional on information about the market.

2 The econometric horizon-based explanation suggested by Levy and Levy (2011), based on the systematic bias of monthly betas relative to annual betas, can provide only a partial explanation to the SFE. Levy (2024) discusses estimation error as a possible explanation.

3 For a review of under-diversification, see Barber and Odean (2013). In an experiment conducted by Levy (1997) students had to choose a portfolio out of 20 available stocks. Despite the fact that there were no transaction costs, the average portfolio included only 4.9 stocks and the median was 3.2 stocks.

4 Several studies empirically document that concentrated portfolios outperform more diversified portfolios (Kacperczyk *et al.*, 2005; Ivkovic *et al.*, 2008; Van Nieuwerburgh and Veldkamp, 2010; Huij and Derwall, 2011; Fulkerson and Riley, 2019; McKay *et al.*, 2018; Brown *et al.* 2020). This may be viewed as evidence supporting the asymmetric information explanation for under-diversification.

5 Merton (1987) discusses the idiosyncratic volatility, and shows that its effect on the SFE may be ambiguous (see also Boehme *et al.*, 2009). Goyal and Santa-Clara (2003) find a positive relation between average stock volatility and market returns. Ang *et al.* (2006) discuss the stock's sensitivity to changes in the market-wide volatility.

6 Typically, under-diversified portfolios are held by individual investors (but Shawky and Smith, 2005 show that many mutual funds also hold a relatively small number of stocks). Direct stock ownership in the US has decreased from about 95% in the 1940s to about 37% today (see Figure 1 in Gârleanu and Pedersen, 2022). However, it is interesting to note that in the past decade the proportion of direct stock ownership has systematically *increased*.

7 Notations are explained as they are introduced, but for convenience, a list of all notations is also provided in the Appendix.

8 Our definition of β_i^* is slightly different than Levy's (1978) in his Equations ((33)–(35)). The advantage of the present formulation is that it yields the simple market-wide risk–return relationship (5), while in Levy's risk return formulation, $\mu_i = r_f + \gamma_{1i}^* \beta_i^*$ (his Eq. (35)), the coefficient γ_{1i}^* is generally different for different stocks.

9 It is well known that in the CAPM framework we have $\sum_{i=1}^n x_i \beta_i = \beta_m = 1$, where x_i is the weight of stock i in the market portfolio, the summation is over all stocks in the market, and m denotes the market portfolio. The same result holds in the GCAPM framework. Multiplying both sides of Equation (5) by the market weights x_i and summing over all n stocks in the market we obtain $\sum_{i=1}^n x_i\, \mu_i = \mu_m \sum_{i=1}^n x_i \beta_i^*$. As by definition we have $\sum_{i=1}^n x_i\, \mu_i = \mu_m$, this implies that $\sum_{i=1}^n x_i \beta_i^* = 1$. Given the market values of all firms, the difference between the CAPM and the GCAPM in this respect is in the division of the total risk (or macroeconomic risk) between the various stocks. Technically, it implies that if for some stocks $\beta_i^* > \beta_i$ it must be that there are other stocks for which $\beta_i^* < \beta_i$. Note that if for a large stock with large x_i we obtain that $\beta_i^* < \beta_i$, we may find

many small stocks with $\beta_i^* > \beta_i$, without violating the constraint $\sum_{i=1}^{n} x_i \beta_i^* = 1$.

[10] We do so only for stocks with complete return records over the previous 120 months, which introduces a potential survivorship bias. This bias is not relevant for the relationship between the CAPM β and the GCAPM β^* (in its two versions: analytical approximation and exact numeric), as these are all calculated for the same set of stocks. Moreover, as small stocks are less likely to survive, the survivorship bias potentially strengthens the empirically measured SFE – small firms appear to be better than they actually are. Below we show that even with this potential bias, under-diversification can explain the observed SFE. Thus, the potential bias only strengthens our results.

[11] For example, Levy and Ritov (2011) find that a stock's mean–variance optimal weight is almost unrelated to its mean return, its variance, or its average correlation with the other stocks included in the portfolio.

Acknowledgment

I am grateful to Gifford Fong, the Editor, and to the anonymous referee for many helpful comments and suggestions.

References

Ang, A., Hodrick, R. J., Xing, Y., and Zhang, X. (2006). "The Cross-Section of Volatility and Expected Returns," *The Journal of Finance* **61**(1), 259–299.

Ang, A., Hodrick, R. J., Xing, Y., and Zhang, X. (2009). "High Idiosyncratic Volatility and Low Returns: International and Further US Evidence," *Journal of Financial Economics* **91**(1), 1–23.

Asness, C., Frazzini, A., Israel, R., Moskowitz, T. J., and Pedersen, L. H. (2018). "Size Matters, if You Control Your Junk," *Journal of Financial Economics* **129**(3), 479–509.

Bali, T. G. and Cakici, N. (2008). "Idiosyncratic Volatility and the Cross Section of Expected Returns," *Journal of Financial and Quantitative Analysis* **43**(1), 29–58.

Banz, R. W. (1981). "The Relationship Between Return and Market Value of Common Stocks," *Journal of Financial Economics* **9**(1), 3–18.

Barber, B. M. and Odean, T. (2000). "Trading Is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors," *The Journal of Finance* **55**(2), 773–806.

Barber, B. M. and Odean, T. (2013). "The Behavior of Individual Investors," In *Handbook of the Economics of Finance*, Vol. 2 (Elsevier), pp. 1533–1570.

Bekaert, G., Hodrick, R. J., and Zhang, X. (2012). "Aggregate Idiosyncratic Volatility," *Journal of Financial and Quantitative Analysis* **47**(6), 1155–1185.

Blume, M. E., Crockett, J., and Friend, I. (1974). "Stock Ownership in the United States: Characteristics and Trends," *Survey of Current Business* **54**(11), 16–40.

Blume, M. and Friend, I. (1975). "The Asset Structure of Individual Portfolios and Some Implications for Utility Functions," *The Journal of Finance* **30**(2), 585–603.

Blume, M. and Friend, I. (1986). *Recent and Prospective Trends in Institutional Ownership and Trading of Exchange and OTC Stocks* by Rodney L. White Center for Financial Research, the Wharton School, University of Pennsylvania.

Boehme, R. D., Danielsen, B. R., Kumar, P., and Sorescu, S. M. (2009). "Idiosyncratic Risk and the Cross-section of Stock Returns: Merton (1987) Meets Miller (1977)," *Journal of Financial Markets* **12**(3), 438–468.

Brown, K. C., Tiu, C., and Yoeli, U. (2020). "The Decision to Concentrate: Active Management, Manager Skill, and Portfolio Size," *The Journal of Portfolio Management* **46**(5), 41–62.

Broekema, S. P. and Kramer, M. M. (2021). "Overconfidence, Financial Advice Seeking and Household Portfolio Under-Diversification," *Journal of Risk and Financial Management* **14**(11), 553.

Bryan, A. (2014). *Does the Small Cap Premium Even Exist?* Chicago IL: Morningstar Inc.

Campbell, J. Y. and Vuolteenaho, T. (2004). "Bad Beta, Good Beta," *American Economic Review* **94**, 1249–1275.

Charness, G., Rustichini, A., and Van de Ven, J. (2018). "Self-confidence and Strategic Behavior," *Experimental Economics* **21**, 72–98.

Crain, M. A. (2011). A Literature Review of the Size Effect. Working Paper.

Daniel, K., Hirshleifer, D., and Subrahmanyam, A. (1997). A Theory of Overconfidence, Self-attribution, and Security Market Under- and Over Reactions. http://dx.doi.org/10.2139.

Daniel, K. and Titman, S. (1997). "Evidence on the Characteristics of Cross Sectional Variation in Stock Returns," *The Journal of Finance* **52**(1), 1–33.

Fama, E. F. and French, K. R. (1992). "The Cross-section of Expected Stock Returns," *The Journal of Finance* **47**(2), 427–465.

Fama, E. F. and French, K. R. (1995). "Size and Book-to-Market Factors in Earnings and Returns," *The Journal of Finance* **50**(1), 131–155.

Fama, E. F. and French, K. R. (2012). "Size, Value, and Momentum in International Stock Returns," *Journal of Financial Economics* **105**(3), 457–472.

Fama, E. F. and French, K. R. (2015). "A Five-Factor Asset Pricing Model," *Journal of Financial Economics* **116**(1), 1–22.

Feldstein, M. (1968). "Mean Variance Analysis in the Theory of Liquidity Preference and Portfolio Selection," *Review of Economic Studies* **36**(1), 5–12.

Fulkerson, J. A. and Riley, T. B. (2019). "Portfolio Concentration and Mutual Fund Performance," *Journal of Empirical Finance* **51**, 1–16.

Gârleanu, N. and Pedersen, L. H. (2022). "Active and Passive Investing: Understanding Samuelson's Dictum," *The Review of Asset Pricing Studies* **12**(2), 389–446.

Goetzmann, W. N. and Kumar, A. (2008). "Equity Portfolio Diversification," *Review of Finance* **12**(3), 433–463.

Goyal, A. and Santa-Clara, P. (2003). "Idiosyncratic Risk Matters!," *The Journal of Finance* **58**(3), 975–1007.

Grinblatt, M., Keloharju, M. and Linnainmaa, J. (2011). "IQ and Stock Market Participation," *The Journal of Finance* **66**(6), 2121–2164.

Guo, L. (2023). "Two Faces of the Size Effect," *Journal of Banking & Finance* **146**, 106708.

Hou, K. and Van Dijk, M. A. (2019). "Resurrecting the Size Effect: Firm Size, Profitability Shocks, and Expected Stock Returns," *The Review of Financial Studies* **32**(7), 2850–2889.

Huij, J. and Derwall, J. (2011). "Global Equity Fund Performance, Portfolio Concentration, and the Fundamental Law of Active Management," *Journal of Banking & Finance* **35**(1), 155–165.

Ivković, Z., Sialm, C., and Weisbenner, S. (2008). "Portfolio Concentration and the Performance of Individual Investors," *Journal of Financial and Quantitative Analysis* **43**(3), 613–655.

Johnson, D. D. and Fowler, J. H. (2011). "The Evolution of Overconfidence," *Nature* **477**(7364), 317–320.

Kacperczyk, M., Sialm, C., and Zheng, L. (2005). "On the Industry Concentration of Actively Managed Equity Mutual Funds," *The Journal of Finance* **60**(4), 1983–2011.

Kimball, M. S. and Shumway, T. (2010). "Investor Sophistication and the Home Bias, Diversification, and Employer Stock Puzzles," http://dx.doi.org/10.2139/ssrn.1572866.

Levy, H. (1978). "Equilibrium in an Imperfect Market: A Constraint on the Number of Securities in the Portfolio," *The American Economic Review* **68**(4), 643–658.

Levy, H. (1997). "Risk and Return: An Experimental Analysis," *International Economic Review*, 119–149.

Levy, H. and Levy, M. (2011). "The Small Firm Effect: A Financial Mirage?," *Journal of Portfolio Management* **37**(2), 129.

Levy, M. (2023). "The Cost of Investment Hubris," *Journal of Portfolio Management* **50**(1), 1–16.

Levy, M. (2024). "Shrinking the Size Effect," *Journal of Portfolio Management*, **50**(8).

Levy, M. and Ritov, Y. A. (2011). "Mean–Variance Efficient Portfolios with Many Assets: 50% Short," *Quantitative Finance* **11**(10), 1461–1471.

Lintner, J. (1965). "Security Prices, Risk, and Maximal Gains from Diversification," *The Journal of Finance* **20**(4), 587–615.

Markowitz, H. M. (1991). *Portfolio Selection: Efficient Diversification of Investment* (New Haven: Yale University Press).

McKay, S., Shapiro, R., and Thomas, R. (2018). "What Free Lunch? The Costs of Overdiversification," *Financial Analysts Journal* **74**(1), 44–58.

Merton, R. C. (1987). "A Simple Model of Capital Market Equilibrium with Incomplete Information," *The Journal of Finance* **42**(3), 483–510.

Perez-Quiros, G. and Timmermann, A. (2000). "Firm Size and Cyclical Variations in Stock Returns," *The Journal of Finance* **55**(3), 1229–1262.

Phan, T. C., Rieger, M. O., and Wang, M. (2018). "What Leads to Overtrading and Under-diversification? Survey Evidence from Retail Investors in an Emerging Market," *Journal of Behavioral and Experimental Finance* **19**, 39–55.

Reinganum, M. R. (1983). "The Anomalous Stock Market Behavior of Small Firms in January: Empirical Tests for Tax-Loss Selling Effects," *Journal of Financial Economics* **12**(1), 89–104.

Roll, R. (1977). "A Critique of the Asset Pricing Theory's Tests Part I: On Past and Potential Testability of the Theory," *Journal of Financial Economics* **4**(2), 129–176.

Schwardmann, P. and Van der Weele, J. (2019). "Deception and Self-deception," *Nature Human Behaviour* **3**(10), 1055–1061.

Schwert, G. W. (2003). "Anomalies and Market Efficiency," *Handbook of the Economics of Finance* **1**, 939–974.

Shawky, H. A. and Smith, D. M. (2005). “Optimal Number of Stock Holdings in Mutual Fund Portfolios Based on Market Performance,” *Financial Review* **40**(4), 481–495.

Stambaugh, R. F., Yu, J., and Yuan, Y. (2015). “Arbitrage Asymmetry and the Idiosyncratic Volatility Puzzle,” *The Journal of Finance* **70**(5), 1903–1948.

Van Nieuwerburgh, S. and Veldkamp, L. (2009). “Information Immobility and the Home Bias Puzzle,” *The Journal of Finance* **64**(3), 1187–1215.

Van Nieuwerburgh, S. and Veldkamp, L. (2010). “Information Acquisition and Under-diversification,” *The Review of Economic Studies* **77**(2), 779–805.

Keywords: Size anomaly; Small Firm Effect; under-diversification; segmented-market model; Generalized CAPM; generalized beta.

JEL Classification: G11, G12