The Galton Correction in China: When Forecasting Learns from the OOS

Maria Roque

Dissertation written under the supervision of Professor Pedro Barroso

Dissertation submitted in partial fulfilment of requirements for the MSc in Finance, at the Universidade Católica Portuguesa, January 2023.

Abstract

In the present dissertation, I apply the Barroso and Saxena (2022) shrinkage methodology – also known as the Galton – to portfolio optimization's inputs in the Chinese market (for the OOS period from January of 2011 until December of 2021). Each portfolio is a selection of the 50 biggest stocks in terms of market capitalization, and weights are rebalanced monthly. The Galton is based on the incorporation of OOS errors into input estimation to overcome the incapability of historical data to reflect the tendency of mean regression that mean returns, variances, correlations and covariances have. I show that the Galton correction's superiority in risk prediction holds for China, even though the baseline method achieves an average negative Sharpe ratio. Nonetheless, when microcaps are excluded, the method produces significantly high Sharpe ratios, being this method the best among all the analyzed optimized strategies in a MV scenario. When the number of portfolio constituents is decreased, however, results for the regular form of the Galton seem to become positive; the changes in estimation windows, in their turn, have a favorable effect on the version where microcaps are excluded, making the Sharpe ratio rise above one.

Key words: Asset allocation, Galton, China, Shrinkage, Regression to the mean, Fama and MacBeth (1973) regressions, Microcaps

Abstrato

Na presente dissertação, aplico a metodologia de shrinkage de Barroso e Saxena (2022) – também conhecida como Galton – a fatores de otimização de portfólios no mercado chinês (para o período de OOS de janeiro de 2011 até dezembro de 2021). Cada portfólio é uma seleção das 50 maiores ações em termos de capitalização de mercado, sendo que os pesos são recalculados mensalmente. O Galton é baseado na incorporação dos erros de OOS na estimativa de variáveis para ultrapassar a incapacidade dos dados históricos de refletir a tendência de regressão à média que os retornos médios, as variâncias, as correlações e as covariâncias têm. Eu mostro que a superioridade da correção de Galton em previsão de risco se mantém para a China, apesar do facto de o método base obter um rácio Sharpe médio negativo. Não obstante, quando as pequenas ações são excluídas, o método atinge rácios Sharpe significativamente altos, sendo este o melhor método entre todas as estratégias otimizadas que foram analisadas num cenário de MV. Quando o número de constituintes dos portfólios diminui, contudo, os resultados para a forma regular do Galton parecem tornar-se positivos; as mudanças nas janelas de estimação, por sua vez, têm um efeito favorável na versão onde as pequenas ações são excluídas, fazendo os rácios Sharpe aumentar para mais que 1.

Palavras-chave: Alocação de ativos, Galton, China, Shrinkage, Regressão à média, Regressões de Fama and MacBeth (1973), Capitalização de mercado

Abbreviations

EB – Empirical Bayesian

EMC – Excluding Microcaps

E – Ex-post window

GMV – Global Minimum Variance

H – Historical window

IS – In Sample

L – Learning window

MV – Mean-variance

N – Number of portfolio constituents

OOS – Out of Sample

SSE – Shanghai Stock Exchange

SR – Sharpe ratio

Contents

Introduction

The present study aims to test the Galton method, presented in the Barroso and Saxena paper of 2022[1], in a different market: China.

The referenced article develops the Galton correction upon the realization that the optimization's metrics – mean returns, variances, correlations and covariances – have a strong regression to the mean, this is, lower historical values of one of these measures lead to higher ex-post values and the same happens for the opposite situation (higher historical values lead to lower ex-post ones). This tendency to the mean is a clear indicator that using historical values directly as estimators in a mean-variance scenario can lead to significant out-of-sample errors.

Therefore, the Galton correction is created as way of taking these, to an extent, predictable errors and incorporating them into the optimization's inputs estimation. In the original paper, the authors, in fact, find that there is a statistically significant pattern for the four mentioned metrics, which can be reflected in the coefficients of a Fama and MacBeth (1973) regression of future values on historical ones.[2]

The Chinese stock market is fairly new, especially when comparing with the US': the Shanghai Stock Exchange (SSE) has only been open since the 1990's[3], while the NYSE[4], USA's most prominent stock exchange, dates back to the late 1700's. In terms of dimensions, the SSE is much smaller than the NYSE, bearing a market capitalization of about one-fourth of the latter's one[5], despite the number of companies listed not being significantly different. This is linked to the fact that only a very small percentage of Chinese citizens actually invest in stocks, which is about 7%, compared to an astonishingly higher portion when analyzing the US, where 53% of people have money invested into the stock market.[6] This, in a sense, leads Chinese companies to be much more reliant on debt, rather than on equity, as it is practice in the US.[7]

[1] The paper applies the method only to the US market.
[2] For each month, the mean of the next 12 months is regressed on the previous 60 months' average. The Fama and MacBeth (1973) particularity of the regression is related to the fact that the values of the produced coefficients are averaged with the previous ones, on a rolling window that corresponds to the learning period of 108 months.
[3] It was open before, but closed in the late 1940's (https://www.investopedia.com/articles/investing/092415/chinas-stock-markets-vs-us-stock-markets.asp).
[4] New York Stock Exchange.
[5] As of October of 2022.
[6] https://www.investopedia.com/articles/investing/092415/chinas-stock-markets-vs-us-stock-markets.asp.
[7] https://www.investopedia.com/articles/investing/092415/chinas-stock-markets-vs-us-stock-markets.asp.

All these diverging factors may heavily affect the stock market's behavior, making the Chinese market an interesting one to contrast with the US, which was studied in Barroso and Saxena (2022).

In fact, I find that there is a weaker regression to the mean in the IS periods of China than in those of the US for mean returns and a substantially stronger one for variances; however, the estimates prove ineffective for the top 50 stocks OOS, producing negative excess returns. When one excludes the bottom 20% stocks in terms of market capitalization, there is an even weaker regression to the mean in China for mean returns, while, in the US, it is stronger; for correlations [and covariances], the opposite happens – in China, there is a stronger mean regression, while in the USA it is weaker. This sample [of top 80% stocks], nonetheless, produces significantly better estimates, resulting in portfolios' Sharpe ratios close to 1 (both in the baseline strategy and in the variations that are part of the robustness tests; some of the latter produce SR that are even above 1, on average). An empirical Bayesian version of the Galton is also tested, resulting in worse monthly performances than its regular counterpart (the Galton where microcaps are excluded), which is contrasting with the outcomes of Barroso and Saxena (2022).

Throughout the present dissertation, the Galton correction will be discussed in more detail. I will also apply the method to the SSE, as well as to some of the other naïve and optimized strategies presented in Barroso and Saxena (2022), to test which conclusions hold in the Chinese market.[8] Some robustness tests' results are also presented in order to further assess the shrinkage technique.

Literature Review

The present study uses the method proposed on Barroso and Saxena (2022) – known as the Galton – for input improvement in the portfolio weights optimization process. The Galton methodology, having at its core portfolio weights optimization with improved – shrank – inputs, is, thus, closely related to some other, in-practice, techniques.

James and Stein (1961) is one of the first papers to showcase the inadequacy of using historical values as estimates for the future. The authors prove that, by using a shrunk version of the previous mean returns as optimization's inputs for a sample with more than two stocks, one can

[8] Many authors have found that changing the sample in analysis can drastically change the results; this was the case of Goyal and Wahal (2015) and McLean and Pontiff (2016), for example.

significantly improve estimations. This shrinkage should be towards the grand, cross-sectional, average of the stocks' returns.

Elton and Gruber (1973) test the performance of inputs' estimation techniques that are based on the metrics' historical averages, comparing them to other, popular, approaches.[9] The input on which the authors focused were the correlations among stocks. They found that assuming the mean correlation for all pairs of stocks, or some variation of this, like assuming the same pattern of correlations for an industry, was consistently better than, for instance, using the historical pairwise correlations as estimates or computing correlations based on stock returns that are constructed as a function of prominent aggregate indices[10], which are some of the most common practices until today.

Jorion (1986) also refuses, based on Stein (1955), the idea that historical metrics should be directly used as estimates for future portfolio optimization. In this case, the author assumes that the grand mean of stock returns is unknown. To overcome this lack of information, one should, according to the paper, direct their attention to the predictive density function of future returns.[11] Once the latter is integrated on the grand, unknown, mean and on the sample's historical mean, the author finds that the portfolio's grand mean is the portfolios' expected return when optimizing for the global minimum variance (GMV). The expected returns used for weights' optimization will, then, be computed as a function of this grand mean and the sample means.[12] The covariance matrix is also corrected to account for risk in the sample mean returns and in the common factor.

Ledoit and Wolf (2004) uses a similar prior to Barroso and Saxena (2022) to achieve optimal portfolio weights. Even though the authors do not consider the mean returns as a possible subject for adjustment, contrarily to the Galton, they impose a correction to the covariance matrix, using this corrected matrix with the plain historical mean returns to optimize for the mean-variance portfolio. The covariance matrix correction is a product of three factors: the historical covariances (the structureless estimator), the average correlation (the target[13] when

[9] The authors also tested the historical model and single and multiple index models.
[10] Sharpe (1963): $R_i = A_i + B_{i1}I_1 + e_i$, where I denotes the index chosen as the basis.
[11] Indirectly dependent on the grand mean of stock returns.
[12] $E[r] = (1 - w)Y + w\iota Y_0$, where w is the shrinkage factor (dependent on the number of stocks and on prior precision), ι is a matrix of ones, Y is the vector of means and Y_0 is the grand mean.
[13] Ledoit and Wolf study other possible shrinkage target in their 2003 paper: the single-factor matrix, proposed by Sharpe in 1963.

shrinking the covariances) and the constant that will serve as a weight for the allocation between the covariance matrix and the target.[14]

More recently, papers like Nard (2020) have further explored the possibility of metrics' shrinkage. The author proposes the application of a correction to the covariance matrix based on clusters, with these being based on external information, like the industries the companies belong to, or purely on the stocks' data itself.

Alternative to applying shrinkage to the optimization's inputs, one could correct the portfolio weights directly. This is the matter of research in the Golosnoy and Okhrin's 2009 paper, which also employs the corrections based on clusters, being these independent from external factors.

The Galton: methodology and different versions

1. The Galton and its Bayesian parallel

Barroso and Saxena (2022), as previously explained, developed the Galton strategy after the realization that stock returns, correlations and covariances regress to their grand, ex-post, mean. This invalidates the commonly trusted idea that historical values are good estimators to use directly in weight optimization in a mean-variance setting. As such, the authors propose an estimator that combines the metrics' grand mean and the historical values, depending on the relationship between ex-post realizations and past historical averages: ex-post values are regressed on historical ones on a monthly basis for a certain number of months; the resulting coefficients of these regressions are used to correct historical values, computing the optimization inputs.

The above-mentioned regressions are Fama and MacBeth (1973) ones. This means that, for each month in the OOS period, the correspondent correction's coefficients are not given by a single monthly, cross-sectional, regression, but by a monthly update of many monthly regressions: a first regression is calculated, as shown by the following formula (for mean returns; the other three metrics follow a similar logic)

$$\mu_{i,t+1|t} = \hat{g}_{0,m,t} + \hat{g}_{1,m,t}\mu_{i,H,t}$$

where $\hat{g}_{0,m,t}$ is the estimated intercept at time t, $\hat{g}_{1,H,t}$ is the average cross-sectional relationship between ex-post and historical mean returns (in a way that sum squared errors are minimized)

[14] The method used in Ledoit and Wolf (2004) can be written as a formula: $\widehat{\Sigma_{Shrink}} = \widehat{\delta^*}F + \left(1 - \widehat{\delta^*}\right)S$, where $\widehat{\delta^*}$ – the constant of shrinkage – is the optimal weight to allocate to the shrinkage target (the average sample correlation), F is the target and S is the unstructured covariance matrix.

and $\mu_{i,H,t}$ is the historical mean return. The resulting coefficients are, then, updated in the next month (and, again, in the following months, repeatedly, until the end of the learning period):

$$\hat{g}_{1,m,t}^{fm} = \hat{g}_{1,m,t} + \hat{g}_{1,m,t-1}/t$$

The regular method analyzed in Barroso and Saxena (2022) assumes an historical period (H) of 60 months, an ex-post (E) period of 12 months and a learning (L) one of 108 months. Thus, once 108 cross-sectional regressions are done, [and, thus, the final regression is computed through an average of the 108 prior ones,] the inputs for the stocks' mean returns are ready for use in the first month of the OOS period.

The just-described method has a Bayesian prior, which is the belief that optimization inputs are similar for every stock. This means that ex-post values are similar to the grand mean of historical values, which is not what the Galton correction assumes, strictly speaking, as the method adjusts this value (the Fama and MacBeth intercept – $g_{0,m,t}$, which is already a percentage of the grand mean, when $g_{1,m,t} \neq 0$) by a percentage ($g_{1,m,t}$) of the historical average. Only when $g_{1,m,t} = 0$, does the expected value correspond to the grand mean.

Weights are then computed in a mean-variance scenario, using the Markowitz (1952) formula:

$$w_{G,t} = \frac{\Sigma_{G,t}^{-1}\mu_{G,t}}{\iota^T \Sigma_{G,t}^{-1}\mu_{G,t}},$$

where ι is a matrix of ones with size N-by-one, $\mu_{G,t}$ is the matrix of predicted mean returns and $\Sigma_{G,t}^{-1}$ is the inverse of the predicted covariance matrix, which is, in its turn, achieved by the formula

$$\Sigma_{G,t} = diag(\sigma_{G,t})\rho_{G,t}diag(\sigma_{G,t}),$$

where $diag(\sigma_{G,t})$ is the diagonalization of the matrix of volatilities (the diagonal corresponds to the Galton estimates of the volatilities, while the other elements of the matrix are zeros; this results in a N-by-N matrix) and $\rho_{G,t}$ is the N-by-N matrix of correlations. Barroso and Saxena (2022) applies the absolute value to the denominator of the weights' formula for the strategies analyzed OOS. This is to avoid a negative denominator switching the relative portfolio weights and is also applied in the present thesis in order to ensure compliance with the original study.

2. Exclusion of microcaps

The same methodology is applied to a sample where the bottom 20% of stocks in terms of market capitalization is excluded. This means that only the top 80% stocks' behavior during the IS periods affects the inputs' predictions, as only its data is used to estimate the coefficients.

In fact, the mean returns' slopes later presented (in Table 1) supports this idea: the exclusion of the bottom 20% of firms increase, in absolute terms, the slope by 0.14 units, i.e., historical mean returns are counted by 14% more for top stocks when applying the shrinkage. If the distribution between the grand mean and the historical one was 126%[15] to -26%, without the microcaps' data it is 140% and -40%.

Data and portfolio selection

1. Data, estimation windows and the OOS

Firm level data is retrieved from Refinitiv Eikon through its API service. Stocks are selected based on the Shanghai Stock Exchange, bearing data from January of 1996 until December of 2021. Risk free rates are based on the *China (Mainland) Interbank lending weighted average interest rate, 1 month* variable of Refinitiv Eikon.

Considering H=60 months, E=12 months and L=108 months, the first cross-sectional Fama and MacBeth (1973) regression used to compute the first Galton coefficients counts with 60 months of H[16] – from January of 1996 until December of 2000 –, then, the cross-sectional regression run in January of 2001 is updated monthly until December of 2009[17], forming the first coefficients estimates, to be used for the first OOS month.

Data until November of 2010 is used to compute the first coefficients used OOS.[18] The first OOS month, for which the first portfolio is formed, is January of 2011. This implies an OOS period of 132 months (from January of 2011 until December of 2021).

[15] The [historical] ex-post average is weighted by 126%, meaning that, if the intercept has the value of 1.56%, the grand mean of returns would have the value of 1.24%: 126%*1.24%=1.56%.
[16] There are 937 stocks eligible for the regression (with 60 months prior and 12 subsequent, at a given time t).
[17] This last cross-sectional regression uses average historical data from December of 2004 until December of 2008 and average ex-post data from December of 2009 until November of 2010 (all interval limits mentioned included).
[18] Portfolios are created at the beginning of each month. Thus, the next 12 months of data that form the ex-post window englobe that same month's data, plus the next 11 months of data. The cross-sectional regressions follow the same logic.

2. Portfolios

Portfolios are created monthly from January of 2011 until December of 2021 based on a MV perception.[19] Each portfolio has N=50 constituents, which correspond to the fifty biggest stocks in December of the previous calendar year in terms of market capitalization. This implies that stocks in the portfolios are the same from January until December of that same year, but their weights are rebalanced each month.

Mean Regression in China

When applying this methodology to China, maintaining the learning (L), the historical (H) and the subsequent (E) periods of L=108, H=60 and E=12 months, the results show different trends of regression to the mean.

Table 1 - Fama and MacBeth (1973) coefficients

Panel A - All stocks				
	Covariance	Correlation	Variance	Mean returns
Intercept	0.01	0.33	0.01	0.02
t-stat (=0)	61.06	79.30	8.43	7.18
Slope	0.37	0.28	0.26	-0.26
t-stat (=0)	42.85	32.59	2.96	-2.38
Greater than 0 (%)	99.17	100.00	70.00	9.17
t-stat (=1)	-140.13	-87.65	-12.04	-11.93
Smaller than 1 (%)	92.50	100.00	94.58	100.00
R-squared (%)	4.59	2.46	2.68	2.98
Min.	17578.00	17578.00	188.00	188.00
Average	117727.05	117727.05	474.04	474.04
Max.	271216.00	271216.00	737.00	737.00
Panel B - Excluding microcaps				
Intercept	0.01	0.40	0.02	0.02
t-stat (=0)	45.44	56.74	7.58	6.55
Slope	0.30	0.20	0.26	-0.40
t-stat (=0)	21.30	14.42	2.55	-2.57
Greater than 0 (%)	92.08	95.83	65.58	3.75
t-stat (=1)	-93.07	-60.59	-10.64	-10.04
Smaller than 1 (%)	93.75	100.00	94.17	100.00
R-squared (%)	3.14	1.40	2.67	3.74
Min.	11476.00	11476.00	152.00	152.00
Average	51621.43	51621.43	305.07	305.07
Max.	95266.00	95266.00	437.00	437.00

Fama and MacBeth (1973) regressions of ex-post values of historical ones. Coefficients were computed from January of 2011 until the same month of 2021 with data ranging from January of 1956 until November of 2021.

[19] This represents a short OOS period when comparing to Barroso and Saxena (2022), which might hinder the results (they are likely to be more random than those presented on the mentioned paper).

The first four rows of Panel A represent: 1) the average intercept of the regressions; 2) its average t-statistic; 3) the average slope; 4) its average t-statistic (both t-stats are Newey-West, 1987, with 12 lags). Rows five to seven are: 5) the proportion of t-statistics that corroborate with the hypothesis that the slope is greater than zero with a significance level of 5%; 6) the Newey-West (1987) t-statistic for the hypothesis that the slope is equal to one; 7) the proportion of t-statistics that, for a significance level of 5%, cannot exclude the hypothesis that the slope is smaller than one. Row eight is the average R-squared of the regressions and the last three rows represent the minimum, average and maximum number of stocks to be included in a monthly regression, respectively. Regressions are done pairwise for covariances and correlations (first two columns) and stock by stock for variances and excess mean returns. Panel B has the equivalent value for the top 80% of market capitalization stocks.

In respect to excess mean returns, the slope is negative, which indicates that, the higher the historical value of this measure, the lowest one can expect it to be when ex-post. This rapid reversion to the mean[20], which is contrasting with those of the other above represented metrics for the fact that the slope is negative, is statistically significant.[21] Firstly, and assuming a normal distribution, the hypothesis that the slope is zero, and, therefore, historical returns are irrelevant for predicting future ones, is rejected at the 5% significance level. Plus, only in 9.17% of the regressions, was it possible to reject the hypothesis that the slope coefficient is smaller or equal to zero. This, together with the rejection of slope equaling zero, further reaffirms the negativity of the relationship between ex-post and historical mean returns. The hypothesis of the slope being one is also clearly rejected at the 1% level, and 100% of regressions are significantly smaller than this threshold at the 5% level.

Regarding variances, the scenario is somewhat contrasting. Instead of the ex-post values decreasing with the historical ones, they increase, but without ever being equal to them. This is noticeable in the fact that in 94.58% of cross-section regressions, the slope coefficient is smaller than one for a significance level of 5%. Moreover, in close to 70.00% of the monthly regressions, this coefficient is statistically believed to be greater than zero for the same level of significance. Also, the hypothesis of the slope being equal to zero is rejected at the 1% level, being the relationship between ex-post and historical values stronger than in mean excess returns. This rejection is even stronger when testing for the slope being equal to one (believed not to be equal at the 1% level).

Concerning pairwise regressions, both regressions of covariances and correlations have greater than zero, but smaller than one, average slopes. These average coefficients are both statistically significantly different than one and zero at the 1% level, rejecting the hypothesis of them being lower than zero or higher than one in almost all regressions at least at the 5% level (for correlations, all Fama and MacBeth regressions are greater than zero and smaller than one,

[20] Only -26% of the historical return's value is counted, on average, to create the optimizations' inputs; this means that the grand, or ex-post, mean has a weight of 126% in the estimate.
[21] This is consistent with the long-term reversal effect found by De Bondt and Thaler (1985).

while for covariances these amount to 99.17% and 92.50%, respectively). Thus, similarly to the variances, the higher the historical values of each of these metrics, the higher will be the expected ex-post values, but never reaching the same level as the former, i.e., ex-post covariances and correlations are expected to be only about 37% and 28%, respectively, of their historical values (corrected by the bias[22]). This will stabilize values in the grand scheme, as higher values will, in proportion, increase less than lower ones[23], reflecting the tendency of mean reversion that these metrics are known to have and making the plug-in approach (using only the recent historical values) inadequate when it comes to MV optimization.

In terms of explicability, the average r-squared is close to, but smaller than, 3% for correlations, variances and mean returns, and higher than 4% for covariances, meaning that this is the metric where historical values better explain the variation in the ex-post ones. Nonetheless, the four measures present relatively low average r-squares.

The results are similar when excluding the 20% smallest stocks in terms of market capitalization: the nature of the relationship between ex-post and historical values is the same as in Panel A for all metrics. However, average slopes for covariances and correlations are smaller in this second approach, indicating that the regression to the grand mean will happen even quicker, as values are expected to be less dependent on the recent history. Significance wise, variance is the only of the first three to diminish, being previously significant at the 1%, and now it is only at the 5% level. This lost in significance is also seen in the percentage of stocks that have a slope that is statistically believed to be above zero at the 5% level, as this proportion decreased from 70.00% to 65.58% of all Fama and MacBeth regressions. Regarding mean returns, the average slope increased in its negativity, meaning that, the higher the historical mean returns, the furthest away from the grand mean it will be. This is opposite of what was verified with the former three metrics in terms of sign and explicability (r-squared), as it is now bigger by 0.76pp. The gain in significance in mean returns can also be verified in the lower percentage of regressions that are statistically believed to have positive slopes at the 5% level (only 3.75%, compared to 9.17% in Panel A).

[22] The regression's intercept.
[23] Imagining a month with Fama and MacBeth intercept covariance's intercept being equal to 0.01 and slope of 0.34, if the historical average of covariances is above the intercept (the average of the 48 months' mean) – 0.012, for example –, the expected value for the ex-post period will be 0.01408; if, instead, it is below the intercept – for, instance, 0.008 – the expected covariance will be 0.01272. The percentual difference between expected and historical values is bigger in the second case, being the former only of about 17%, significantly low when comparing to the latter one of 59%.

When comparing these results to those obtained by Barroso and Saxena (2022), one can easily understand that one of the biggest differences is in the variances' regressions: the relationship between ex-post and historical values is much greater both in terms of explicability and of the proportion of the dependency for the US (1952 – 2016) than for China (1996 – 2021). The average slope for the US variances was reported to be 0.55, which means that more than half of the historical value, on average, is used to correct the grand [ex-post] mean variance. As a result, ex-post variances (or the expected variances, for the sake of weight allocation) will be much closer to the historical ones, computed with the previous 60 months of data, than in China. This is also reflected in the r-squared, which is significantly bigger than for China (12.17% for the US comparing to 2.68% for China), indicating a much higher explicability of the variation in ex-post variances by the historical ones. Regarding the pairwise metrics, the slope and the r-squares are similar for both countries in the time periods in analysis. However, while, in the US, all Fama and MacBeth regressions are significantly above zero for a level of 5%, in China, a small portion of them is not positive enough to exclude the possibility of being equal or smaller than zero (0.83%, in covariances). When comparing mean returns' results in China versus the US, the greatest difference is in terms of how much historical values are expected to influence ex-post ones: in the present study, an average slope of -0.26 is reported, while in Barroso and Saxena (2022) finds a mean slope of -0.16. This suggests a stronger convergence to the grand ex-post mean in the US, as it places less importance on the historical excess mean returns. Besides the slopes, when focusing on the intercepts, it is noticeable that ex-post means are, on average, greater for China than for the US. This is true for all metrics except variance, but more significantly for correlations, which report an extremely large difference of 0.20: the average intercept for China's Fama and MacBeth's regressions is 0.33, while that of the US is about 0.13; the higher correlation between stocks in the ex-post periods in China might negatively affect the performance – via the increased risk – of portfolios created using solely Chinese stocks, if these somewhat highly related movements are also prevalent among the selected instruments. In terms of differences between Panel A and B, both in China and the US, one can see a trend of lower slopes when excluding microcaps in variances' regressions. However, in the pairwise metrics, while in China, as previously developed, slopes tend to be greater for Panel A, the US reflects the opposite tendency. Regarding mean returns, this matter is also different between the countries and periods, as, for China, when one excludes microcaps, there is an increase in the negativity of the slope, while for the US, on the contrary, there is a decrease.[24]

Out-of-sample performance: the Galton vs. other strategies

1. Comparable strategies

For a thorough comparison of the Galton methodology applied to China versus to the US, some of the other portfolio MV allocation strategies that were also used in Barroso and Saxena (2022) are also represented in the present dissertation. These are either simple, standard, strategies, which is the case for the value weighted, the Talmud's 1/N and the Markowitz MV, or corrected strategies, namely the James-Stein version of the Markowitz, of the Elton and Gruber's and of the Ledoit and Wolf's.

The standard strategies are those that do not implement any correction to the optimization inputs. The three above mentioned methods are some of the most popular and used ways to allocate portfolio weights, either for their simplicity or for having been proven to have good, consistent, results.[25] However, for the Markowitz MV ("plug-in") approach, despite its simplicity and common use, it is known to frequently generate unproportionally high weights, which can be a major risk for an investor.

To control for some of these risks, the correction proposed by James and Stein in their 1961 paper was applied to excess mean returns: it consists in balancing the cross-sectional grand mean and the historical average; the latter one is weighted by the maximum between zero and the following expression:

$$1 - \frac{\frac{(N-3)\sigma_0^2}{T}}{\|\mu_H - \mu_0 \iota\|^2}, N \geq 4,$$

where μ_H is the historical, individual, excess mean return, μ_0 is the grand, cross-sectional, mean, σ_0^2 is the grand variance across a time-series of T periods, ι is the matrix of ones with shape N × 1 and *N* is the number of stocks. This is the case for the Markowitz, but also for the two other, optimized, strategies – the Ledoit and Wolf and the Elton and Gruber –, which already perform shrinkage by definition: Ledoit and Wolf applies the correction to the

[24] Barroso and Saxena (2022) reports an average slope of -0.16 using all stocks and -0.10 when excluding microcaps. This reflects a decreased relevance of historical excess mean returns when only considering the 80% biggest stocks, meaning that this selection (Panel B) has a stronger reversion to the mean. The opposite is proven for China (in the period in analysis) in this study.
[25] DeMiguel, Garlappi and Uppal (2009) proves that the Talmud's 1/N is not consistently beaten by any other allocation method in analysis.

covariance matrix, having as a target the average, cross-sectional, correlation[26]; the Elton and Gruber approach corrects only the correlation matrix, using historical volatilities as inputs for the expected ones. The latter method is also reliant on constant correlations, as, for each period, pairwise correlations are expected to be equal to the cross-sectional average of the historical ones.

2. Portfolios' performance

Table 2 shows the performances of the monthly portfolios formed on the 50 highest market cap stocks for the above-mentioned strategies. The metrics reported are the average Sharpe ratio of every month, the active share, computed with the value weighted portfolio weights as a basis, the turnover, the bankruptcy rate, which is the percentage of months whose returns were below -100%, and the characterization of the average portfolios' weights. These are the average minimum and maximum weight of the month, the average of the monthly volatility of weights and its standard deviation, and the average sum of negative weights by month. All these metrics are relevant when taking decisions about which strategy to implement and all were used in the Barroso and Saxena (2022) paper, allowing for a closer comparison.

Table 2 - OOS performance

	Sharpe	Active Share	Turnover	Bankruptcy rate	Min *w*	Max *w*	Mean StDev *w*	StDev StDev *w*	Sum Negative *w*
Value weighted	0.39	0.00	6.89	0.00	0.32	13.89	2.80	0.51	0.00
Talmud 1/N	0.17	44.92	2.30	0.00	2.00	2.00	0.00	0.00	0.00
Markowitz	-1.88	7052.03	18965.94	8.33	-1146.31	955.48	378.92	1456.41	-7047.65
Galton	-0.02	362.34	161.48	0.76	-44.46	41.94	18.35	7.19	-320.40
Galton (Exc. MicCaps)	0.93	536.84	327.41	0.76	-55.05	62.51	26.32	18.13	-486.26
Galton EB	0.22	75.51	25.69	0.00	-7.09	10.65	3.89	0.54	-36.83
Markowitz (JS)	-0.76	2945.52	8558.36	2.27	-400.33	425.95	156.47	679.40	-2917.05
Elton Gruber (JS)	0.39	252.62	470.06	0.76	-30.04	43.39	14.13	27.29	-229.51
Ledoit Wolf (JS)	-0.28	388.88	941.72	2.27	-51.04	59.30	21.50	55.61	-362.61

OOS performance. The first portfolio is formed in January of 2011 and the last in December of 2021. Historical periods range from January of 2006 until November of 2021. The first four columns are the Sharpe ratio (average of the OOS months), the average active share (in percentage and with the value weighted strategy as a benchmark), the average turnover (in percentage) and the bankruptcy rate (in percentage; portion of months with realized exc ess portfolio's return below -100%). The last five columns are the average minimum and maximum weights (maximum and minimum is observed for each of the 132 months separately and then averaged), the mean and the standard deviation of the monthly weights' volatilities, and the average of the sum of negative weights (sum by month and then average the 132 months). All these five columns' values are in percentage.

[26] The constant correlation model, proposed in their 2004 paper *Honey, I shrunk the sample covariance matrix*.

Regarding mean-variance performance, the Sharpe ratios indicate that the best strategy would be the Galton, but excluding microcaps.[27] This method achieves an average ratio close to one, which indicates that the realized excess returns are close to their volatility, meaning that risk is almost compensated by the actual portfolio's returns. There is a high discrepancy when comparing to the other two Galton approaches: Galton EB is significantly inferior to its non-constrained counterpart; the regular Galton strategy achieves negative results. The former approach – Galton EB – is also applied to the sample where microcap stocks are excluded; this did not guarantee a superior performance, as verified in the second Galton strategy, meaning that applying the restrictions to the Galton coefficients leads to decreased results. Both Markowitz's strategies achieved negative Sharpe ratios, with the second one (the one with the James-Stein correction on mean returns) having a less negative ratio, as it would be expected; the Elton Gruber method was the only one out of the last three optimized strategies to obtain an average positive Sharpe ratio; the values weighted and the Talmud 1/N both achieve average ratios above zero, even though the latter method presents a very modest value of 0.17.

The active share is a measure of how much its weights diverge from the benchmark, more standard, portfolio. In this case, just like in Barroso and Saxena (2022), the basis of analysis are the value weighted portfolios. Thus, a higher active share indicates that the strategy at hand deviates more from the strategy where weights are allocated to stocks as a proportion to their market capitalizations; the more it deviates, the more it might be justifiable the implementation of the more optimized strategy (which might be more costly), depending on whether it generates better results or not. On the other hand, if weights are very disparate from the benchmark, which, in principle, is characterized by moderate weighs, it might be indicative of a higher risk being taken by the investor when applying the alternative strategy.

For the regular Galton strategy, the average monthly active share is about 362.34%, meaning that, for the average month, weights greatly deviate from those of the value weighted portfolio. This is also reflected on the average maximum and minimum weights, which are more extreme than those of the benchmark strategy. The latter is especially true for the minimum weight, which is negative for the Galton and positive for the value weighted; this fact is also reflected in the average sum of negative weights, which is null for the benchmark and about -320.40% for the Galon. The second Galton strategy has an even larger active share, being also noticeable

[27] This could be expected as, as seen in Table 1, smaller stocks have different patterns of regression to the mean than larger ones. Thus, the sample where microcaps are excluded produces more accurate estimates for the top 50 stocks.

the even more extreme maximum and minimums and sum of negative weights. The average monthly volatility of weights is also affected by the wider interval, as well as its standard deviation, which is unproportionally larger than that of the regular Galton. The Galton EB, once again, despite being performed on the same sample of the second Galton strategy, displays contrasting results: from the three Galtons [and out of all the optimized strategies], this is the one with the lowest active share (significantly lower), resulting in much more moderate [or controlled] weights metrics (Galton EB's average weights' maximums and minimums are, in absolute terms, greatly lower than those of the former two strategies, with the average maximum being even smaller than that of the value weighted method; its mean and standard deviation of weights' volatilities very low, being closer to those of the benchmark method; the average sum of negative weights is smaller in absolute terms, making it a more attractive strategy to investors with negative weights' limitations).

Both Markowitz's strategies require an exponentially larger mean active share, which meets expectations, as this method is known to produce extreme weights. Thus, maximum and minimums are also very large (smaller, in absolute terms, for the method were the James-Stein was applied), and so are their average volatility and its standard deviations (these two strategies present the highest *Mean StDev w*, which is a direct reflect of the extremely high active shares; this metric has the mere value of 2.80% for the value weighted portfolios, compared to 378.92% for the plug-in Markowitz and 156.47% for the James-Stein adjusted one; this indicates a bigger management of the portfolios' weights) and the sum of negative weights, being these the most extreme out of all the represented methods. The Elton-Gruber (JS) and the Ledoit-Wolf (JS) both produce active shares close to those of the first two, unconstrained, Galton strategies.

The average turnover is a measure of how much, within the same portfolio and strategy, an investor has to manage the allocation of stocks' weights. The values are, thus, very small for the strategies whose weights depend on moderately stable characteristics, like the market capitalization of a firm (the case of the value weighted strategy, which present a turnover of 6.89%), or even on absolute prior (the Talmud's 1/N keeps weights stable at, in this case, 2.30%; these is only a turnover of stock weights when the portfolio's composition is renewed at the beginning of each year). Much like the active share metric, the portfolio turnover of the regular Galton is lower than that of the second alike strategy, being, nonetheless, moderately high, indicating non-stable portfolios that were already hinted by the high standard deviations of weight volatilities, together with their moderately high means. The Galton EB has, as well as a lower active share and absolute maximum and minimum weights, a lower turnover – the turnover is 92.15% lower than that of the second Galton's strategy; this is actually more than proportional to the decrease in the weights' mean volatility (-85.22%) from the second Galton to the EB one, and might be indirectly justified by the significant decrease in the standard deviation of this volatility (-97.02%), which means that weights have a similar deviation from each other every month, sticking close to the average volatility of 3.89%; a lower monthly standard deviation of volatilities is likely due, in this case, to a lesser weight variation across the months. Markowitz, once again due to its extreme weights, generate very high turnovers in both strategies (a small variation in the inputs of a certain stock within the same year, for example, might produce significantly different weights for that same instrument in the portfolio). Elton-Gruber (JS) and Ledoit-Wolf (JS) produce higher turnovers than any of the Galton methods.

In terms of bankruptcy rates, from the three Galtons, both the regular and the first where microcaps are excluded verified one month with realized excess returns lower than -100%. This is a consequence to bear in mind that might hinder some of the benefits of the high Sharpe ratio obtained by these same strategies. The Elton-Gruber (JS) strategy obtained the same result, which also works as a possible downturn of the relatively high average Sharpe ratio. The Ledoit-Wolf (JS), just like the Markowitz (JS), had three monthly returns lower than -100%, and the plain, plug-in, Markowitz strategy produced bankruptcy in 11 out of the 132 months in the OOS.

Comparing these results to those obtained in Barroso and Saxena (2022) for the USA between 1967 and 2017 (OOS; including both limit years), one can notice some major differences:

- The first has to do with the fact that, in the US, none of the tried strategies beat the regular Galton in terms of average Sharpe ratio. This is not the case for China's sample, as only the Galtons strategies where microcaps are excluded achieve the highest ratios. Moreover, the regular Galton achieves, on average, positive SR in the US, while in China, for the selected period, the ratio is negative.
- In the US sample, Galton EB performs better in terms of Sharpe ratio than the second Galton strategy, meaning that imposing the coefficients' restrictions ended up being beneficial for the portfolios' overall performances. For China, the Sharpe ratio decreases with the imposition of these limits. This suggests that the long-term reversal of China positively contributed to the unconstrained Galton, more than it did in the US.

- In China, the weights are more extreme and less stable. This is noticeable by the higher minimum and maximum weights on absolute terms on average and by the bigger mean weight volatility and its standard deviation in China. Absolute sums of negative weights are also higher in China for each of the represented methods. Active shares and turnovers are also, on average, bigger than their US counterparts. Bankruptcy rates are smaller in the US for every strategy.
- The Talmud 1/N portfolios achieve a lower mean Sharpe ratio in the present sample.
- Ledoit-Wolf (JS), which, from the shrinkage strategies here represented, was the only one to achieve a positive Sharpe ratio of 0.11 in the US sample, in China, it generated a negative ratio of -0.28. The opposite happened for Elton-Gruber (JS), which, in China, led to a positive average Sharpe ratio of 0.39, while, in the US, it had the value of -0.16.

OOS returns and risk analysis

Table 3 reports the expected and realized returns of the strategies in analysis. Furthermore, there are represented, for each method, the hit rates for the first and fifth – bottom and top – quantiles of the excess returns distribution. This latter analysis allows for an assessment of the value-at-risk (VaR), which might be determining for an investor when deciding on the best portfolio strategy for their needs.

Table 3 - OOS returns, volatilities and hit rates

	Expected r	Realized r	RMSFE	Expected σ	Realized σ	r<Qz(1%)	r<Qz(5%)	r>Qz(95%)	r>Qz(99%)
Value weighted		7.22			18.49				
Talmud 1/N		3.50			20.35				
Markowitz	672.87	-714.64	1387.51	106.45	380.10	50.76	53.03	28.79	25.76
p-value		(.02)				(.00)	(.00)	(.00)	(.00)
Galton	43.82	-0.82	44.64	52.29	49.37	2.27	8.33	3.03	0.00
p-value		(.02)				(.33)	(.17)	(.19)	-
Galton (Exc. MicCaps)	68.44	60.78	7.66	71.95	65.24	0.00	4.55	6.06	0.00
p-value		(.74)				-	(.80)	(.61)	-
Galton EB	19.71	4.43	15.28	33.06	20.48	0.76	1.52	0.76	0.00
p-value		(.04)				(.75)	(.00)	(.00)	-
Markowitz (JS)	624.81	-120.82	745.63	29.14	158.56	52.27	56.06	28.03	24.24
p-value		(.06)				(.00)	(.00)	(.00)	(.00)
Elton Gruber (JS)	127.25	15.00	112.25	24.66	38.94	21.97	31.06	9.85	3.79
p-value		(.00)				(.00)	(.00)	(.06)	(.09)
Ledoit Wolf (JS)	168.92	-11.64	180.56	31.54	41.70	23.48	29.55	9.85	6.06
p-value		(.00)				(.00)	(.00)	(.06)	(.01)

OOS results: expected versus realized returns and risk metrics. First portfolio created in January of 2011 and the last in December of 2021. First two columns represent the annualized, average, expected excess returns and the realized ones of the respective portfolio. P-values for the second column (in parenthesis) are referent to the difference between realized and expected returns being equal to zero. The third column is the root mean square forecast error (RMSFE) for the excess returns. The fourth and fifth columns are, respectively, the annualized expected volatility of returns and the realized one. The last four columns are the hit rates for the first and fifth bottom and top quantiles of the excess returns distribution. P-values for these columns are in parenthesis and are generated for the hypothesis of the hit rate being equal to the corresponding quantile. All values besides p-values are in percentage.

The regular Galton strategy presents a significantly large RMSFE of 44.64%, being this the greatest one out of all three Galton methodologies. This is due to a big discrepancy between expected and realized excess returns (being the expected one 43.82% higher than the actual one, which is negative), which are statistically believed to be different at the 5% significance level. This, together with an expected volatility of 52.29%, would lead one to expect a Sharpe ratio of about 0.84, which is very discrepant from the ex-post period (as previously presented, the average realized ratio is of about -0.02 for this strategy; this represents a difference of -102.38%). This mismatch between realized and expected Sharpe ratios can be explained by the significant percentual decrease in returns that was not compensated by the change in volatility: although realized volatility is smaller than the expected one, this difference only amounts to 5.58%[28], which is smaller than the 102.38% decrease in excess returns previously mentioned.

In terms of hit rates, the regular Galton strategy does a moderately decent job in predicting the distribution of returns: for the lowest percentile (r<Qz(1%)), where 1% of realized returns should be if one wanted to precisely match the expected distribution, [but, ideally, the lowest possible percentage would be concentrated,] a percentage of 2.27 of the realized returns is found. This can represent a problem when it comes to VaR, as an investor is expecting only 1% of monthly returns to be below the corresponding 1% value, but, in reality, happens to obtain 2.27% of their monthly returns below that threshold. Similarly, below the lowest 5% percentile expected return value, there are concentrated 8.33% of monthly returns of the realized distribution. The opposite does not happen, as above the highest quantiles' values are only concentrated a small percentage of the monthly expected returns, fact that, together with the surpassing of the lowest quantiles, was already hinted by the significant RMSFE (with expected being higher than realized returns; realized returns distribution is to the left of the expected one).

The Galton strategy that excludes microcaps, on the contrary, obtains a very low average RMSFE. The average realized volatility is, similarly to the first Galton strategy, smaller than

[28] $\frac{\sigma}{E[\sigma]} - 1 = \frac{49.37}{52.29} - 1 = -5.58\%.$

the expected one, representing a conservative bias. Besides the latter fact, the average realized volatility of the current method is bigger than that of the regular Galton by 15.87pp. In terms of VaR, the realized returns distribution does not register any month below the lowest quantile's threshold of the expected returns distribution. Moreover, only six of the months[29] in the realized returns distribution are below the threshold of the bottom 5% quintile for the expected values. Opposingly to the regular Galton strategy, the highest quintiles register bigger hit rates than the bottom ones (although the top 1% has a hit rate of zero).

The Empirical Bayesian Galton is the one with the lowest realized volatility from the three Galton strategies and among all the shrinkage strategies. However, the percentual difference between realized and expected volatilities is the biggest, in absolute terms, out of all three Galtons (-38.05%); the standard difference is also the largest among the three strategies (Galton EB is the Galton strategy that most poorly predicts risk): 20.48%-33.06%=-12.58pp. This strategy has a higher hit rate for the bottom quintile than the Galton excluding microcaps (not EB), but lower values for the bottom 5% and for the top quintiles. Regarding the RMSFE, it is the second highest among the Galtons. However, it is still lower than any of the other non-naïve strategies.

Out of the shrinkage strategies, the Galtons are the only to overestimate volatility. This is a safer, more conservative, outcome for a risk averse investor than the opposite scenario, where risk would have been higher than what had been predicted. Furthermore, the absolute differences for these three strategies between realized and expected volatilities are all lower than those of the two Markowitz and of the Elton Gruber (JS), and the first two Galtons are better at predicting risk than the Ledoit Wolf (JS). Also, the three Galtons achieve the lowest RMSFE out of the corrected strategies, which is a positive point considering that realized returns are lower than the expected ones. Still comparing to the other strategies, all three Galtons have closer hit rates to the corresponding quantile, which, especially when it comes to the bottom extreme, is a great indicator of risk prediction.

Comparing these results to those obtain for the US in Barroso and Saxena (2022), one can observe the following patterns:

- The RMSFE are smaller for the Galtons than for the other optimized strategies and the plain Markowitz. Regardless of this fact, all RMSFE are larger in China than for the

[29] 6months/132months=4.55% in the *r<Qz(5%)* variable for the *Galton (Exc. MicCaps)* strategy. The denominator comes from the 132 months in the OOS period (January of 2011 until December of 2021).

corresponding strategy in the US. The regular Galton has the biggest discrepancy in RMSFE's between the US and China, being the realized return in the former positive (9.95%, on average) and negative for China.

- Both in China and in the US, the Galton strategies are the only ones to overestimate volatilities, which is an indicator of a more conservative approach. Nevertheless, China registers higher volatilities and bigger gaps between expected and realized volatilities, even for the Galtons. Plus, the difference in risk prediction between the Galtons and the other optimized strategies represented in the present study is smaller for China than for the US. In China, the absolute average difference between realized and expected volatilities is 51.29pp for the other shrinkage strategies; this average for the Galtons is of about 7.40pp. This amounts to a percentual advantage of 85.57% (the Galtons are, on average, 85.57% better at predicting risk than the other three optimized strategies). The latter advantage is of approximately 98.53% for the US.[30]
- The Ledoit-Wolf (JS), in the current study, presents a negative average realized return, while in Barroso and Saxena (2022), it achieves positive returns. The opposite happens for the Elton-Gruber (JS), as, in China, the realized returns are, on average, positive, while in the US they are negative.
- The Galtons seem better at predicting VaR for the US data (according to the results reported in Barroso and Saxena (2022)), as hit rates in the bottom tails are closer to the corresponding percentile value. Nonetheless, both Galton strategies that exclude microcaps have lower hit rates in the bottom quantiles than in the US and even than the percentile value, which reflects a lower skewness to the left on the realized distribution of returns than on that of the expected ones.

Variations of the Galton's method

1. Galton applied to covariances

The Galton's correction is, as previously explained, applied to the volatilities and to the correlations in order to ensure that the resulting covariance matrices are positive semidefinite. Nonetheless, it is possible to apply this correction directly to the covariance matrices.[31] Table 4 displays the results of this alternative methodology.

[30] The average absolute difference between expected and realized volatilities in the US for the Markowitz (JS), the Elton-Gruber (JS) and the Ledoit-Wolf (JS) is 84.03pp. This average for the three Galtons is 1.24pp.
[31] Using the coefficients regarding the *Covariances* column on Table 1.

Table 4 - Galton applied to covariances

Panel A - OOS performance

	Sharpe	Active Share	Turnover	Bankruptcy rate	Min *w*	Max *w*	Mean StDev *w*	StDev StDev *w*	Sum Negative *w*
Galton (covariances)	0.13	335.02	156.93	0.00	-42.98	37.50	17.13	6.71	-295.34

Panel B - OOS returns, volatilities and hit rates

	Expected r	Realized r	RMSFE	Expected σ	Realized σ	r<Qz(1%)	r<Qz(5%)	r>Qz(95%)	r>Qz(99%)
Galton (covariances)	43.13	5.83	37.30	51.18	46.31	1.52	6.06	5.30	1.52
p-value		(.03)				(.63)	(.61)	(.88)	(.63)

OOS results for the Galton strategy applied to covariances matrices, using the complete sample of stocks. OOS period ranges from January 2011 until December 2021. Panel A displays average weight and performance statistics: the first four columns are the Sharpe ratio (average of the OOS months), the average active share (in percentage and with the value weighted strategy as a benchmark), the average turnover (in percentage) and the bankruptcy rate (in percentage; portion of months with realized excess portfolio's return below -100%). The last five columns are the average minimum and maximum weights (maximum and minimum is observed for each of the 132 months separately and then averaged), the mean and the standard deviation of the monthly weights' volatilities, and the average of the sum of negative weights (sum by month and then average the 132 months). All these five columns' values are in percentage. Panel B displays average expected and realized statistics for the monthly portfolios OOS: the first two columns represent the annualized, average, expected excess returns and the realized ones of the respective portfolio. P-values for the second column (in parenthesis) are referent to the difference between realized and expected returns being equal to zero. The third column is the root mean square forecast error (RMSFE) for the excess returns. The fourth and fifth columns are, respectively, the annualized expected volatility of returns and the realized one. The last four columns are the hit rates for the first and fifth bottom and top quantiles of the excess returns distribution. P-values for these columns are in parenthesis and are generated for the hypothesis of the hit rate being equal to the corresponding quantile. All values besides p-values are in percentage.

When computed with the corrected covariance matrix, the Galton achieves a very distinctive performance than its regular counterpart. This is mainly due to the average realized excess return, which is positive with *Galton (covariances)*, opposingly to the regular Galton. The difference is also noticeable in the hit rates, which are smaller for the bottom quintiles (leading to smaller actual losses than the predicted 1% and 5% VaR) and bigger for the top ones with the present strategy, as well as in the fact that this method produces a bankruptcy rate of zero (the regular Galton's one is 0.76%) and that it has a smaller average sum of negative weights, on absolute terms. The better allocation that produced the increase in performance is likely to be due to the higher explicability of ex-post values by historical ones in covariances when compared to variances and correlations (bigger average r-squared for the former metric than for any of the latter). Regarding risk prediction, this method still overestimates volatilities, but the difference between realized and estimated values is larger.

2. Global Minimum Variance Galton

Although the Galton's correction was thought mainly to be applied in a mean-variance setting, principally because the method goes to the extent of applying shrinkage to mean returns, as well as to correlations and variances, it has other uses. One of them, which was also covered in Barroso and Saxena (2022), is the creation of global minimum variance (GMV) portfolios. In this application, only the Galton covariances matrices are needed to allocate the portfolio weights, meaning that only the Galton coefficients related to variances and correlations are used. The optimal weights are given by the following formula:

$$w_{G,t} = \frac{\Sigma_{G,t}^{-1}\iota}{\iota\Sigma_{G,t}^{-1}\iota}$$

where $\Sigma_{G,t}^{-1}$ is the inverse covariance matrix[32] and ι is an N-by-one matrix of ones.

Table 5 shows the results when optimizing for volatilities.

Table 5 - GMV Galton

Panel A - OOS performance

	Sharpe	Active Share	Turnover	Bankruptcy rate	Min *w*	Max *w*	Mean StDev *w*	StDev StDev *w*	Sum Negative *w*
Galton GMV	0.22	93.22	23.19	0.00	-7.11	12.31	4.27	0.81	-46.25
Markowitz GMV	0.39	475.58	495.31	0.00	-65.04	79.59	26.07	7.32	-429.04
Elton Gruber GMV	0.42	111.60	23.00	0.00	-4.69	27.52	7.07	0.62	-70.77
Ledoit Wolf GMV	0.60	124.45	36.00	0.00	-6.70	31.49	7.75	0.91	-75.83

Panel B - OOS returns, volatilities and hit rates

	Realized r	Expected σ	Realized σ
Galton GMV	2.93	29.54	13.40
Markowitz GMV	8.60	5.54	22.11
Elton Gruber GMV	7.15	11.87	17.01
Ledoit Wolf GMV	7.67	10.20	12.78

Galton corrections applied in a GMV optimization setting, using the complete sample of stocks. OOS period ranges from January of 2011 until December of 2021. Panel A displays average weight and performance statistics: the first four columns are the Sharpe ratio (average of the OOS months), the average active share (in percentage and with the value weighted strategy as a benchmark), the average turnover (in percentage) and the bankruptcy rate (in percentage; portion of months with realized excess portfolio's return below -100%). The last five columns are the average minimum and maximum weights (maximum and minimum is observed for each of the 132 months separately and then averaged), the mean and the standard deviation of the monthly weights' volatilities, and the average of the sum of negative weights (sum by month and then average the 132 months). All these five columns' values are in percentage. Panel B displays average expected and realized statistics for the monthly portfolios OOS: the first column represents the annualized, average, realized excess returns; the second and third columns are, respectively, the annualized expected volatility of returns and the realized one. All values are in percentage.

[32] Achieved at by the formula $\Sigma_{G,t} = diag(\sigma_{G,t})\rho_{G,t}diag(\sigma_{G,t})$, where $\sigma_{G,t}$ is the volatility matrix and $\rho_{G,t}$ the correlations' one.

For the GMV Galton strategy, the Sharpe ratio is positive, contrarily to what happens with the MV one. A smaller SR in the GMV version would be expected given the fact that the regular Galton is strictly optimizing for mean-variance, which ends up being reflected in the metric, but this increment might be explained by something else – the in-sample periods predict a higher regression to the grand mean than it actually happens in the OOS for the fifty selected stocks. This is hinted by the significant absolute slope increase when microcaps are excluded. As the Galton GMV's weight allocation does not rely on this [returns] prediction, it achieves a higher SR, even though the MV method is the one optimizing for it.

Average minimum and maximum weights are substantially less extreme for the GMV Galton than they are for the MV one. Plus, the mean weight volatility is less than half of that of the MV Galton, which is also linked to a largely lower active share, turnover and average sum of negative weights.

Regarding risk, this strategy achieves a lower realized average volatility by about 35.97pp. This is a meaningful decrease in risk, being even higher than what could had been predicted before the ex-post period (the difference between expected volatilities for the MV and the GMV Galton's is 22.75pp). In spite of this advantage, the risk prediction is inferior for the present method, being the difference between expected and realized volatilities of about 16.14pp, compared to one of 2.92pp with the MV Galton. Nevertheless, this method, just like the other variants of the Galton, overestimates risk.

The other three strategies above represented also register [substantially] lower realized volatilities than their MV counterparts, which would be expected, and all achieved better SR that in the previous MV scenario (with the Markowitz and the Elston-Gruber's performances even becoming positive, like the regular Galton). This denotes a problem of predictability for Chinese stocks in the period in analysis, as both the regular Galton and these three strategies – previously computed using the James-Stein estimation method for returns – achieve better overall results when not accounting for mean returns' estimations.

Robustness tests – variation of hyperparameters

In the present section, some of the Galton's hyperparameters will be changed in order to understand how the method would perform. The hyperparameters in question are the windows' length (historical, learning and ex-post windows) and the number of constituents of the portfolios. Table 6 displays the strategies and their OOS results.

The first strategy is a variation of the Galton where only the number of firms is changed: instead of the usual 50 stocks, the 40 stocks with highest market capitalization in the previous December are picked to be part of the monthly portfolios for the following year. This decrease in N was also tested in Barroso and Saxena (2022), although not with 40 stocks (instead, it was either 30). In the second one, the ex-post window in enlarged by one more year – E=24 months –, making the OOS period shorter: as L=108 months, and the first month of data available is January of 1996, the last cross-sectional regression to be part of the first Fama and MacBeth (1973) one is in December of 2009; this requires the use of data until November of 2011; the first OOS month is January of 2012 (the OOS period is of 120 months). An increase in E was tested in Barroso and Saxena (2022). However, E was made to be 60 months, and L was decreased also to 60 months. The third strategy was tested in the mentioned paper and is a simple increase in the learning period from 108 to 180 months.[33] An increase in the historical window, as done in the Barroso and Saxena (2022), would have been an interesting variation to test. However, the stocks did not have enough monthly data to extend H in even one year.

It is relevant to note that the different OOS lengths, being the latter one a substantially small period of five year (monthly data from January of 2017 until December of 2021), might significantly hinder the comparison – the lengthier ones produce stabler results – among variations of hyperparameters.

Table 6 - Variations in the hyperparameters

Panel A - OOS performance

		Sharpe	Active Share	Turnover	Bankruptcy rate	Min *w*	Max *w*	Mean StDev *w*	StDev StDev *w*	Sum Negative *w*
H=60, E=12, L=108, N=40	Value weighted	0.36	0.00	6.76	0.00	0.44	14.58	3.15	0.61	0.00
	Galton	0.11	305.37	137.18	0.00	-43.30	44.22	19.24	8.41	-264.75
	Galton EMC	0.64	498.03	342.04	1.52	-60.80	64.10	30.27	30.30	-446.93
	Galton EB	0.14	61.88	21.34	0.00	-6.57	11.03	4.01	0.58	-25.73
H=60, E=24, L=108, N=50	Value weighted	0.52	0.00	6.22	0.00	0.30	13.18	2.77	0.52	0.00
	Galton	-0.03	309.38	133.99	0.00	-38.85	36.64	15.79	5.02	-267.68
	Galton EMC	1.32	389.13	215.27	0.00	-37.83	42.46	18.83	5.41	-338.84
	Galton EB	0.46	71.09	24.38	0.00	-7.18	10.21	3.71	0.45	-34.17
H=60, E=12, L=180, N=50	Value weighted	0.66	0.00	6.43	0.00	0.32	10.84	2.41	0.12	0.00
	Galton	-0.57	301.17	125.56	0.00	-37.40	37.24	15.25	2.12	-253.46
	Galton EMC	1.75	368.33	211.91	0.00	-37.20	41.94	17.84	4.89	-318.04
	Galton EB	0.28	73.07	22.25	0.00	-7.99	10.10	3.86	0.36	-36.41

[33] The OOS period is, therefore, reduced, as the last year used for the Fama and MacBeth (1973) regression for the first OOS month (excluding the ex-post window) is 2015, making the E 2016.

Panel B - OOS returns, volatilities and hit rates

		Expected r	Realized r	RMSFE	Expected σ	Realized σ	r<Qz(1%)	r<Qz(5%)	r>Qz(95%)	r>Qz(99%)
H=60, E=12, L=108, N=40	Value weighted		6.67			18.63				
	Galton	38.52	4.96	33.56	50.63	44.19	1.52	5.30	1.52	0.00
	p-value		(.04)				(.63)	(.88)	(.00)	-
	Galton EMC	64.37	43.09	21.28	75.28	67.76	0.76	1.52	5.30	0.00
	p-value		(.52)				(.75)	(.00)	(.88)	-
	Galton EB	19.71	2.99	16.72	33.34	20.78	0.76	2.27	0.76	0.00
	p-value		(.02)				(.75)	(.04)	(.00)	-
H=60, E=24, L=108, N=50	Value weighted		9.87			18.85				
	Galton	36.16	-1.29	37.45	48.16	43.49	2.50	6.67	3.33	0.00
	p-value		(.04)				(.29)	(.46)	(.31)	-
	Galton EMC	44.23	61.66	17.43	55.31	46.78	0.00	2.50	7.50	0.00
	p-value		(.28)				-	(.08)	(.30)	-
	Galton EB	19.02	9.11	9.91	33.99	19.80	0.83	1.67	0.83	0.00
	p-value		(.18)				(.84)	(.00)	(.00)	-
H=60, E=12, L=180, N=50	Value weighted		9.57			14.55				
	Galton	35.06	-21.66	56.72	45.50	37.72	1.67	8.33	1.67	0.00
	p-value		(.00)				(.69)	(.35)	(.04)	-
	Galton EMC	43.97	68.34	24.37	52.25	39.02	0.00	3.33	3.33	0.00
	p-value		(.16)				-	(.47)	(.47)	-
	Galton EB	19.00	4.16	14.84	32.47	14.93	0.00	0.00	0.00	0.00
	p-value		(.06)				-	-	-	-

Variation in the regular hyperparameters H=60 moths (historical window), E=12 months (ex-post window), L=108 months (learning window) and N=50 (number of stocks with the biggest market caps). All three main Galton strategies are represented in the table (Galton EMC corresponds to the Galton where microcaps are excluded), as well as the value weighted. The OOS period ranges from January of 2011 until December of 2021 for the first variation, but begins in January of 2012 for the second Galton and of 2017 for the last one (H=60, E=12, L=180, N=50). Panel A displays average weight and performance statistics: the first four columns are the Sharpe ratio (average of the OOS months), the average active share (in percentage and with the value weighted strategy as a benchmark), the average turnover (in percentage) and the bankruptcy rate (in percentage; portion of months with realized excess portfolio's return below -100%). The last five columns are the average minimum and maximum weights (maximum and minimum is observed for each of the OOS months separately and then averaged), the mean and the standard deviation of the monthly weights' volatilities, and the average of the sum of negative weights (sum by month and then average the OOS months). All these five columns' values are in percentage. Panel B displays average expected and realized statistics for the monthly portfolios OOS: the first two columns represent the annualized, average, expected excess returns and the realized ones of the respective portfolio. P-values for the second column (in parenthesis) are referent to the difference between realized and expected returns being equal to zero. The third column is the RMSFE for the excess returns. The fourth and fifth columns are the annualized expected volatility of returns and the realized one. The last four columns are the hit rates for the first and fifth bottom and top quantiles of the excess returns distribution. P-values for these columns are in parenthesis and are generated for the hypothesis of the hit rate being equal to the corresponding quantile. All values besides p-values are in percentage.

The first Galton strategy of Table 6 achieves a higher Sharpe ratio (by 0.13 units) than the regular, MV, Galton (SR is, now, positive, contrarily to the main strategy's), fact that is due to a positive realized return of 4.96%. Risk-wise, the method still overestimates volatility, which indicates a conservative strategy. In terms of weight statistics, there is not a significant difference for the regular Galton between N=50 and N=40, although the active share, turnover and absolute sum of negative weights are lower for the latter. However, when looking at the VaR, the regular Galton with N=40 does a better job in terms of the bottom 1%, as 1.52% of realized returns are below the 1% VaR, computed with the expected returns. Contrarily to the Galton with N=50, the one with N=40 does not produce any months out of the OOS period where the strategy goes bankrupt.

Still analyzing the first variation of the method (H=60, E=12, L=108, N=40)[34], both Galtons where microcap stocks are excluded achieve relatively similar weight statistics in comparison to their N=50 counterparts, and both present a higher RMSFE; also, in the N=40 version, Galton EMC is worse in terms of Sharpe ratio in respect to the N=50 one, and the same happens with Galton EB. Apart from these similarities, the Galton EMC (excluding microcaps) differs in behavior from the EB one in the fact that Galton EMC is a worse risk predictor when N=40, while Galton EB is better (even though error in risk estimation is very similar between N=40 and N=50 for this last strategy).

The fact that the Sharpe ratios increase [and becomes positive] for the regular strategy when N=40 can be considered unexpected: the sample where microcaps are excluded performing better than the one where all stocks are used might lead one to believe that the lower the N, the worse the effect of the inclusion of the bottom 20% stocks. Instead, the regular Galton's SR increased, while the ones of the two strategies where only the top 80% is included decreased (although they are both still higher that that of the regular strategy). This might be indicative of higher differences in behavior between microcaps and some of the stocks that were excluded from the portfolios in this robustness test [in comparison to the top 40 that remained in the portfolios].

The second variation of the hyperparameters bears somewhat similar results to the corresponding main strategy in terms of SR[35]: Galton EMC also achieves a very high Sharpe ratio, even though it is now higher than one; the regular Galton obtains a negative ratio, just

[34] Table A in the Appendix displays all methods previously presented in this dissertation with the H=60, E=12, L=108, N=40 adjustment.

[35] Table C in the Appendix displays the obtained results for all the strategies when H=60, E=24, L=108, N=50.

like its E=12 counterpart. Nevertheless, the Galton EB obtains a ratio that is more than double the E=12 one.

The average negative realized excess returns of the regular Galton with E=24, together with the positive results for the other two Galtons, are a further reinforcement that the bottom 20% stocks in terms of market capitalization behaves significantly differently to the rest of the stocks. While, for the Galton EMC, there is less reversal to the ex-post mean – higher mean returns' slope in absolute terms[36] – in the leaning period and this trend continues in the OOS one (the average slope found in the learning period is still a good predictor of the impact that historical mean returns have on ex-post ones, hence the small RMSFE), the regular Galton estimates a stronger mean reversal in returns (Table B in the Appendix; smaller slope in absolute terms, which hints a lesser impact of the past 60 months' data in predicting ex-post mean returns). This is not a good prediction for the top 50 stocks in the OOS period, leading to a large RMSFE of 37.45%). This fact is corroborated by the fact that the GMV Galton results in a positive Sharpe ratio (Panel A of Table C in the Appendix): once expected returns are ignored for weight optimization, this metric is positive, even though the SR is a measure of MV and the method in analysis minimizes expected variances.

In regard to the third variation, the regular Galton achieves significantly worse results than the main one (or even than the previously two versions represented in Table 6). This was due to an even worse prediction of the mean returns, which is curious, as only L was changed and the historical period remains at 60 months. The higher learning period might have led to an 'overfitting' of the metrics' coefficients (only with the sample that includes microcaps): the coefficients[37] captured a trend [which is learnt throughout 180 months] that is not followed by the 50 biggest stocks during the OOS period [of 60 month – from January of 2017 until December of 2021]. When excluding the bottom 20% stocks in terms of size, the larger learning period actually helps: a Sharpe ratio of 1.75 is achieved, which is bigger than the corresponding one with L=108 months (0.93). It is interesting to notice that, once again, the Galton GMV achieves a positive SR, despite the negative regular Galton's result.

When comparing the Galton strategies to the naïve, value weighted, strategy, some trends are noticeable. The Sharpe ratios of the Galton methods do not increase with that of the benchmark strategy. From the first variation of the version (H=60, E=12, L=108, N=40) to the second

[36] Table B in the Appendix displays the obtained coefficients when H=36, E=12, L=132, N=50.
[37] Table D in the Appendix displays the obtained coefficients when H=60, E=12, L=180, N=50.

(H=60, E=24, L=108, N=50), the SR of the value weighted increases by 0.16 units; contrarily, the regular Galton turns negative; the Galton EB more than doubles its SR, mostly due to a significantly lower RMSFE. However, from the second to the third version (H=60, E=12, L=180, N=50), while the value weighted' SR still increases, the Galton EB' ratio decreases. The regular Galton's SR increases in its negativity. Like the Galton EB, the Galton EMC significantly increases its Sharpe from the first to the second variation, but also increases in the third one. This shows that, even though a better value weighted performance might hint a better sample of top stocks, the Galton estimates (coefficients) actively change the monthly portfolios' results: for example, when, from the second to the third variation, the value weighted gets better, the Galton EMC gets better, while the Galton EB also gets worse results; this shows that, in this case, the limits imposed on the slopes prevented Galton [EB] from increasing with the naïve portfolio, contrarily to the unrestricted version.

Limitations

The present dissertation presents some relevant and potentially a hindrance to the accuracy of the findings and conclusions that ought to be taken from them.

Firstly, the number of stocks to contribute to the estimation of the Galton coefficients is substantially smaller in the present paper than in Barroso and Saxena (2022). This is because Chinese stocks have had shorter lifespans than US', making the sample of stocks with full data on the previous 60 months and on the subsequent 12 months less numerous than the corresponding USA one – the average month in the main sample in analysis has 474.04 stocks that fulfill this requisite, while the US, for the period analyzed in Barroso and Saxena (2022), has an average of 2991.97 stocks. The above-mentioned limitation also led to the impossibility of running robustness tests that involved extending the historical period in one calendar year.[38]

The fact that the empirical Bayesian version of the Galton has only been performed on the sample of the 80% biggest stocks is also a shortcoming for the analysis. However, even though this analysis has not been made, one could expect the EB Galton with the microcaps to produce a lower SR than the regular strategy, as well as a worse risk prediction, as it was the case when comparing the Galton EB that excludes microcaps to the main Galton without the bottom 20% stocks.

[38] In the original paper, the authors find a better Sharpe ratio when H is changed to 180 (only H changes).

Another relevant drawback of the present analysis that is important to be born in mind is the fact the dataset contains a smaller timespan than the one used in Barroso and Saxena (2022), and it does not start or end in the same dates as the latter. This is a barrier to a cleaner comparison of results, as, ideally, only one aspect would be altered – either the date interval or the market. More specifically, the shorter OOS periods of the present study might also hinder the robustness of the results, as they are implicitly more random than when computed for longer timespans.

Conclusion & future research suggestions

The Galton correction of the MV optimization's inputs produces polarizing results: the method, on its regular form, achieves a negative Sharpe ratio; when microcaps are excluded, however, the estimates produce a significantly high average ratio. This strategy was also more successful in terms of SR in Barroso and Saxena (2022), however at a lower scale – in the just mentioned paper, the authors find that this difference is of only 0.05 units[39], compared to a difference of 0.95 units in the present dissertation. I also find that the Galton EB, which also excludes microcaps, produces, on average, a worse ratio than the simple Galton where the bottom 20% stocks are excluded, fact that is contradictory to the Barroso and Saxena (2022)'s results. This might mean that the long-term reversal that is reflected in the Galton corrections actually add value to the optimization process: when the coefficients are restricted between 0 and 1, implying that most of the historical mean returns will be multiplied by zero [as slopes are, mostly, negative], and that 100% of the estimation will be the grand mean, the SR decreases when compared to its unrestricted counterpart; thus, the reversal that happens when a negative slope is multiplied by the historical mean substantially contributes to the reduction of the OOS error and, therefore, to better estimates.

In terms of risk prediction, the Galton is the only shrinkage strategy to consistently overestimate volatilities. Furthermore, even though the risk prediction is worse for the present sample than for that of Barroso and Saxena (2022), all three Galtons do better at this than the other optimized strategies and the plain Markowitz.

Moreover, the robustness tests performed in this dissertation suggest that, for China [and in the period in analysis], one should not decrease the number of the portfolio constituents, as, when N=40, the unrestricted Galton where microcaps are excluded' Sharpe ratio increases (although

[39] Difference between Sharpe ratios.

the regular Galton's ratio becomes positive, the Galton EMC's performance is consistently the highest; thus it should be the preferred strategy among the three); extending E to 24 months also seems beneficial to both strategies where microcaps are excluded (this is not the case for the regular Galton, as the Sharpe becomes even more negative); extending L to 180 months is also recommendable for the two samples where the bottom 20% of stocks is excluded.

As the robustness checks, as well as the main study, show a tendency for the Galton that excludes the bottom 20% stocks to outperform the regular Galton in terms of SR, I would suggest further reducing the sample used to compute the coefficients to try to understand if the OOS results are better. This is especially true for China [in comparison to the US], as the difference in performance between the Galton EMC and the regular one is substantially high, to the point where excluding the bottom stocks turns a negative Sharpe into a positive one (close to or even higher than 1).

A further exercise of regressing the returns of the Galton strategy where microcaps are excluded on a factor model – CAPM, FF3, for instance – would be interesting to understand if this method produces abnormal results.

Appendix

Table A - OOS results when H=60, E=12, L=108, N=40

Panel A - OOS performance

	Sharpe	Active Share	Turnover	Bankruptcy rate	Min *w*	Max *w*	Mean StDev *w*	StDev StDev *w*	Sum Negative *w*
Value weighted	0.36	0.00	6.76	0.00	0.44	14.58	3.15	0.61	0.00
Talmud 1/N	0.13	42.63	2.23	0.00	2.00	2.00	0.00	0.00	0.00
Markowitz	-2.01	5039.00	10203.44	9.85	-981.06	860.72	339.45	782.66	-5038.16
Galton	0.11	305.37	137.18	0.00	-43.30	44.22	19.24	8.41	-264.75
Galton EMC	0.64	498.03	342.04	1.52	-60.80	64.10	30.27	30.30	-446.93
Galton EB	0.14	61.88	21.34	0.00	-6.57	11.03	4.01	0.58	-25.73
Galton Cov	0.19	275.25	129.95	0.00	-41.62	38.08	17.51	8.34	-237.52
Galton GMV	0.15	83.09	19.57	0.00	-7.15	13.98	4.68	0.90	-36.73
Markowitz (JS)	-0.76	1561.07	4737.05	2.27	-277.23	296.49	109.79	684.24	-1534.44
Elton Gruber (JS)	0.42	223.01	399.50	0.76	-30.39	44.50	15.48	25.25	-200.83
Ledoit Wolf (JS)	0.66	293.57	614.23	0.76	-51.85	54.46	20.75	40.76	-267.61

Panel B - OOS returns, volatilities and hit rates

	Expected r	Realized r	RMSFE	Expected σ	Realized σ	r<Qz(1%)	r<Qz(5%)	r>Qz(95%)	r>Qz(99%)
Value weighted		6.67			18.63				
Talmud 1/N		2.75			20.94				
Markowitz	455.23	-503.14	958.37	108.00	250.02	34.85	44.70	15.91	12.88
p-value		(.00)				(.00)	(.00)	(.00)	(.00)
Galton	38.52	4.96	33.56	50.63	44.19	1.52	5.30	1.52	0.00
p-value		(.04)				(.63)	(.88)	(.00)	-
Galton EMC	64.37	43.09	21.28	75.28	67.76	0.76	1.52	5.30	0.00
p-value		(.52)				(.75)	(.00)	(.88)	-
Galton EB	19.71	2.99	16.72	33.34	20.78	0.76	2.27	0.76	0.00
p-value		(.02)				(.75)	(.04)	(.00)	-
Galton Cov	37.49	7.83	29.66	48.50	41.80	1.52	5.30	3.30	1.52
p-value		(.05)				(.63)	(.88)	(.19)	(.63)
Galton GMV	15.70	2.21	13.49	30.08	14.89	0.00	0.00	0.76	0.00
p-value		(.11)				-	-	(.75)	-
Markowitz (JS)	536.59	-62.98	599.57	38.53	83.37	37.88	45.45	18.94	17.42
p-value		(.12)				(.00)	(.00)	(.00)	(.00)
Elton Gruber (JS)	125.20	15.00	110.20	24.06	35.93	21.21	29.55	6.82	4.55
p-value		(.00)				(.00)	(.00)	(.41)	(.05)
Ledoit Wolf (JS)	147.15	23.12	124.03	24.65	35.10	21.97	30.30	6.82	5.30
p-value		(.00)				(.00)	(.00)	(.41)	(.03)

Panel A represents the OOS performance. The first portfolio is formed in January of 2011 and the last in December of 2021. Historical periods range from January of 2006 until November of 2021. The first four columns are the average Sharpe ratio, the average active share (value weighted strategy as a benchmark), the average turnover and the bankruptcy rate (portion of months with realized excess portfolio's return below -100%). The last five columns are the average minimum and maximum weights (observed for each of the 132 months separately and then averaged), the mean and the standard deviation of the monthly weights' volatilities, and the average of the sum of negative weights. All values except SR in percentage. Panel B represent the OOS results: expected versus realized returns and risk metrics. First two columns represent the annualized, average, expected excess returns and the realized ones of the respective portfolio. P-values for the second column (in parenthesis) are referent to the difference between realized and expected returns being equal to zero. The third column is the root mean square forecast error for the excess returns. The fourth and fifth columns are the annualized expected volatility of returns and the realized one. The last four columns are the hit rates for the first and fifth bottom and top quantiles of the excess returns distribution. P-values for these columns are in parenthesis and are generated for the hypothesis of the hit rate being equal to the corresponding quantile. All values besides p-values are in percentage.

Table B - Fama and MacBeth (1973) coefficients for when H=60, E=24, L=108, N=50

Panel A - All stocks

	Covariance	Correlation	Variance	Mean returns
Intercept	0.01	0.36	0.02	0.01
t-stat (=0)	83.01	104.07	11.09	10.52
Slope	0.33	0.25	0.23	-0.23
t-stat (=0)	48.46	34.86	3.37	-3.13
Greater than 0 (%)	100.00	100.00	75.44	4.39
t-stat (=1)	-164.39	-110.86	-14.21	-16.54
Smaller than 1 (%)	99.56	100.00	100.00	100.00
R-squared (%)	6.75	3.28	3.38	3.36
Min.	16290.00	16290.00	181.00	181.00
Average	96684.25	96684.25	431.38	431.38
Max.	174936.00	174936.00	592.00	592.00

Panel B - Excluding microcaps

	Covariance	Correlation	Variance	Mean returns
Intercept	0.01	0.43	0.02	0.02
t-stat (=0)	61.50	74.78	9.88	9.08
Slope	0.28	0.17	0.25	-0.32
t-stat (=0)	24.90	14.36	2.91	-2.99
Greater than 0 (%)	100.00	96.05	78.07	3.07
t-stat (=1)	-111.79	-77.20	-11.99	-13.34
Smaller than 1 (%)	98.25	100.00	98.25	100.00
R-squared (%)	4.62	1.65	3.18	4.42
Min.	10153.00	10153.00	143.00	143.00
Average	41174.92	41174.92	272.15	272.15
Max.	76636.00	76636.00	392.00	392.00

Fama and MacBeth (1973) regressions of ex-post values of historical ones. Coefficients were computed from January of 2012 until the same month of 2021 with data ranging from January of 1956 until November of 2021. The first four rows of Panel A represent: 1) the average intercept of the regressions; 2) its average t-statistic; 3) the average slope; 4) its average t-statistic (both t-stats are Newey-West, 1987, with 12 lags). Rows five to seven are: 5) the proportion of t-statistics that corroborate with the hypothesis that the slope is greater than zero with a significance level of 5%; 6) the Newey-West (1987) t-statistic for the hypothesis that the slope is equal to one; 7) the proportion of t-statistics that, for a significance level of 5%, cannot exclude the hypothesis that the slope is smaller than one. Row eight is the average R-squared of the regressions and the last three rows represent the minimum, average and maximum number of stocks to be included in a monthly regression, respectively. Regressions are done pairwise for covariances and correlations (first two columns) and stock by stock for variances and excess mean returns. Panel B has the equivalent value for the top 80% of market capitalization stocks.

Table C - OOS results when H=60, E=24, L=108, N=50

Panel A - OOS performance

	Sharpe	Active Share	Turnover	Bankruptcy rate	Min *w*	Max *w*	Mean StDev *w*	StDev StDev *w*	Sum Negative *w*
Value weighted	0.52	0.00	6.22	0.00	0.30	13.18	2.77	0.52	0.00
Talmud 1/N	0.34	45.52	2.23	0.00	2.00	2.00	0.00	0.00	0.00
Markowitz	-0.88	4793.04	11500.81	5.83	-881.14	748.77	269.69	980.41	-4784.19
Galton	-0.03	309.38	133.99	0.00	-38.85	36.64	15.79	5.02	-267.68
Galton EMC	1.32	389.13	215.27	0.00	-37.83	42.46	18.83	5.41	-338.84
Galton EB	0.46	71.09	24.38	0.00	-7.18	10.21	3.71	0.45	-34.17
Galton Cov	0.06	275.81	122.61	0.00	-36.29	32.24	14.29	3.18	-237.25
Galton GMV	0.25	87.50	21.15	0.00	-6.23	11.73	3.96	0.73	-40.81
Markowitz (JS)	-0.68	3133.53	9206.69	2.50	-425.04	456.75	166.85	711.49	-3104.75
Elton Gruber (JS)	1.00	233.23	410.91	0.00	-28.54	42.18	13.37	25.55	-210.98
Ledoit Wolf (JS)	0.10	361.57	842.23	1.67	-49.54	57.73	20.40	52.96	-336.10

Panel B - OOS returns, volatilities and hit rates

	Expected r	Realized r	RMSFE	Expected σ	Realized σ	r<Qz(1%)	r<Qz(5%)	r>Qz(95%)	r>Qz(99%)
Value weighted		9.87			18.85				
Talmud 1/N		7.02			20.65				
Markowitz	332.05	-159.61	491.66	47.34	180.62	50.00	51.67	29.17	26.67
p-value		(.03)				(.00)	(.00)	(.00)	(.00)
Galton	36.16	-1.29	37.45	48.16	43.49	2.50	6.67	3.33	0.00
p-value		(.04)				(.29)	(.46)	(.31)	-
Galton EMC	44.23	61.66	17.43	55.31	46.78	0.00	2.50	7.50	0.00
p-value		(.28)				-	(.08)	(.30)	-
Galton EB	19.02	9.11	9.91	33.99	19.80	0.83	1.67	0.83	0.00
p-value		(.18)				(.84)	(.00)	(.00)	-
Galton Cov	35.30	2.60	32.70	46.81	40.60	1.67	6.67	4.17	2.50
p-value		(.03)				(.57)	(.46)	(.65)	(.29)
Galton GMV	15.97	3.47	12.50	30.68	13.70	0.00	0.00	0.83	0.00
p-value		(.43)				-	-	(.84)	-
Markowitz (JS)	667.20	-113.97	781.17	29.62	167.05	52.50	56.67	30.83	26.67
p-value		(.07)				(.00)	(.00)	(.00)	(.00)
Elton Gruber (JS)	113.60	35.06	78.54	20.72	35.14	23.33	31.67	10.83	4.17
p-value		(.00)				(.00)	(.00)	(.04)	(.08)
Ledoit Wolf (JS)	149.54	3.83	145.71	26.09	36.73	24.17	30.00	10.83	6.67
p-value		(.00)				(.00)	(.00)	(.04)	(.01)

Panel A represents the OOS performance. The first portfolio is formed in January of 2012 and the last in December of 2021. Historical periods range from January of 2007 until November of 2021. The first four columns are the Sharpe ratio (average of the OOS months), the average active share (with the value weighted strategy as a benchmark), the average turnover and the bankruptcy rate (portion of months with realized excess portfolio's return below -100%). The last five columns are the average minimum and maximum weights (maximum and minimum is observed for each of the 120 months separately and then averaged), the mean and the standard deviation of the monthly weights' volatilities, and the average of the sum of negative weights (sum by month and then average the 120 months). Panel B represent the OOS results: expected versus realized returns and risk metrics. First two columns represent the annualized, average, expected excess returns and the realized ones of the respective portfolio. P-values for the second column (in parenthesis) are referent to the difference between realized and expected returns being equal to zero. The third column is the root mean square forecast error (RMSFE) for the excess returns. The fourth and fifth columns are, respectively, the annualized expected volatility of returns and the realized one. The last four columns are the hit rates for the first and fifth bottom and top quantiles of the excess returns distribution. P-values for these columns are in parenthesis and are generated for the hypothesis of the hit rate being equal to the corresponding quantile. All values besides p-values and SR are in percentage.

Table D - Fama and MacBeth (1973) coefficients when H=60, E=12, L=180, N=50

Panel A - All stocks

	Covariance	Correlation	Variance	Mean returns
Intercept	0.01	0.32	0.01	0.01
t-stat (=0)	59.54	77.74	8.65	6.98
Slope	0.39	0.28	0.27	-0.26
t-stat (=0)	40.23	31.98	2.79	-2.36
Greater than 0 (%)	99.17	100.00	70.00	9.17
t-stat (=1)	-128.69	-85.64	-11.66	-11.65
Smaller than 1 (%)	92.50	100.00	94.58	100.00
R-squared (%)	4.53	2.54	2.64	3.12
Min.	17578.00	17578.00	188.00	188.00
Average	117727.05	117727.05	474.04	474.04
Max.	271216.00	271216.00	737.00	737.00

Panel B - Excluding microcaps

	Covariance	Correlation	Variance	Mean returns
Intercept	0.01	0.39	0.02	0.02
t-stat (=0)	43.79	54.20	7.83	6.17
Slope	0.32	0.21	0.28	-0.40
t-stat (=0)	19.26	13.99	2.55	-2.53
Greater than 0 (%)	92.08	95.83	65.83	3.75
t-stat (=1)	-83.17	-57.10	-10.35	-9.85
Smaller than 1 (%)	93.75	100.00	94.17	100.00
R-squared (%)	3.20	1.52	2.83	4.00
Min.	11476.00	11476.00	152.00	152.00
Average	51621.43	51621.43	305.07	305.07
Max.	95266.00	95266.00	437.00	437.00

Fama and MacBeth (1973) regressions of ex-post values of historical ones. Coefficients were computed from January of 2017 until the same month of 2021 with data ranging from January of 1956 until November of 2021. The first four rows of Panel A represent: 1) the average intercept of the regressions; 2) its average t-statistic; 3) the average slope; 4) its average t-statistic (both t-stats are Newey-West, 1987, with 12 lags). Rows five to seven are: 5) the proportion of t-statistics that corroborate with the hypothesis that the slope is greater than zero with a significance level of 5%; 6) the Newey-West (1987) t-statistic for the hypothesis that the slope is equal to one; 7) the proportion of t-statistics that, for a significance level of 5%, cannot exclude the hypothesis that the slope is smaller than one. Row eight is the average R-squared of the regressions and the last three rows represent the minimum, average and maximum number of stocks to be included in a monthly regression, respectively. Regressions are done pairwise for covariances and correlations (first two columns) and stock by stock for variances and excess mean returns. Panel B has the equivalent value for the top 80% of market capitalization stocks.

Table E - OOS results when H=60, E=12, L=180, N=50

Panel A - OOS performance

	Sharpe	Active Share	Turnover	Bankruptcy rate	Min *w*	Max *w*	Mean StDev *w*	StDev StDev *w*	Sum Negative *w*
Value weighted	0.66	0.00	6.43	0.00	0.32	10.84	2.41	0.12	0.00
Talmud 1/N	0.40	43.77	2.77	0.00	2.00	2.00	0.00	0.00	0.00
Markowitz	-0.06	5337.92	13829.68	3.33	-1206.88	943.99	319.46	1332.55	-5315.06
Galton	-0.57	301.17	125.56	0.00	-37.40	37.24	15.25	2.12	-253.46
Galton EMC	1.75	368.33	211.91	0.00	-37.20	41.94	17.84	4.89	-318.04
Galton EB	0.28	73.07	22.25	0.00	-7.99	10.10	3.86	0.36	-36.41
Galton Cov	-0.62	288.52	119.46	0.00	-35.73	34.03	14.52	1.95	-244.23
Galton GMV	0.18	85.57	21.10	0.00	-5.78	10.79	3.82	0.74	-38.78
Markowitz (JS)	1.44	1465.14	3332.70	0.00	-272.22	226.60	82.91	160.16	-1429.92
Elton Gruber (JS)	1.63	285.58	571.54	0.00	-38.51	50.34	16.12	35.07	-258.43
Ledoit Wolf (JS)	0.88	332.84	689.73	1.67	-52.25	51.14	19.12	38.13	-302.02

Panel B - OOS returns, volatilities and hit rates

	Expected r	Realized r	RMSFE	Expected σ	Realized σ	r<Qz(1%)	r<Qz(5%)	r>Qz(95%)	r>Qz(99%)
Value weighted		9.57			14.55				
Talmud 1/N		5.55			13.75				
Markowitz	364.14	-9.61	373.75	43.99	171.30	41.67	43.33	40.00	36.66
p-value		(.15)				(.00)	(.00)	(.00)	(.00)
Galton	35.06	-21.66	56.72	45.50	37.72	1.67	8.33	1.67	0.00
p-value		(.00)				(.69)	(.35)	(.04)	-
Galton EMC	43.97	68.34	24.37	52.25	39.02	0.00	3.33	3.33	0.00
p-value		(.16)				-	(.47)	(.47)	-
Galton EB	19.00	4.16	14.84	32.47	14.93	0.00	0.00	0.00	0.00
p-value		(.06)				-	-	-	-
Galton Cov	35.58	-23.33	58.91	45.55	37.49	1.67	5.00	1.67	0.00
p-value		(.00)				(.69)	(1.00)	(.04)	-
Galton GMV	14.52	2.40	12.12	29.25	13.44	0.00	0.00	0.00	0.00
p-value		(.37)				-	-	-	-
Markowitz (JS)	311.69	109.89	201.80	12.93	76.20	53.33	58.33	35.00	26.67
p-value		(.16)				(.00)	(.00)	(.00)	(.00)
Elton Gruber (JS)	144.99	66.57	78.42	24.13	40.79	25.00	35.00	11.67	1.67
p-value		(.10)				(.00)	(.00)	(.11)	(.69)
Ledoit Wolf (JS)	169.92	33.93	135.99	23.72	38.49	25.00	31.67	11.67	5.00
p-value		(.03)				(.00)	(.00)	(.11)	(.16)

Panel A represents the OOS performance. The first portfolio is formed in January of 2017 and the last in December of 2021. Historical periods range from January of 2012 until November of 2021. The first four columns are the Sharpe ratio (average of the OOS months), the average active share (in percentage and with the value weighted strategy as a benchmark), the average turnover (in percentage) and the bankruptcy rate (in percentage; portion of months with realized excess portfolio's return below -100%). The last five columns are the average minimum and maximum weights (maximum and minimum is observed for each of the 60 months separately and then averaged), the mean and the standard deviation of the monthly weights' volatilities, and the average of the sum of negative weights (sum by month and then average the 60 months). All these five columns' values are in percentage. Panel B represent the OOS results: expected versus realized returns and risk metrics. First two columns represent the annualized, average, expected excess returns and the realized ones of the respective portfolio. P-values for the second column (in parenthesis) are referent to the difference between realized and expected returns being equal to zero. The third column is the root mean square forecast error (RMSFE) for the excess returns. The fourth and fifth columns are, respectively, the annualized expected volatility of returns and the realized one. The last four columns are the hit rates for the first and fifth bottom and top quantiles of the excess returns distribution. P-values for these columns are in parenthesis and are generated for the hypothesis of the hit rate being equal to the corresponding quantile. All values besides p-values are in percentage.

References

Barroso, P., and Saxena, K., 2022. Lest We Forget: Learn from Out-of-Sample Forecast Errors When Optimizing Portfolios. *The Review of Financial Studies* 35:1222–1278.

De Bondt, W. F., and Thaler, R. (1985). Does the stock market overreact?. *The Journal of finance* 40(3):793-805.

DeMiguel, V., Garlappi, L., and Uppal, R., 2009. Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *Review of Financial Studies* 22:1915–53.

De Nard, G., 2022. Oops! I Shrunk the Sample Covariance Matrix Again: Blockbuster Meets Shrinkage. *Journal of Financial Econometrics* 20: 569–611.

Elton, E., and Gruber, M., 1973. Estimating the dependence structure of share prices – implications for portfolio selection. *Journal of Finance* 28:1203–32.

Fama, E., and MacBeth, J., 1973. Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy* 81:607–36.

Golosnoy, V., and Okhrin, Y., 2009. Flexible shrinkage in portfolio selection. *Journal of Economic Dynamics & Control* 33: 317-328.

Goyal, A., and Wahal, S., 2015. Is Momentum an Echo? *Journal of Financial and Quantitative Analysis* 50:1237-1267.

James, W., and Stein, C., 1961. Estimation with quadratic loss. In *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability 1*. Berkeley: University of California Press.

Johnston, M., 2022. China's Stock Markets vs. U.S. Stock Markets. *Investopedia*, at https://www.investopedia.com/articles/investing/092415/chinas-stock-markets-vs-us-stock-markets.asp.

Jorion, P., 1986. Bayes-Stein estimation for portfolio analysis. *Journal of Financial and Quantitative Analysis* 21:279–92.

Ledoit, O. and Wolf, M., 2003. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance* 10: 603-621.

Ledoit, O., and Wolf, M., 2004. Honey, I shrunk the sample covariance matrix. *Journal of Portfolio Management* 30:110–19.

Markowitz, H., 1952. Portfolio selection. *Journal of Finance* 7:77–91.

Mclean, R., and Pontiff, J., 2016. Does academic research destroy stock return predictability? *Journal of Finance* 71:5–32.

Stein, C., 1964. Inadmissibility of the usual estimator for the variance of a normal distribution with unknown mean. *Annals of the Institute of Statistical Mathematics* 16:155–60.