Lest we forget: using out-of-sample errors in portfolio optimization [1]

Pedro Barroso[2]

First version: December 2015

This version: June 2017

[1]The present work has benefited from comments and suggestions of Andrew Detzel, David Feldman, Grigory Vilkov, José Corrêa Guedes, Konark Saxena, Li Yang, Paul Karehnke, Rossen Valkanov, Pedro Santa-Clara, Victor DeMiguel, and Xing Jin. I also thank numerous seminar participants at Australian National University, Católica-Lisbon School of Business and Economics, Deakin University, Schroders' Official Institutions Seminar (Sydney) and the European Financial Management Association, Financial Management Association (FMA) Asia, FMA Europe, and Portuguese Finance Network conferences in 2016. I thank in particular the discussants Dawei Fang, Luis Vicente, Martin Young, and Ralf Kellner for their suggestions.

[2]University of New South Wales. E-mail: p.barroso@unsw.edu.au.

Abstract

Portfolio optimization usually struggles in realistic out of sample contexts. I deconstruct this stylized fact comparing historical estimates of the inputs of portfolio optimization with their subsequent out of sample counterparts. I confirm that historical estimates are often very imprecise guides of subsequent values but also find this lack of persistence varies significantly both across inputs and sets of assets. The resulting estimation errors are not entirely random. They have predictable patterns and can be partially reduced using their own history. Correcting inputs using past errors results in portfolio performance that reinforces the case for optimization versus naïve allocation rules.

JEL classification: G11; G12; G17.

Keywords: Portfolio optimization; out-of-sample robustness; covariance matrix; risk management; big data.

1. Introduction

"Those who cannot remember the past are condemned to repeat it."

George Santayana, *The Life of Reason*

The entire field of asset pricing is built on the foundation of modern portfolio theory laid out by Markowitz (1952). Markowitz shows how an investor with mean-variance utility should form portfolios given the expected returns on a set of assets and the respective covariance matrix. In a more realistic setting though, investors have to make their portfolio decisions in a context of uncertainty, with estimates learned from a sample instead of the true inputs (Detemple (1986), Dothan and Feldman (1986), and Brennan (1998)). It is well documented that the resulting estimation errors pose serious challenges to anyone pursuing the potential benefits of optimal diversification (e.g. Jobson and Korkie (1980), Michaud (1989)).

The implications are far reaching. A prominent line of research in asset pricing consists on searching new factors with significant alphas with respect to other previously known sets of factors. That discovery, when successful, shows that there is some ex post linear combination of the new factor with the old ones that improves the risk-return trade-off of the overall portfolio. The state of the literature on portfolio optimization suggests finding that same combination ex ante is far from straightforward. This raises the real possibility that an investor endowed with knowledge of some newly found priced factor does not necessarily achieve any improvement in his overall portfolio, at least in feasible optimization conditions.

The empirical challenge of portfolio optimization has spurred considerable research addressing the limitations of the initial Markowitz method. Still, DeMiguel et al. (2009b) show that in a demanding out-of-sample (OOS) environment, using only information available in real time, such optimized portfolios struggle to outperform simple benchmarks such as the 1/N allocation rule.[1]

The present paper proposes an alternative approach to OOS tests. In typical OOS tests, inputs are estimated from an historical sample, for example a window of the previous 60 months. The optimal portfolio is chosen in month t with inputs estimated from months $t-60$ to t and its subsequent performance in month $t+1$ recorded. The following period the procedure is re-iterated, with the historical sample rolled over one period, from months $t-59$ to $t+1$ and a new portfolio formed for month $t+2$. The motivation for this paper starts from observing that the OOS errors in these tests are recorded but not used in subsequent estimations, they are implicitly thrown away. After a long history of (usually large) OOS errors these should be of some use to correct the estimates obtained only from the historical sample. I examine this possibility and find that such correction has the potential to attenuate many of the well known limitations of portfolio optimization. In fact, the plain Markowitz method works quite well, OOS, once given the corrected inputs.[2] The correction only uses past OOS errors that could be known in real time, so it is still an OOS method. While an extensive research documents the existence of

[1] Some dismiss such OOS tests claiming they also benefit from hindsight and as such they are not necessarily preferable to in-sample tests. In my view, the claim about hindsight is likely true but its logic consequence should be exactly the opposite: OOS tests understate the real extent of uncertainty surrounding the portfolio decision. As such, resorting to these pseudo-OOS tests is the very least one should do when testing optimal strategies. In a sense it is encouraging to note the often dismal performance of optimized portfolios in those tests. It suggests they do not belittle true uncertainties, at least to implausible extremes as their in-sample counterparts do.

[2] Other methods that improve on the plain Markowitz optimization could in principle achieve even better results with the corrected inputs, that is a question outside the scope of the present version of this paper. I simply show that the pioneer version of portfolio optimization, with all of its known limitations, performs reasonably well with inputs corrected for past OOS errors.

large OOS errors, to the best of my knowledge this is the first paper that makes explicit use of those errors to improve the portfolio allocation decision.

The benefits of the resulting corrected covariance matrix are particularly significant for risk management. In an OOS exercise with the 50 largest stocks by capitalization, the mean-variance optimal risky portfolio (MV) has losses exceeding the 1% VaR approximately 40% of the months. This provides an extreme example of the substantial bias in risk estimates obtained from the historical sample. The hypothetical investor represented in this OOS test has to be endowed with an heroic persistence in the face of dis-confirming evidence: he continues estimating risk the same way month after month without ever correcting the large distance between his in-sample risk estimates and the respective ex post realizations. In sharp contrast, the corrected covariance matrix produces risk estimates very close to actual OOS risk. For example, losses exceeding the 1% VaR only happen in 1.33% of the OOS observations. This illustrates well the potential of correcting past OOS errors for risk management.

Past OOS errors in the inputs of optimization should be of no particular use if they had no structure. But it turns out they have a pattern that makes them predictable to some extent. Figure 1 illustrates the main stylized fact about the inputs of portfolio optimization for individual stocks.

[Insert figure 1 near here]

Suppose an investor uses the Fama and French (1993) risk factors to model the risk and expected return of a set of stocks. He estimates in a historical sample the betas of each stock with respect to each factor, their alphas and the correlation of residuals between all stocks in the universe. But as the figure shows all of these inputs regress considerably to the mean in the subsequent out of sample period.

Some of these patterns are well known. The mean regression in the betas is reminiscent of the results in Vasicek (1973), that of expected returns is known at least since Bondt and Thaler (1985). The fact that future alphas are nearly unrelated to their historical estimates should be expected by even the most lenient form of market efficiency. But surprisingly, the classic approach to portfolio optimization relies exactly on simply plugging the historical estimates in the optimization problem. Graphically, this amounts to expect the ex post values to lie on the 45 degree line with respect to the historical estimates. The simple patterns presented in figure 1 clearly advise otherwise.

Some optimization inputs are reasonably close, on average, to their past values (such as variances), while others mean reverse (e.g. stocks with above average alphas in the past tend to have below average alphas OOS). These stylized facts suggest a correction specific to each input based on its past reliability. I call this the Galton correction after Sir Francis Galton, who first proposed the concept of regression to the mean (Galton (1894)). This correction provides a simple and flexible method to estimate the covariance matrix of a set of stocks and also filters out most of the noise in estimating mean returns.

Besides individual stocks, I also examine the OOS persistence of optimization inputs in other four sets of test assets. Each set of test assets is composed of 25 double-sorted portfolios on stock characteristics (e.g. operating profitability and investment, the two new factors of Fama and French (2015)). I find different sets show different patterns of predictability. Generally, past returns tend to be much more informative about future returns than for individual stocks. As such the Galton correction produces more impressive results than for individual stocks (in terms of Sharpe ratio).[3] For example, in the set of portfolios sorted on size and momentum, the annualized Sharpe ratio of the optimized portfolio OOS is 1.37, versus 0.54 for the 1/N benchmark. On average across the four sets of assets considered the Sharpe ratio improves 80% on the 1/N rule. This illustrates the flexibility of the Galton correction. It filters out most of the information on expected returns in individual stocks (where it is mostly noise) but retains it in other test assets where it matters (as portfolios of stocks sorted on characteristics).

The paper is organized as follows. Section 2 briefly discusses the closely related literature. Section 3 takes a closer look to the evidence on regression to the mean in the inputs of Markowitz optimization. Section 4 proposes the method to correct for past OOS errors in the inputs and explains the construction of the alternative portfolio strategies. Section 5 shows the OOS performance of optimized portfolios with the largest 50 stocks by capitalization. Section 6 examines the predictability of the risk of those optimal stock portfolios. Section 7 shows the results of a bootstrap of 1000 simulations of similar sequences of samples of stocks randomly drawn every 12 months from the 500 largest firms by market capitalization. Section 8 examines the usefulness of the Galton correction on portfolios sorted on size, book to market, operating profitability, investment and other characteristics. Section 9 concludes.

[3]The optimization with individual stocks does not use any of the abundant evidence on the predictability of stock returns in the cross section (see for example Harvey et al. (2015), Green et al. (2013), and Lewellen (2014)). Therefore with individual stocks the most impressive gains are in risk management. Stock portfolios formed on characteristics already benefit from some return predictability in the cross section and so their gains go beyond risk management.

2. Related literature

This work is related to a recent literature that proposes robust optimization methods. DeMiguel et al. (2009a) show that imposing constraints on the vector of portfolio weights substantially improves OOS performance. Brandt et al. (2009) use asset characteristics to model weights directly, avoiding the issue of estimating both expected returns and the covariance matrix altogether. These two methods have one trait in common: they circumvent the issue of estimation error in the covariance matrix and focus instead on the final output of the optimization process: the portfolio weights. However, even if an investor is successful in estimating in real time sensible portfolio weights (quite a non-trivial task), he is still left with the problem of managing and estimating the risk of that portfolio. Comparatively this work focuses more on the estimation of the covariance matrix and using it for estimating risk.

Correcting the inputs for past OOS errors can be seen as a form of shrinkage.[4] In that sense, this paper is related to a vast literature on portfolio methods that rely on Bayesian approaches to estimation error (Barry (1974), Bawa et al. (1979), Jobson and Korkie (1980), Jobson and Korkie (1981), Jobson et al. (1979), Jorion (1985), Jorion (1986)). It is also related to the literature on shrinking the estimation error of the covariance matrix (Best and Grauer (1992), Ledoit and Wolf (2004a), Ledoit and Wolf (2004b), Disatnik and Benninga (2007)). Generally that literature shows that one effective way to reduce the estimation error in the covariance matrix

[4]This happens with the examined sets of assets because the most recurrent pattern is regression to the mean. Therefore the cross-sectional differences in the corrected inputs are smaller than the initial estimates. In general, correcting for past OOS errors in inputs does not forcefully result in shrinkage per se.

is to shrink the estimates to some target. The exact target to chose and the extent of shrinkage to apply is a matter of ongoing research (Benninga (2014), Ledoit and Wolf (2014)). By comparison, the approach I propose does not specify a target to shrink to and the amount of “shrinkage” is specific to each input (variances, correlations, expected returns) and each set of assets. I just let the data speak for itself and chose the adjustment in inputs that would most reduce past OOS errors. My results suggest choosing a priori an appropriate target is challenging as persistence varies according to the set of assets and the input variables considered.

The idea of using past out-of-sample errors to correct estimates is akin to cross-validation (e.g., Stone (1974),Shao (1993)) and jackknife estimators (e.g., Efron (1983)). Generally, these approaches examine different methods of splitting the sample, inferring estimates from one part of the sample and testing them on the other. I also estimate optimization inputs from historical samples each moment in time and test them on a different sample (“out-of-sample”) window of time. Notwithstanding, the Galton method differs from cross-validation in two important ways. First, given a specific sample of, say, 5 years of monthly returns for N stocks, the method does not split this sample in any way or throw observations out for out-of-sample testing. Instead, the method infers a workable correction from a grand sample of many other past stock returns (or pairs of stocks), most no longer active and so not even present in the sample used for the optimization. Second, the methodology focuses on out-of-sample testing to motivate the correction. This form of testing is at the crux of the literature on portfolio optimization and in this sense distinct from other possible ways to split data. It is also closer to the investors problem of estimating inputs of optimization in real time.

The Galton correction uses an extensive data set of variances and covariances for every individual stock and pair of individual stocks in the CRSP universe over a 60 year period. Given the high dimensionality of the dataset, especially in the covariance matrix, the approach can be classified as *big data*.

3. Regression to the mean in optimization inputs

The Markowitz (1952) approach shows how to solve for the optimal weights of a portfolio given the information on the expected returns of the assets available and the respective covariance matrix. The vector of relative weights of the optimal risky portfolio is:

$$w_t = \frac{\Sigma_t^{-1}\mu}{1_N \Sigma_t^{-1}\mu} \tag{1}$$

where μ is a N-by-1 vector of mean returns, 1_N is a N-by-1 vectors of ones, N is the number of assets, and Σ is the covariance matrix.[5] The inputs to solve the optimization are unknown in practice and have to be estimated. As DeMiguel et al. (2009b) mention, the classic "plug-in" approach solves this problem replacing the true mean and variances by their sample counterparts in some rolling window. In one out-of-sample testing framework, where the weights must be determined using only information available to each point in time, this amounts to estimating the inputs $\hat{\mu}$ and $\hat{\Sigma}$ in the historical sample. Implicitly, the approach relies on the strong assumption that historical sample moments offer the best estimate of their true unobservable counterparts. Throughout this paper, I call this the historical (or plug-in) method and denote the respective estimates as μ_H and Σ_H.

Goyal and Welch (2008) show that OOS the historical mean performs quite well

[5] In the results in the empirical sections I divide by $|1_N \Sigma_t^{-1}\mu|$. This prevents the cases when the negative denominator switches the sign of the relative weights in the complete portfolio.

predicting the equity premium when compared to most alternative methods based on predictive regressions. So it is not the general case that using the moment from a historical sample results in poor OOS estimates. But in the case of portfolio optimization, it is well known that historical estimates are plagued with large sampling errors (e.g. Michaud (1989), Kan and Smith (2008)) and hence result in poor OOS performance (DeMiguel et al. (2009b)). This motivates a comparison of the inputs of Markowitz optimization in historical samples with their ex post, out-of-sample, counterparts.

[Insert figure 2 about here]

Figure 2 shows the relation between the historical sample and the ex post periods for the entire universe of US stocks. This figure is similar to figure 1 above, but focuses on the more general case of estimating risk and returns without any particular risk model. Given the large number of asset pricing models available in the literature today, I choose to focus on this agnostic approach where no model is assumed as the truth and the optimization simply relies on past correlations and variances.[6] For each period and variable, the observations are sorted into deciles according to their values over the previous 60 months. The y-axis shows the average value for each decile in the subsequent 12 months.

It is apparent that past covariances and correlations are positively related to their future counterparts, but the slope is clearly below one. This shows that assuming the past value is the correct estimate, as in the historical approach, is on average excessive. But it also shows it should be sub-optimal to assume past correlations are best forecast by their cross section mean, as is the case in the

[6]The correction proposed in this paper could be equally implemented with a factor model though.

constant correlation matrix of Elton and Gruber (1973). The slope of the OOS values is clearly not zero either.

Panel C shows the variances are relatively well approximated by their past values. Panel D shows there is mean reversion in mean returns (a result known at least since Bondt and Thaler (1985)). So μ_H is actually negatively correlated with true expected returns. This result offers a simple explanation for the fact that the MV performs badly not only in terms of Sharpe ratio but also in expected returns. The method, when estimated using the plug-in approach, suffers from pervasive estimation error issues but this should not explain its consistently low returns. On top of this problem, the plot in panel D shows the method tends to overweight stocks with low expected returns and short those with high expected returns.

Table 1 presents the results of Fama and MacBeth (1973) predictive regressions for covariances, correlations, variances, and mean returns on their historical estimates. For each month, I run a regression of ex post OOS values on their respective historical estimates. This regression draws power from a very large cross sample. For instance, a set of N assets has N(N-1)/2 covariances. This implies, as the last row of the table shows, that the cross section of covariances is quite large, its maximum number of observations in a single monthly regression exceeds six million. Usually, the large number of covariances to estimate is pointed out as a limitation of optimization methods. But in this regression exercise, it is quite the opposite. The large cross section leads to a more accurate estimation of the correction to make.

[Insert table 1 about here]

The reported slopes and t-statistics are inferred from the time series averages of the Fama-MacBeth regression coefficients. For covariances, correlations, and variances the null hypothesis that the slope is zero (no predictability in the variable) is clearly rejected with t-statistics of 10.03, 21.44, and 11.07 respectively. The hypothesis of a slope of zero is rejected at the 5% level in every cross section regression for the covariance and correlation, and in 98.18% of the regressions for the variance. This strongly suggests optimization could do better than just using the constant correlation matrix of Elton and Gruber (1973).[7] So effectively almost all monthly regressions confirm some predictability in the optimization inputs.

On the other hand, the hypothesis that historical inputs are on average close to their true values is also strongly rejected. The intercepts of the regressions are all significantly positive at the 1% level. The null hypothesis that the slope coefficients of these variables are one (implying the historical approach is correct on average) is clearly rejected too. The t-statistics for covariances, correlations, and variances are, respectively, -17.12, -71.16, and -8.01. This is illustrative of the problems of the plug in approach that implicitly assumes a value of one for the slope (and zero for the intercept). In the case of mean returns, the slope coefficient is also significantly negative with a t-statistic of -4.83.

For covariances, correlations, variances, and mean returns, it is striking how almost all regressions reject the null hypothesis of a slope of one. This shows an hypothetical investor following the plug in approach should rapidly realize there is something wrong with his estimates. For most variables, one single regression of ex post values on ex ante historical inputs would be enough to strongly suspect of the

[7]For correlations, the R-square of the regression is quite low (only 1.59%). This shows that making inference with a relatively small number of assets, as in Elton and Gruber (1973), the constant correlation matrix should provide a very good approximation. By contrast, the high statistical significance of the slope coefficient in my results benefits from much larger cross section of pair-wise correlations.

existence of regression to the mean.

All in all, this section shows there is a clear regression to the mean in the covariance matrix. The best estimate for the future correlation between a pair of stocks is somewhere between the past correlation of the same pair of stocks and the mean correlation between all pairs of stocks. In the case of mean returns, there is even mean reversion. Those stocks the historical approach estimates to have high-expected returns assets, are in fact, on average, those with lower expected returns.

4. Correcting past out of sample errors

The previous section shows there are large differences between historical estimates of the optimization inputs and the values they assume on average OOS. But those differences are also consistent and predictable to some extent. For instance, above average historical correlations tend to become smaller OOS[8]. This leads to the possibility that correcting past OOS errors in the inputs can lead to more robust portfolio optimization.

Typical OOS tests use either an expanding or a rolling window, with observations up till time t, to produce a forecast for time $t+1$. Then this forecast is compared with the value observed at time $t+1$ and the corresponding OOS error is recorded. The following period the estimation window is either rolled over or expanded one period and a new forecast is produced for time $t+2$. But this forecast still only uses the in-sample information to produce the forecast. It ignores the OOS error

[8]This does not imply that there is some break in true correlations between the in-sample and the OOS period. Even if all true correlations are the same and do not change, some pairs of stocks should have high (small) sample correlations by randomness and this bares no information about their subsequent correlations.

obtained in the previous period, it is implicitly discarded. The approach proposed here consists in using those OOS errors to improve forecasting.

For any individual variable of interest X (variance, pairwise correlation, or mean return) let $X_{H,t}$ denote its historical estimate at time t computed from a rolling window of H observations. The value it assumes in a subsequent ex post window of E months is denoted by $X_{E,t}$. The results shown throughout this paper are for $H = 60$ and $E = 12$ (note that $X_{E,t}$ only becomes known at $t + E$).

For each period t in the sample, I run the cross-section regression:

$$X_{E,t,j} = g_{0,t} + g_{1,t} X_{H,t,j} + \epsilon_{t,j} \tag{2}$$

for $j = 1, ..., N_t$, where N_t is the number of stocks (for variances or mean returns) or pairs of stocks (for correlations) available in the sample at time t. If the historical approach is correct, the best estimate of $X_{E,t}$ is $X_{H,t}$ and so $g_{0,t} = 0$ and $g_{1,t} = 1$. If regression to the mean is total, the best estimate of $X_{E,t}$ is the cross section average, so $E(X_{E,t,j}) = g_{0,t}$ and $g_{1,t} = 0$.

The history of cross sectional estimates $\widehat{g_{0,t}}$ and $\widehat{g_{1,t}}$ can be used to form a corrected expectation of $X_{E,t,j}$, which I denote as $X_{G,t,j}$:

$$X_{G,t,j} = \bar{G}_{0,t-E} + \bar{G}_{1,t-E} X_{H,t,j} \tag{3}$$

where $\bar{G}_{0,t-E} = \sum_{s=1}^{t-E} \widehat{g_{0,s}}/(t-E)$ and $\bar{G}_{1,t-E} = \sum_{s=1}^{t-E} \widehat{g_{1,s}}/(t-E)$. This estimate, $X_{G,t,j}$, is the historical estimate, $X_{H,t,j}$, corrected by how close (or how far) all known past historical estimates were of their subsequent OOS values ($X_{E,t,j}$). It is totally agnostic about the data generating process and the distribution of

OOS errors. It consists in a linear correction of past OOS errors. Other more sophisticated methods would likely produce better corrections, but I refrain from that pursuit and focus instead on a straightforward linear function for its simplicity. As $X_{G,t,j}$ only uses information available until time t, it can be used for OOS tests. The correction should become more accurate as more past OOS errors are available. As such, besides the usual initial in-sample period needed in OOS tests, the results in this paper require one additional learning period (L) to correct for past OOS errors. Given this, the first truly OOS return for a strategy using $X_{G,t,j}$ will occur at time $H + L + E + 1$. I pick an arbitrary initial period (L) of 120 months for the correction. For the chosen values of H, L, and E that amounts to 193 months.[9]

Using the correction above, I compute corrected inputs for the Markowitz optimization. The corrected correlation matrix, $\rho_{G,t}$, has in each entry the corrected pairwise correlation and ones in the diagonal. Similarly, I obtain the N-by-1 vector of corrected estimates of the variances, $\sigma^2_{G,t}$, and mean returns, $\mu_{G,t}$. The Galton corrected covariance matrix is:

$$\Sigma_{G,t} = diag(\sigma_{G,t})\rho_{G,t}diag(\sigma_{G,t}) \tag{4}$$

A covariance matrix must meet the important requirement of being positive semi-definite, otherwise one could form portfolios with negative variances. This imposes constraints on the covariance matrix and shrinkage methods applied therein. This requirement is generally met for the correction I propose in a natural manner.

[9]Please note that after the initial learning period, for a given set of assets, the only requirement is to have H past observations. The method 'learns' from past OOS errors of similar assets. It does not require a record of past observations of 193 months for every asset. In fact, the correction converges quite fast to relatively stable values, just a pair of constants to correct each type of input.

This happens because $\bar{G}_{0,t-E} \approx \bar{\rho}_{t-E} - \bar{G}_{1,t-E} * \bar{\rho}_{t-E}$, where $\bar{\rho}_{t-E}$ is the grand mean of pairwise correlations in past samples up to time $t - E$. As a result, $\rho_{G,t} = (1 - \bar{G}_{1,t-E})\bar{\rho}_{t-E} + \bar{G}_{1,t-E} * \rho_{H,t}$. As $\bar{G}_{1,t-E}$ is empirically a number between 0 and 1 (see section 3.), this is the convex combination of two positive semi-definite matrices. As a result $\rho_{G,t}$ is a positive semi-definite matrix itself. Furthermore, I restrict the $\sigma_{G,t}$ to be strictly positive to machine precision.

From this the weights of the Galton mean-variance (MV) portfolio are:

$$w_{G,t}^{MV} = \frac{\Sigma_{G,t}^{-1}\mu_{G,t}}{1_N \Sigma_{G,t}^{-1}\mu_{G,t}} \tag{5}$$

Similarly, the Galton global minimum variance (GMV) portfolio is:

$$w_{G,t}^{GMV} = \frac{\Sigma_{G,t}^{-1}1_N}{1_N \Sigma_{G,t}^{-1}1_N} \tag{6}$$

The weights of these portfolios do not use any information about the characteristics of the stocks and are not adjusted ex post to respect any constraints. They are simply the result of plain Markowitz optimization applied to corrected inputs. I compare the results of these two portfolios with the MV and GMV obtainable using the historical approach and also the Elton and Gruber (1973) constant-correlation approach.

In the Elton and Gruber (1973) constant correlation approach, $\rho_{EG,t}$ consists of a matrix where the non-diagonal elements are the average of pairwise correlations in the rolling historical sample and the diagonal elements are all ones. Then the covariance matrix is:

$$\Sigma_{EG,t} = diag(\sigma_{H,t})\rho_{EG,t}diag(\sigma_{H,t}) \tag{7}$$

and $\sigma_{H,t}$ is the N-by-1 vector of estimated volatilities from the historical sample. The GMV portfolio of the Elton-Gruber approach is determined as in equation 6 but with $\Sigma_{EG,t}$ instead of $\Sigma_{G,t}$. I combine $\Sigma_{EG,t}$ with $\mu_{H,t}$ to obtain the weights of a Elton-Gruber MV portfolio. This is not the portfolio optimization method proposed in Elton and Gruber (1973) but I include it in the comparison for completeness.

The Elton and Gruber approach is related to a Galton correction in correlations only. If one sets the slope of the Galton correction to zero and the intercept equal to a constant then the Galton-corrected correlation matrix would be similar to the Elton-Gruber correlation matrix. The deviation from the Elton-Gruber correlation matrix reflects the usefulness of past correlations to predict subsequent correlations in the data.

Besides these four portfolios, I also compare the two Galton-corrected portfolios with the 1/N benchmark that DeMiguel et al. (2009b) show compares favourably with most optimization methods.

5. The OOS performance

The dataset consists of monthly returns of the entire universe of US listed stocks on the Center for Research in Security Prices (CRSP). The monthly returns data start in 1950:03 and end in 2010:12. At the start of the sample I pick the 50 stocks with the largest market capitalization with a complete history of returns in the previous 60 months and the subsequent 12 months. The set of stocks is kept fixed for 12 months and then renewed each 12 months until the end of the time series. For the OOS exercise, one initial period of 193 months is needed, this implies that the first OOS return is in 1966:04. Also, the requirement of a subsequent history of 12 months means that no OOS returns can be computed for the last 11 months, so the OOS period is of 526 monthly returns from 1966:04 to 2010:01. In total 43 universes of 50 stocks are sequentially chosen in the OOS period, one for each year ($526/12 = 43.83$).

Table 2 shows the OOS performance of the 6 portfolios. The historical GMV portfolio has a Sharpe ratio of 0.19, below the 1/N benchmark. The historical MV has a negative Sharpe ratio of -0.14 and a very high excess kurtosis of 515.78. This is illustrative of the well known problems of Markowitz optimization, at least when using the classic plug in approach. The simple 1/N portfolio compares favourably with historical approach with a Sharpe ratio of 0.29. This confirms with portfolios of individual stocks the result DeMiguel et al. (2009b) obtain with industry and size / book-to-market sorted portfolios.

The relative robustness of the 1/N rule is usually interpreted as a consequence of estimation error. Past returns are so noisy that an investor is better off totally ignoring the sample and chose a naïve allocation instead. Notwithstanding, Plyakha et al. (2012) show the 1/N rule with monthly rebalancing is not that naïve for individual stocks. The strategy effectively implies a constant rebalancing that buys (sells) recent one-month losers (winners). This amounts to systematically exploiting the short term reversal effect of Jegadeesh (1990). They further show that with less frequent rebalancing the outperformance of the 1/N rule diminishes considerably. In my discussion of the results I disregard this result and stick to the general interpretation of the 1/N as a natural naïve benchmark for optimized portfolios. This raises the bar for optimized portfolios.

The Elton-Gruber GMV portfolio has a very interesting performance, with a Sharpe ratio OOS of 0.45, more than 50% higher than the 1/N. On the other hand, the Elton-Gruber MV has a very poor performance, with a Sharpe ratio of -0.15 and, more importantly, with an extremely high standard deviation of 1224.86.

Both the Galton methods, the GMV and the MV, perform well with Sharpe ratios of 0.48 for the GMV and 0.43 for the MV, 65% and 48% higher than the 1/N benchmarks, respectively. Most strikingly, the method achieves sensible ex post volatilities of 12.71 (GMV) and 18.68 (MV).[10]

Ledoit and Wolf (2004a) derives an optimal shrinkage method for the covariance matrix assuming a multivariate normal distribution for asset returns. Stock returns are far from normal but still, in unreported results (available upon request), I find in this setting Ledoit and Wolf (2004a)'s covariance matrix successfully reduces the noise in the estimation of the covariance matrix. The results of the method are generally analogous to those of the Elton-Gruber method for individual stocks.

6. The predictability of risk of individual stock portfolios

Institutional investors often have relatively concentrated stock portfolios. Agarwal et al. (2013) show that the Herfindhal index of a typical mutual fund stock portfolio is 0.018 and that of a hedge fund is 0.047. This implies that the equivalent number of holdings, defined as the reciprocal of the Herfindhal index, is respectively 56 and 21 stocks. This concentration of the bulk of a portfolio in a relatively small set of securities can seem inefficient from a diversification perspective, but

[10]In unreported results, omitted for brevity but available upon request, I optimize the portfolios correcting only variances, correlations, or mean returns, both in isolation and two at a time. Generally, correcting mean returns and correlations jointly is the most important source of gains for the method.

Kacperczyk et al. (2005) show it is associated with superior performance once controlling for risk. So relatively small stock portfolios are important for institutional investors and perhaps for good reasons. To manage the risk of those portfolios, whatever the information set used to estimate returns - factors, characteristics, or privately produced fundamental analysis -institutional investors always need a reasonable estimate of the covariance matrix on its own merits, not just to pick the weights of the optimal portfolio. This should be particularly relevant to estimate the value-at-risk (VaR) for extreme quantiles of the distribution, particularly in the case of hedge funds pursuing long-short strategies or central clearing counterparties estimating margins requirements for its members.

The search of robust mean variance portfolios in OOS tests has received considerable more research efforts than the predictability of risk of those same portfolios. Yet, according with the two-fund separation theorem, in a conventional Markowitz setting, a hypothetical investor solves the allocation problem in two steps: i) identify ex ante the tangency risky portfolio; ii) determine the allocation of wealth between the tangent portfolio and the risk free rate depending on the portfolio's risk and his own preferences. The first step has received a lot of attention, and the difficulties of finding the mean-variance efficient portfolio in a realistic OOS setting are well documented. There are successful methods available that achieve robust OOS performance in terms of Sharpe ratio (Brandt et al. (2009), DeMiguel et al. (2009a), Kirby and Ostdiek (2012)). But even if an investor is able to solve the first step using one of those methods, he is still left with the problem of how to estimate, ex ante, the risk of his chosen portfolio. For that second step, it does not matter if the standard deviation of the optimal portfolio is high or low, but it does matter if it is predictable.

Basak (2005) show that using the historical method to estimate the risk of the GMV results in a dramatical understatement of its true risk OOS. Also, assuming a multivariate normal distribution, Kan and Smith (2008) show analytically that historical estimates of the risk and mean return of the GMV portfolio are systematically overly optimistic. Below, I examine this problem in an OOS setting with real stock data and add to the historical sample estimation two other methods: the constant correlation matrix of Elton and Gruber (1973) and the Galton correction using past OOS errors.

[Insert table 3 near here]

Table 3 shows the ex ante risk of each of the portfolios and the respective ex post OOS risk. An investor wary of the fact that mean returns are difficult to estimate might decide to follow the advice of Jobson et al. (1979) and pursue a GMV strategy. This investor would be quite surprised to see that the strategy, in real time, has about 7 times the risk he anticipates. The ex ante standard deviation of the historical GMV is of 3.60 percentage points (annualized) but the ex post standard deviation of the strategy is 26.50.

The problem is even worse for the historical approach MV portfolio with an OOS risk 113 times higher than the ex ante estimate. A hypothetical investor following the historical approach to estimate risk should soon conclude there is something wrong with his estimates. A clear illustration of this is that 39.67 percent of the OOS returns are losses exceeding the investor's ex ante estimate of the 1% level value-at-risk (VaR). This is particularly disturbing from a regulatory perspective as the vast majority of commercial banks rely on historical simulation methods to estimate value at risk (Perignon and Smith (2007), Pérignon and Smith

(2010)) and they feature prominently in regulations (see e.g. EMIR). This result shows that for some portfolios, designed to obtain an optimal risk-return trade-off in a historical sample, past hit rates are very misleading of true OOS risk.

The Elton-Gruber constant correlation approach performs much better predicting the risk of the GMV. The standard deviation of the GMV portfolio is 15.38 ex post versus 8.85 expected ex ante. So in the case of stock portfolios, the Elton-Gruber approach substantially reduces the dramatic problems shown in Basak (2005) with historical sample estimates. Still, an investor in the robust Elton-Gruber GMV portoflio, would find on average 74% more volatility OOS than anticipated. The number of occurrences in the extreme quantiles of the distribution is much higher than anticipated too. In the left tail, losses exceeding his estimate of 1% level VaR would occur 5.93 times more often than anticipated, the null hypothesis that the true hit rate is 1% is clearly rejected. For the MV portfolio, the constant correlation matrix does not capture accurately the OOS risk either. The standard deviation OOS is more than 7 times higher than the ex ante estimate. So while the constant correlation matrix achieves a somehow acceptable performance describing risk of the GMV, an investor tempted to use it to estimate the risk of a MV portfolio would be dramatically surprised.

The third column shows the performance of the portfolios that use the covariance matrix (and mean returns) corrected for past OOS errors. The most striking result is that there is no significant difference between ex ante and out of sample risk. For the GMV portfolio, the expected standard deviation is 15.24 percentage points while the standard deviation in the OOS period is 12.71. So risk is actually lower OOS than expected for this portfolio. None of the hit rates is significantly higher than the respective target rates, two are even significantly lower in a statistical sense.

Even for the case of the MV portfolio, OOS standard deviation (18.68 percentage points) is very close in magnitude to the estimated ex ante (18.20 percentage points). Most hit rates are not significantly different than their respective targets. So the covariance matrix corrected for past OOS errors captures well the risk of these optimized stock portfolios. This contrasts starkly with the other methods examined.

[Insert figure 3 about here]

Figure 3 shows the minimum variance estimates using the historical and Galton methods for the largest stocks in the end of the sample. Panel A shows that the minimum variance frontier estimated with the historical approach is very close to the y-axis and seems almost vertical when compared to the shape of the frontier using the Galton method.[11]

The historical method overestimates the potential benefits of optimal risk diversification. For instance, the GMV, with a 100% net exposure to the stock market, has an estimated annualized volatility of only 2.38%. This is far less than its historical volatility. Panel B shows the minimum variance frontier with a risk free rate asset available for the investor. The historical method estimates ex ante an annualized Sharpe ratio attainable for the investor of 10.98. This should overstate by a great extent the true risk-return trade-off available to the investor. The Galton method by comparison estimates a Sharpe ratio of 0.62, an excessively optimistic estimate too but one order of magnitude closer to a sensible value.

[11] As a side note, the contrast between the two methods is so extreme that it is hard to find a scale where both share the familiar textbook shape of a minimum variance frontier. If presented in an interval too wide the Galton method looks like a horizontal line, with a too narrow interval the historical method produces an estimate that looks more like a vertical line parallel and almost overlapping with the y-axis.

7. The result of simulations

The results in the previous section are based upon only one sequence of 43 stock universes (each universe comprising either 30 or 50 stocks) covering an OOS period of 526 months. In spite of the long period, there is still significant sampling error. To handle this I simulate 1000 such sequences of 526 months, resulting in 43,000 stock universes with a total of 526,000 OOS returns. All returns are OOS, so within each sequence, the investor following a strategy only uses the available data up to the month in question. The simulations use the actual OOS returns of the stocks in each portfolio, so they do not assume a multivariate normal distribution data generating process as in, for example, Jobson and Korkie (1980) or Kan and Smith (2008).

Table 4 shows the summary of the performance of these strategies in the OOS simulations. In panel A, with portoflios of 30 stocks, on average the Sharpe ratio of the historical GMV is 0.24. This less than the 1/N strategy and it only outperforms the naïve strategy in 22% of the simulations.[12] The investor expects, ex ante, a standard deviation of 7.48 percentage points, but OOS the actual standard deviation is 17.08 percentage points, more than double his expectation. In 100% of the simulations, OOS risk (as measured by the standard deviation) is higher than the ex ante expectation. So even if the strategy delivers a reasonable performance, it is consistently inferior to the naïve portfolio in OOS simulations and it also surprises investors with risks substantially higher than anticipated.

The Elton-Gruber constant correlation matrix GMV has a Sharpe ratio very

[12]In unreported results I found that on average the 1/N strategy loads more on the size and value Fama-French factors than the other portfolios. This possibly also contributes partially to its consistent performance.

close to the 1/N on average. In fact, in 54% of the simulations it outperforms the naïve benchmark.[13] It is noteworthy that this approach to portfolio management, proposed in 1973, has performed so well out of sample in the context of individual stock portfolios. Still the ex-post risk of the strategy is 44% higher on average than the ex ante estimates. Hence the constant correlation matrix systematically underestimates the risk of the GMV.

Both the historical and Elton-Gruber mean variance portfolios show a consistently bad performance OOS. The average Sharpe ratio is close to 0 and the risk OOS is more than 10 times higher the ex ante estimates from the respective covariance matrices. This shows that even the constant-correlation matrix has non negligible difficulties estimating the risk of concentrated stock portfolios.

The methods that correct past OOS errors have, on average, Sharpe ratios of 0.40 (GMV) and 0.36 (MV). Both are higher than the Sharpe ratio of the 1/N strategy (0.33). The differences in OOS performance are statistically significant at the 1% level and the second column shows the outperformance occurs in most simulations (85% for the GMV and 63% for the MV). While statistically significant, the economic gains in terms of Sharpe ratio are relatively small (9% and 21% higher than the 1/N for the MV and the GMV portfolios respectively). This should be partially expected as the optimization intentionally ignores all stock characteristics and these are relevant forming portfolios (Brandt et al. (2009)).

The performance of the MV portfolio is not as impressive as the GMV. This suggests that for individual stocks the most relevant information from the correction for the optimization is in the covariance matrix and not the vector of mean past

[13] All values in the second column are statistically different from 50% in two-tailed tests at the 1% level.

returns.

For risk management purposes, the most relevant issue is the OOS predictability of risk for each stock portfolio. The results show ex ante estimated risk is, on average, 16.21 percentage points for the GMV and 18.99 percentage points for the MV. This compares to ex post OOS risk of 14.21 percentage points and 19.72 percentage points, respectively. So the most noticeable result is that OOS risk is close to the ex ante estimate when using the corrected covariance matrix. This contrasts sharply with the historical and constant correlation approaches. Even the MV portfolio, that shows dismal performance OOS and unpredictable risk with the other methods, achieves to outperform the 1/N on average. This performance is achieved without portfolio constraints or using stock characteristics.

The results in Panel B with portfolios of 50 stocks are broadly consistent with those obtained with portfolios of 30 stocks. The most striking difference is that the historical GMV performs worse with more stocks to form the portfolio. Its average Sharpe ratio decreases substantially in simulations from 0.24 to 0.11. The method also becomes less accurate estimating risk with ex-ante volatility of only 3.96 versus 28.51 observed ex post on average (comparing to 7.48 versus 17.08 with $N = 30$). This illustrates that the higher the dimensionality of the covariance matrix the higher also are the chances of in-sample over-fitting and finding portfolios with implausible (and misleading!) low risk.

Table 5 shows the hit rates in the OOS period for the 1000 simulations. For the historical and the constant correlation approaches, almost all hit rates for the extreme quantiles are statistically different than the target. Extreme observations happen consistently more frequently than the ex ante risk estimate would suggest. Typically, the problem is more pronounced in the left tail than in the right tail and further exacerbated in portfolios with more stocks. Losses that should only happen 1% of the time, according with ex ante estimation of risk, occur with a frequency between 5.30% and 48.64% on average. This shows both historical and constant correlation covariance matrices leave considerable scope for investors to be mislead (or to mislead) about the true OOS risk of their stock portfolios - a conclusion with potentially important implications for regulators and Central Clearing Counterparties.

For the portfolios using past OOS errors to correct the inputs, the hit rates are, on average, close to the target levels. In fact, they are significantly below the target in 7 cases and insignificantly different from the target in 10 cases. They still capture insufficiently the risk in the extreme left tail (hits at the 1% level or below). In panel A, losses that should happen with 1% probability occur in the OOS period with 1.40% and 1.80% frequency for the GMV and MV portfolios, respectively. This is consistent with the interpretation that the distribution of returns is not multivariate normal. Still this is a relatively minor deviation in the VaR estimate when compared to the other methods. This shows that correcting the covariance matrix for past OOS errors has the potential to improve the estimation of risk of concentrated stock portfolios, in particular in extreme quantiles of the distribution.

8. Optimization with characteristic-sorted portfolios

As a robustness check, I examine the OOS performance of the Galton correction with other test assets, namely the portfolios sorted on different characteristics in

Kenneth French's website.

These test assets are very different from the individual stocks discussed in the sections before. The characteristics used to sort the stocks into different portfolios are purposely chosen to better capture a dispersion in expected returns (and a corresponding factor structure in risk). As such the persistence in correlations, variances, and mean returns should be higher for these portfolios. Table 6 shows the slope of ex post OOS values of the inputs of optimization on their ex ante estimates from the historical sample.

[Insert table 6 about here]

Generally, correlations between portfolios do not seem to regress toward the mean as much as for individual stocks. The slope is significantly below one for the double sorted portfolios on operating profitability and investment (panel B) and those double sorted on size and momentum (panel D) but not for the other two sets of portfolios. There is no evidence of mean reversion in returns in any of the test assets. For three of the sets, the slope coefficient is significantly positive. Still, for all mean returns, the slope is below one with statistically significance, so there is prevalent regression to the mean in returns. All of these results suggest the sorting exercises are successful in creating a set of portfolios with strong structures of both risk and return.

But there are important differences in the degrees of persistence between the sets of portfolios. For instance, returns are highly predictable in the set of size and momentum double-sorted portfolios. The R-square of the regression of ex post returns on ex ante regressions is 30.89 percentage points, the highest among the sets of assets compared. The slope coefficient is also closer to unity than 0, unlike the other sets.

On the other extreme, the portfolios sorted on size and beta have the most predictable risk. The slope coefficients for the covariances, correlations, and variances are all very close to one and the R-squares for each of these regressions is the highest when compared to the other test assets. So past risk is a very good predictor of future risk in this set of assets. On the other hand, regression to the mean in returns is practically total. The slope coefficient for this variable is not even significantly above zero implying that past returns are not informative of future returns on average.

All in all, this shows that these sets of assets are very different from individual stocks and also have important differences between themselves. Given that they differ in the predictability of their risk and returns, they provide a natural test of the flexibility of the Galton method.

[Insert table 7 about here]

Table 7 shows the results for each set of test assets. For each set of assets I use a rolling window of 120 months to estimate the moments and correct those estimates using past OOS errors.[14] The 1/N rule produces a consistent performance in terms of Sharpe ratio between 0.54 and 0.59, depending on the test assets. The historical GMV portfolio performs quite well with these sets of assets. It benefits from the fact that the covariance matrix estimated from past data is informative about OOS risk. But the performance of the historical MV portfolio is far from consistent across panels. It has the worst Sharpe ratios in panels A and C, but a very high Sharpe ratio in panel D of 1.29.

[14](DeMiguel et al. 2009b) uses the same window for the Kenneth French portfolios in their study. Using 60 months as for individual stocks does not change substantially the results.

The covariance matrix estimated with the historical method fails substantially in estimating the OOS risk of the ex ante tangent portfolio. In the case of operating profitability and investment portfolios, the H-MV has an OOS volatility of 28.83 percentage points, considerably below the ex ante estimate of 70.60. With the portfolios double-sorted on size and beta the opposite occurs: the ex ante volatility is of 117.87 percentage points while ex post the portfolio has a volatility of 340.01, almost three times the expected risk. Either way, the estimates are far from ex post risk using the historical estimates for the mean-variance optimal risky portfolio.

The Elton-Gruber constant correlation matrix produces far from accurate risk estimates for characteristic-sorted portfolios. For example, in panel D, the EG-MV ex post risk is more than 19 times higher its ex ante estimate. This contrasts with individual stocks where the constant correlation approach produces generally sensible results. This underscores the difficulty in estimating the appropriate amount of shrinkage to the covariance matrix of a set of assets. The results suggest a shrinkage method that serves its purpose with a set of test assets does not necessarily achieve the same in other sets.

The Galton method is quite flexible and for each set of test assets it seems to produce sensible corrections. The Sharpe ratio of both the tangent and the minimum variance portfolios are always above the 1/N and among the best performing methods. In most of these sets of assets past returns predict future returns in the cross section. The minimum variance portfolio does not use this predictability but the ex ante MV tangent portfolio does. Due to this the G-MV always has the highest Sharpe ratio compared to the other methods. In the case of the portfolios sorted on size and momentum the OOS Sharpe ratio on the G-MV portfolio is 1.37, more than double the 0.54 of the 1/N benchmark. Besides the G-MV portfolio, the

G-GMV also performs very well. On average across the four panels, the Sharpe ratio of the G-MV is 80% higher than that of the 1/N. Furthermore, the respective estimated covariance matrix produces risk estimates that are reasonably close to ex post risk. This reinforces the main conclusion of the analyses with individual stocks: equipped with an appropriate shrinkage method for the covariance matrix, the predictability in risk per se is enough to produce optimized portfolios that beat substantially the 1/N strategy OOS.

In conclusion, this robustness exercise shows that the benefits of the Galton method are not specific to portfolios of individual stocks. In fact, looking at the Sharpe ratios, the Galton optimization produces much better results with these sets of assets where returns and risk show more predictability.

9. Conclusion

The historical and constant correlation approaches to the estimation of the covariance matrix systematically underestimate the risk of optimal portfolios of individual stocks. The actual OOS standard deviation of the strategies is up to more than 100 times higher the ex ante estimate would suggest. The problem is more severe for the MV portfolios but is also pertinent in the GMV. This leads to a gross mis-estimation of risk, particularly at the extreme quantiles of the distribution. In simulations with real stock returns, losses exceeding the 1% level VaR occur in as much as 48.64% of the OOS periods (for the historical MV portfolio).

The ex ante estimates of the minimum variance frontier using the "plug in" approach are counter-intuitive. They go against both economic fundamentals and historical evidence on the risk-return trade-off. Hypothetical investors using the “plug-in” approach to Markowitz optimization should strongly suspect the optimization inputs even before being surprised with the subsequent dismal OOS performance of the respective optimized portfolios. The ex ante estimates of risk and return are extremely implausible. They suggest annualized Sharpe ratios as high as 11 are attainable investing in equities while the historical reward for holding the stock market has a whole produces a Sharpe ratio of around 0.4. The actual OOS Sharpe ratios of those portfolios are close to zero.

Correcting the covariance matrix for past OOS errors dramatically reduces this estimation problem. The ex post risk of the optimal portfolios is very close in magnitude to its ex ante risk. For most of the extreme quantiles, the hit rates are either not statistically different from the respective targets or even below them. This suggests that correcting past OOS errors provides a simple method to estimate the covariance matrix, with potentially useful applications in risk management.

In sets of assets where returns show some predictability, the Galton correction uses this on top of the predictability of risk. There the mean-variance efficient portfolio outperforms both the 1/N benchmark and even the global minimum variance portfolio.

Table 1

Regression to the mean

Regression of ex post values on historical estimates. For each month in the sample from 1955:03 to 2009:12, I regress the ex post value of some variable (computed using the subsequent 12 months of data) on its estimate from the historical sample in the previous 60 months. The variables in columns are: i) the pairwise covariance of stock returns; ii) the pairwise correlation of stock returns; iii) the variance of individual stocks; iv) the mean return of the individual stocks. The rows show the output of Fama-MacBeth (1973) regressions of the ex post values on the historical estimates. The outputs are: i) the average intercept coefficient in the monthly regressions; ii) the Newey-West (1987) (NW) t-statistic of the slope coefficient (computed with 12 lags); iii) the average slope coefficient in the monthly regressions; iv) the NW t-statistic of the slope coefficient; v) the percentage of cross sectional regressions where the slope coefficient is significantly positive in a one-tailed test at a significance level of 5%; vi) the NW t-statistic of the null that the true coefficient is one; vii) the percentage of regressions where the slope coefficient is significantly smaller than one in a one-tailed test at a significance level of 5%; viii) the average R-square of the regressions. Rows 9 to 11 show, respectively, the minimum, average, and maximum number of observations in the monthly regressions.

	Covariance	Correlation	Variance	Mean return
Intercept	0.0018	0.1490	0.0105	0.0093
t-stat (=0)	8.19	12.12	7.16	5.19
Slope	0.37	0.23	0.58	-0.18
t-stat(=0)	10.03	21.44	11.07	-4.83
Greater than 0 (%)	100.00%	100.00%	98.18%	18.54%
t-stat(=1)	-17.12	-71.16	-8.01	-32.06
Smaller than 1 (%)	93.01%	100.00%	86.93%	100.00%
R-square	4.16%	1.59%	13.24%	2.14%
Min	371091	371091	862	862
Average	3297770	3297770	2353	2353
Max	6579378	6579378	3628	3628

Table 2

OOS performance of the portfolio

Each twelve months I select the 50 stocks of the firms with the largest market capitalization for which there is a complete return history over the previous 60 months and the subsequent 12 months. The sample is kept fixed for the subsequent twelve months. The weights are re-balanced monthly, each month I use a rolling window of 60 months to estimate the covariance matrix hence obtaining the global minimum variance (GMV) and the mean-variance (MV) portfolios for three different methods: the historical method, the Elton-Gruber method, and the Galton method. The columns show descriptive statistics of the out-of-sample performance of each portfolio. These are: i) the mean annual return of the portfolio; ii) the annualized standard deviation of the portfolio; iii) the excess kurtosis of the portfolio; iv) the skewness of the portfolio; and v) the Sharpe ratio of the portfolio. The sample returns are from 1955:03 to 2010:12.

Strategies	Mean	STD	KURT	SKEW	Sharpe
1/N	4.70	16.05	1.92	-0.22	0.29
Historical GMV	5.15	26.50	1.46	0.20	0.19
Historical MV	-2051.12	15105.21	515.78	-22.69	-0.14
Elton-Gruber GMV	6.95	15.38	1.49	-0.10	0.45
Elton-Gruber MV	-189.69	1224.86	284.40	-13.94	-0.15
Galton GMV	6.14	12.71	1.21	-0.27	0.48
Galton MV	7.95	18.68	5.83	0.54	0.43

Table 3

Accuracy of risk estimates

The first row shows the expected standard deviation of a strategy in the OOS period using the weights of the strategy and the respective estimate of the covariance matrix. The second row shows the realized standard deviation of the strategy computed from its OOS monthly returns. Both measures of standard deviation are annualized. Rows 3 to 5 show how often the strategy delivered returns smaller than the 1st, 5th, and 10th quantile, respectively, according to the ex-ante estimate of volatility and assuming a normal distribution. Rows 6 to 8 show the same information for returns above the 90th, 95th, and 99th quantile, respectively. In the first column the variance is estimated from the historical sample, in the second column it is estimated with the Elton-Gruber correlation matrix, and in the third column with the Galton correction. Panel A shows the results for the global minimum variance portfolios and panel B for the mean-variance portfolios. The out-of-sample returns are from 1966:04 to 2010:01. All values are in percentage points. One star denotes significance at the 10% level, two stars at the 5% level, and three stars at the 1% level for the (two-tailed) test that the OOS hit rate is different than the expected ex ante.

	Historical	Elton-Gruber	Galton
Panel A: The global minimum variance portfolio			
STD(expt)	3.60	8.85	15.24
STD(realized)	26.50	15.38	12.71
r<Qz(0.01)	28.72***	5.93***	1.52
r<Qz(0.05)	32.37***	10.49***	4.56
r<Qz(0.10)	34.04***	14.13***	7.41*
r>Qz(0.90)	32.98***	17.63***	4.94***
r>Qz(0.95)	31.91***	14.44***	2.85**
r>Qz(0.99)	28.88***	8.51***	0.76
Panel B: The mean variance portfolios			
STD(expt)	133.52	172.95	18.20
STD(realized)	15105.21	1224.86	18.68
r<Qz(0.01)	39.67***	15.35***	1.33
r<Qz(0.05)	43.47***	25.23***	5.32
r<Qz(0.10)	45.59***	31.61***	7.98
r>Qz(0.90)	22.49***	8.21	6.65**
r>Qz(0.95)	21.58***	5.62	4.75
r>Qz(0.99)	17.93***	3.19***	1.14

Table 4
Results of simulations (OOS)

Each simulation selects randomly a set of stocks (30 in panel A and 50 in panel B). The stocks are selected out the 500 stocks with largest market capitalization. It forms optimal portfolios with information available up to each moment in time and rebalances them monthly, reflecting the update in the information available. Every twelve months I select another set of stocks until the end of the 526 months OOS period. So a simulation comprises 43 stock universes and a total of 526 OOS monthly returns. The results are based on 1000 such simulations totalling 526,000 OOS monthly returns. The first column shows the average Sharpe ratio across the 1000 simulations for the respective strategy. The second column shows the percentage of simulations the Sharpe ratio of the strategy was superior to that of the 1/N strategy. The third column shows the average expected volatility of the strategy using the ex-ante covariance matrix. Column 4 shows the actual ex post volatility of the strategy. Column 5 shows the percentage of simulations the ex post risk was higher than the one expected ex ante. The values in columns 3 to 5 are in percentage points.

Strategies	$\overline{SR}$	$SR > SR_{1/N}$	$E\sigma$	σ_{expost}	$\sigma_{expost} > E\sigma$
Panel A: Portfolios of 30 stocks					
1/N	0.33	-	-	-	-
Historical GMV	0.24	0.22	7.48	17.08	100.00
Historical MV	-.05	0.00	193.02	5556.92	99.90
Elton-Gruber GMV	0.34	0.54	9.90	14.25	100.00
Elton-Gruber MV	-0.06	0.00	243.15	3178.24	100.00
Galton GMV	0.40	0.85	16.21	14.21	0.00
Galton MV	0.36	0.63	18.99	19.72	80.20
Panel B: Portfolios of 50 stocks					
1/N	0.34	-	-	-	-
Historical GMV	0.11	0.06	3.96	28.51	100.00
Historical MV	-0.03	0.00	428.00	22681.05	100.00
Elton-Gruber GMV	0.33	0.45	8.83	15.07	100.00
Elton-Gruber MV	-0.07	0.00	307.07	4832.25	99.90
Galton GMV	0.40	0.82	15.15	13.36	0.00
Galton MV	0.36	0.57	18.52	20.92	100.00

Table 5

Hit rates in the OOS simulations

Each simulation selects randomly a set of stocks (30 in panel A and 50 in panel B) and uses the information available up to each point in time to form a portfolio. It rebalances the portfolio monthly, reflecting the update in the information available, and every twelve months selects another set of stocks until the end of the 526 months OOS period. So a simulation comprises 43 stock universes and a total of 526 OOS monthly returns. The results are based on 1000 simulations totalling 526,000 OOS monthly returns. Columns 1 to 3 show the average hit rates of each simulation for the shown critical level (the average percentage of observations with losses greater than the ex-ante estimated VaR). Columns 4 to 6 show the same information for hit rates above the respective critical level in the right tail of the distribution (that is gains exceeding the respective quantile of the ex-ante distribution). Three stars denote significance at the 1% level, two stars at the 5% level and one star at the 10% level. All tests are two-tailed tests. The null hypothesis is that the hit rate is equal to the respective critical level. The rows show the results for the following portfolios: i-ii) the historical global minimum variance ('H-GMV') and mean variance('H-MV'); iii-iv) the constant correlation matrix global minimum variance ('EG-GMV')and mean variance ('EG-MV'); v-vi) Galton corrected global minimum variance ('G-GMV') and mean variance ('G-MV').

Strategies	Hit <1%	Hit<5%	Hit<10%	Hit>90%	Hit>95%	Hit>99%
Panel A: Portfolios of 30 stocks						
H-GMV	14.58***	22.44***	27.70***	25.93***	20.84***	13.37***
H-MV	25.09***	35.89***	42.49***	15.27***	11.86***	7.15***
EG-GMV	5.30***	11.15***	16.20***	17.24***	11.73***	5.39***
EG-MV	11.49***	22.38***	30.94***	7.22***	4.53	1.85***
G-GMV	1.40***	4.36	7.73**	5.42***	2.62***	0.74
G-MV	1.80**	5.32	9.64	7.67**	4.39	1.46
Panel B: Portfolios of 50 stocks						
H-GMV	36.42***	40.54***	42.79***	40.93***	38.72***	34.73***
H-MV	48.64***	53.01***	55.33***	29.54***	27.64***	24.35***
EG-GMV	7.91***	14.19***	19.05***	21.96***	16.18***	8.75***
EG-MV	17.73***	30.02***	38.62***	7.72**	5.30	2.60***
G-GMV	1.46	4.49	8.15*	5.41***	2.65***	0.78
G-MV	2.25***	6.43**	11.28	8.79	5.26	1.93***

Table 6

Persistence(s) in characteristic-sorted portfolios

The test assets in each panel are 25 double-sorted portfolios on two different characteristics obtained from Kenneth French's online data library. I use a rolling window of 120 months to estimate moments and regress the subsequent values in the following 12 months on the historical estimates. The table shows the results of Fama-MacBeth regressions of ex post values on ex ante estimates. The first column shows the estimate of the slope (that is, the average slope across all cross sectional regressions); column 2 shows the t-statistic for the null hypothesis of the true slope being zero; the third column shows the t-statistic for the null of the slope being equal to 1; the fourth column shows the average R-square of the cross sectional regressions in percentage points. In panel A the test assets are the 25 portfolios sorted on size and book to market; in panel B the portfolios sorted on operating profitability and investment; in panel C the portfolios sorted on size and beta; and in panel D the portfolios sorted on size and momentum.

	Slope	tstat(=0)	tstat(=1)	R-square
Panel A: Size and value				
covariance	0.83	7.18	-1.47	46.03
correlation	0.92	13.82	-1.27	37.95
variance	0.83	6.76	-1.34	47.76
mean return	0.38	3.46	-5.66	16.82
Panel B: Operating profitability and investment				
covariance	0.80	6.60	-1.65	23.43
correlation	0.49	8.82	-9.35	5.23
variance	0.86	6.23	-1.06	28.70
mean return	0.30	3.38	-8.01	9.64
Panel C: Size and beta				
covariance	0.99	7.96	-0.10	70.00
correlation	0.98	9.67	-0.20	46.37
variance	1.00	7.39	0.00	70.84%
mean return	0.23	1.27	-4.30	17.63
Panel D: size and momentum				
covariance	0.83	7.79	-1.57	34.16
correlation	0.80	16.56	-4.18	37.27
variance	0.84	7.53	-1.39	35.43
mean return	0.69	8.71	-3.84	30.89

Table 7

Performance with characteristic-sorted portfolios

The out-of-sample performance of the optimized portfolios with different test assets. The test assets in each panel are 25 double-sorted portfolios on two different characteristics obtained from Kenneth French's online data library. I use a rolling window of 120 months to estimate moments and correct those based on their historical ability to forecast out-of-sample moments in the subsequent 12 months. The estimation requires an additional period of 120 months to start correcting the past OOS errors. In panel A the test assets are the 25 portfolios sorted on size and book to market; in panel B the portfolios sorted on operating profitability and investment; in panel C the portfolios sorted on size and beta; and in panel D the portfolios sorted on size and momentum. The strategies shown are the $1/N$, the historical global minimum variance ('H-GMV'), the historical mean variance('H-MV'), the constant correlation matrix global minimum variance ('EG-GMV')and mean variance ('EG-MV'); the Galton-corrected global minimum variance ('G-GMV') and mean variance ('G-MV'). For each strategy and set of test assets the information shown is the Sharpe ratio, the ex-ante volatility and its ex-post counterpart. The volatilities are annualized and in percentage points.

	Panel A: size and value			Panel B: OP and investment		
Strategies	SR	$E\sigma$	σ_{expost}	SR	$E\sigma$	σ_{expost}
1/N	0.58	-	16.76	0.58	-	15.51
H-GMV	0.86	10.51	12.97	0.89	10.18	13.54
H-MV	-0.11	191.12	516.21	0.68	70.60	28.83
EG-GMV	0.48	7.26	18.37	0.93	7.54	13.84
EG-MV	0.23	94.46	403.40	0.68	152.12	29.84
G-GMV	0.84	9.80	12.35	0.97	11.83	12.71
G-MV	1.01	13.38	16.18	0.98	15.17	16.55
	Panel C: size and beta			Panel D: size and mom		
Strategies	SR	$E\sigma$	σ_{expost}	SR	$E\sigma$	σ_{expost}
1/N	0.59	-	16.92	0.54	-	17.22
H-GMV	0.69	8.01	11.04	0.95	10.29	12.49
H-MV	0.06	117.87	340.01	1.29	35.11	44.44
EG-GMV	0.58	4.85	14.05	0.57	7.90	15.30
EG-MV	0.27	26.09	84.26	0.21	188.10	3611.82
G-GMV	0.70	8.97	10.61	0.95	11.97	11.77
G-MV	0.75	10.42	11.81	1.37	33.01	45.60

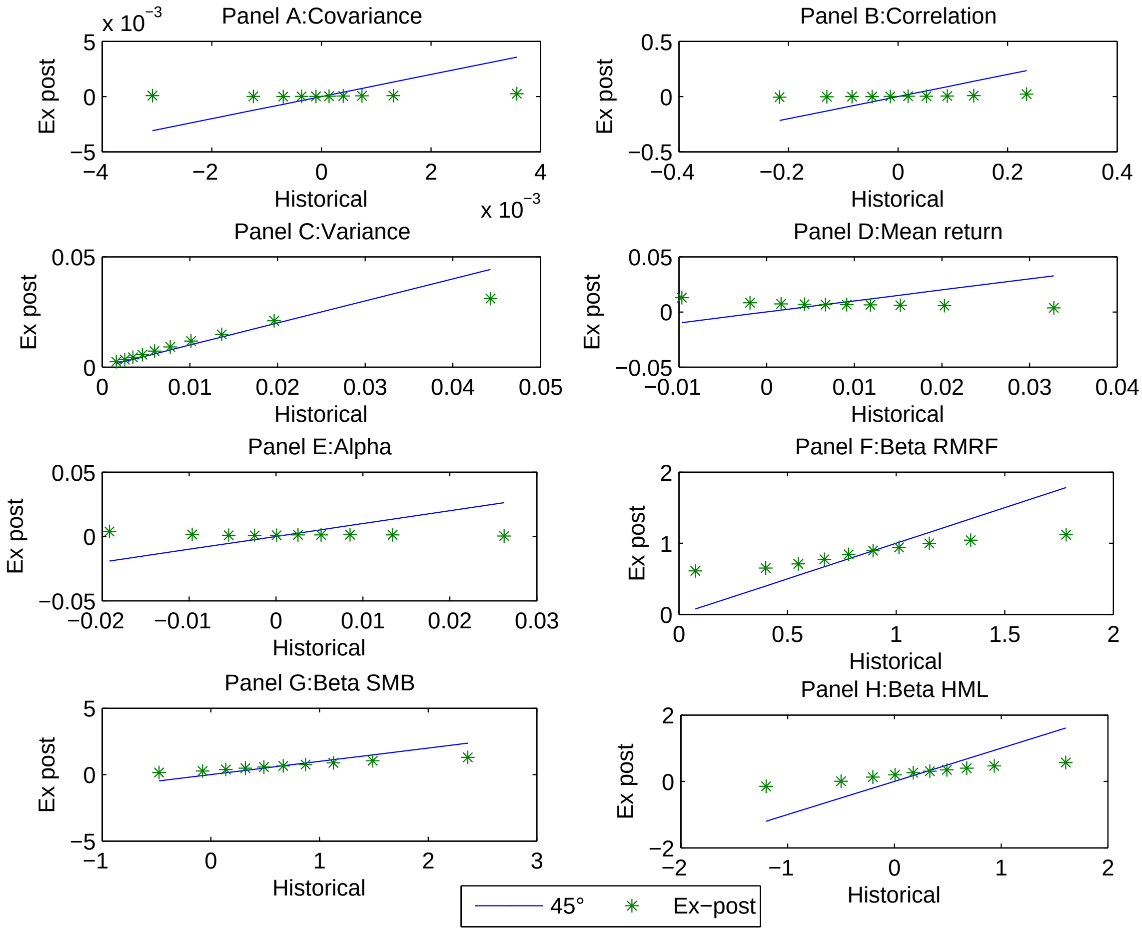

Fig. 1. Historical versus ex post realizations. For each moment in time and observation (either a stock or a pair of stocks) I compute the value of a variable in the historical sample - the previous 60 months of observations -and its realized value in the future -the subsequent 24 months. Then observations are classified into deciles according to their values in the historical sample each month. For each combination of month-variable-decile, I compute the mean value in the historical sample and in the following 24 months. The figure shows, for different variables, the time series average of the values in each decile of historical and respective ex post realizations. In panel A and B the observations are pairs of stocks while from C to H they are the stocks themselves. Panel A shows the historical covariance versus the ex post realized covariance of the idiosyncratic returns of individual stocks. The idiosyncratic returns of each stock are estimated regressing the stock monthly returns on the Fama-French (1992) factors. Panel B shows the same comparizon for the correlation of idiosyncratic returns. Panels C and D show the same comparizon for, respectively, the variance of residuals of stocks and the mean total return of each stock. Panels E to H show the same comparison for the alphas, and the betas with respect to the market ('RMRF'), the size ('SMB'), and the value ('HML') factors respectively. The data consists of monthly observations from 1955:03 to 2009:12.

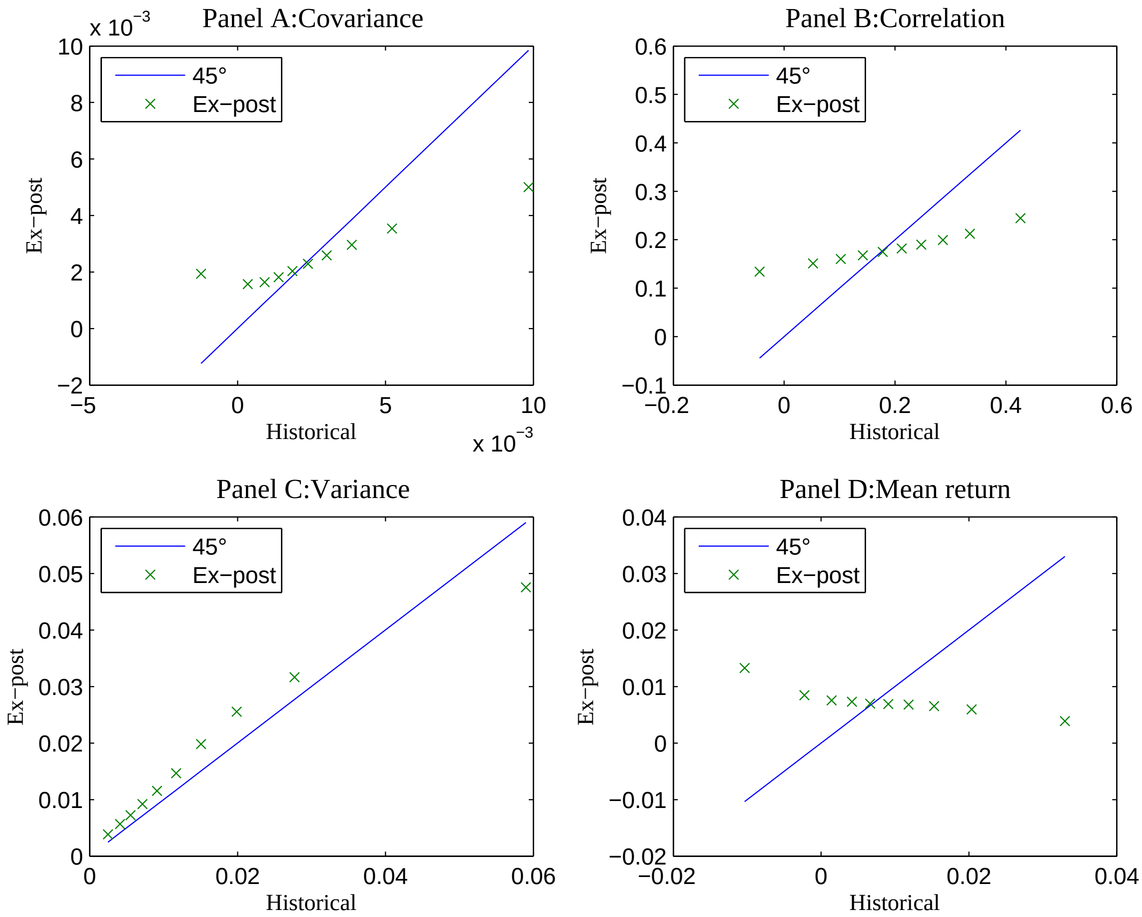

Fig. 2. Historical versus ex post values. For each moment in time and observation (either a stock or a pair of stocks) I compute the value of a variable in the historical sample - the previous 60 months of observations -and its respective value in the future -the subsequent 12 months. Then observations are classified into deciles according to their values in the historical sample each month. The figure shows, for different variables, the time series average of the values in each decile of historical and respective ex post realizations. In panel A and B the observations are pairs of individual stocks while in C and D they are the stocks themselves. Panel A shows the historical covariances versus the ex post covariances of the returns of individual stocks. Panel B shows the same comparizon for the pairwise correlations of returns. Panels C and D show the same comparizon for, respectively, the variance of stocks and the mean total return. The data consists of monthly observations from 1955:03 to 2009:12.

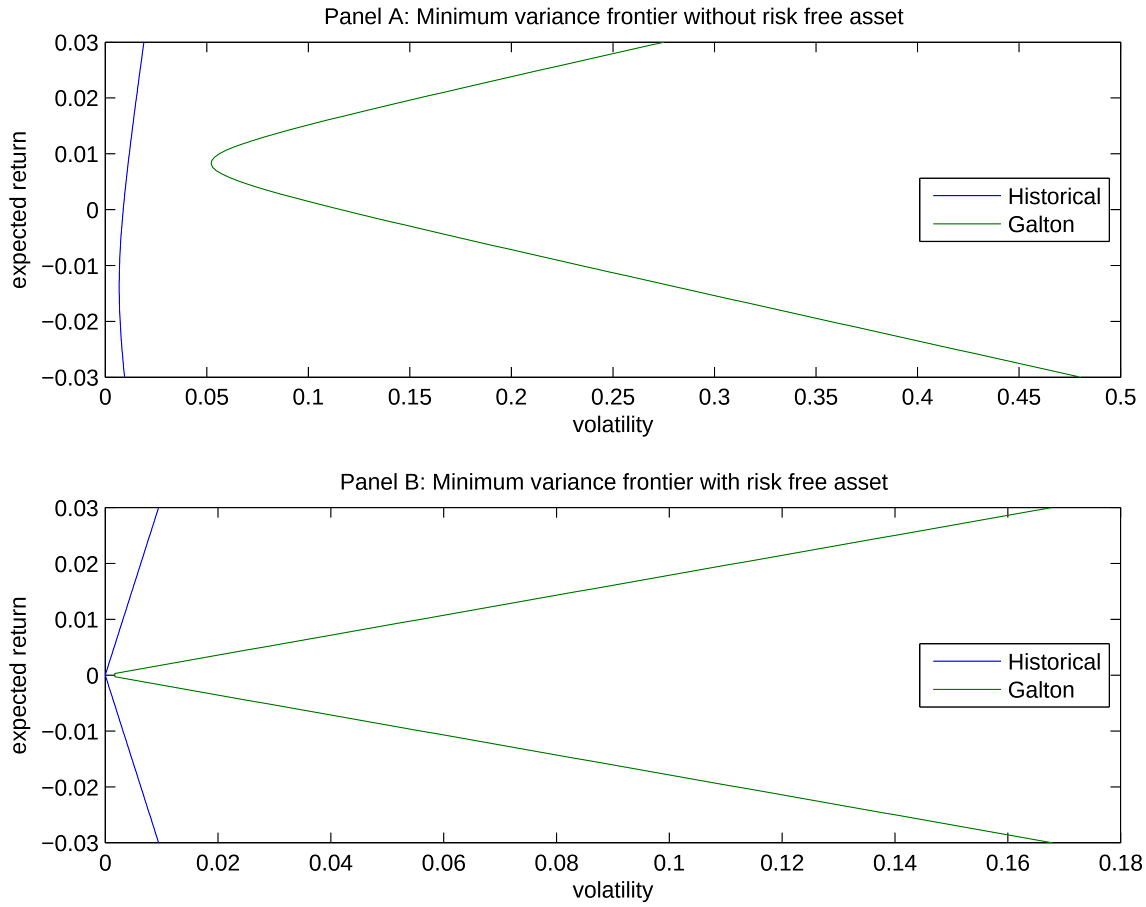

Fig. 3. Minimum variance frontiers in 2009:12. The ex ante minimum-variance frontiers estimated using the historical and the Galton methods for the 50 largest stocks by market capitalization in 2009:12. The historical estimate uses the previous 60 months of returns to estimate the covariance matrix and the vector of expected returns while the Galton method corrects these estimates using past OOS errors. Panel A shows the minimum variance frontiers without a risk free asset in the asset space and the panel B with a risk free asset.

References

Agarwal, V., Fos, V., and Jiang, W. (2013). Inferring reporting-related biases in hedge fund databases from hedge fund equity holdings. *Management Science*, 59(6):1271–1289.

Barry, C. B. (1974). Portfolio analysis under uncertain means, variances, and covariances. *The Journal of Finance*, 29(2):515–522.

Basak, G. K. (2005). *Estimating the Risk in Sample Efficient Portfolios1*. PhD thesis, Citeseer.

Bawa, V. S., Brown, S. J., and Klein, R. W. (1979). Estimation risk and optimal portfolio choice. *NORTH-HOLLAND PUBL. CO., N. Y., 190 pp.*

Benninga, S. (2014). *Financial modeling*. MIT press.

Best, M. J. and Grauer, R. R. (1992). Positively weighted minimum-variance portfolios and the structure of asset expected returns. *Journal of Financial and Quantitative Analysis*, 27(04):513–537.

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

Brandt, M. W., Santa-Clara, P., and Valkanov, R. (2009). Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns. *Review of Financial Studies*, page hhp003.

Brennan, M. J. (1998). The role of learning in dynamic portfolio decisions. *European finance review*, 1(3):295–306.

DeMiguel, V., Garlappi, L., Nogales, F. J., and Uppal, R. (2009a). A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Science*, 55(5):798–812.

DeMiguel, V., Garlappi, L., and Uppal, R. (2009b). Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *Review of Financial Studies*, 22(5):1915–1953.

Detemple, J. B. (1986). Asset pricing in a production economy with incomplete information. *The Journal of Finance*, 41(2):383–391.

Disatnik, D. J. and Benninga, S. (2007). Shrinking the covariance matrix. *The Journal of Portfolio Management*, 33(4):55–63.

Dothan, M. U. and Feldman, D. (1986). Equilibrium interest rates and multiperiod bonds in a partially observable economy. *The Journal of Finance*, 41(2):369–382.

Efron, B. (1983). Estimating the error rate of a prediction rule: improvement on cross-validation. *Journal of the American Statistical Association*, 78(382):316–331.

Elton, E. J. and Gruber, M. J. (1973). Estimating the dependence structure of share pricesimplications for portfolio selection. *The Journal of Finance*, 28(5):1203–1232.

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, 33(1):3–56.

Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1–22.

Fama, E. F. and MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *The Journal of Political Economy*, pages 607–636.

Galton, F. (1894). *Natural inheritance*. Macmillan.

Goyal, A. and Welch, I. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21(4):1455–1508.

Green, J., Hand, J. R., and Zhang, X. F. (2013). The supraview of return predictive signals. *Review of Accounting Studies*, 18(3):692–730.

Harvey, C. R., Liu, Y., and Zhu, H. (2015). ... and the cross-section of expected returns. *Review of Financial Studies*, page hhv059.

Jegadeesh, N. (1990). Evidence of predictable behavior of security returns. *The Journal of Finance*, 45(3):881–898.

Jobson, J., Korkie, B., and Ratti, V. (1979). Improved estimation for markowitz portfolios using james-stein type estimators. In *Proceedings of the American Statistical Association, Business and Economics Statistics Section*, volume 41, pages 279–284. American Statistical Association Washington DC.

Jobson, J. D. and Korkie, B. (1980). Estimation for markowitz efficient portfolios. *Journal of the American Statistical Association*, 75(371):544–554.

Jobson, J. D. and Korkie, B. M. (1981). Performance hypothesis testing with the sharpe and treynor measures. *Journal of Finance*, pages 889–908.

Jorion, P. (1985). International portfolio diversification with estimation risk. *Journal of Business*, pages 259–278.

Jorion, P. (1986). Bayes-stein estimation for portfolio analysis. *Journal of Financial and Quantitative Analysis*, 21(03):279–292.

Kacperczyk, M., Sialm, C., and Zheng, L. (2005). On the industry concentration of actively managed equity mutual funds. *The Journal of Finance*, 60(4):1983–2011.

Kan, R. and Smith, D. R. (2008). The distribution of the sample minimum-variance frontier. *Management Science*, 54(7):1364–1380.

Kirby, C. and Ostdiek, B. (2012). Its all in the timing: simple active portfolio strategies that outperform naive diversification. *Journal of Financial and Quantitative Analysis*, 47(02):437–467.

Ledoit, O. and Wolf, M. (2004a). Honey, i shrunk the sample covariance matrix. *The Journal of Portfolio Management*, 30(4):110–119.

Ledoit, O. and Wolf, M. (2004b). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of multivariate analysis*, 88(2):365–411.

Ledoit, O. and Wolf, M. (2014). Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets goldilocks. *Available at SSRN 2383361*.

Lewellen, J. (2014). The cross section of expected stock returns. *Forthcoming in Critical Finance Review*.

Markowitz, H. (1952). Portfolio selection*. *The journal of finance*, 7(1):77–91.

Michaud, R. O. (1989). The markowitz optimization enigma: is 'optimized'optimal? *Financial Analysts Journal*, 45(1):31–42.

Perignon, C. and Smith, D. (2007). Which value-at-risk method works best for bank trading revenues? Technical report, Working Paper, Simon Fraser University.

Pérignon, C. and Smith, D. R. (2010). The level and quality of value-at-risk disclosure by commercial banks. *Journal of Banking & Finance*, 34(2):362–377.

Plyakha, Y., Uppal, R., and Vilkov, G. (2012). Why does an equal-weighted portfolio outperform value-and price-weighted portfolios. *Available at SSRN*.

Shao, J. (1993). Linear model selection by cross-validation. *Journal of the American statistical Association*, 88(422):486–494.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the royal statistical society. Series B (Methodological)*, pages 111–147.

Vasicek, O. A. (1973). A note on using cross-sectional information in bayesian estimation of security betas. *The Journal of Finance*, 28(5):1233–1239.